THE DESIGN PROCESS OF AN AUTOSTEREOSCOPIC VIEWING INTERFACE FOR COMPUTER-ASSISTED MICROSURGERY

POLYDOROS CHIOS

UNIVERSITY COLLEGE LONDON

A Thesis Submitted to the University of London
for the Degree of

DOCTOR OF PHILOSOPHY

AUGUST 2002

ABSTRACT

This thesis describes and evaluates the design process of an autostereoscopic viewing interface for computer-assisted microsurgery using augmented reality techniques. The augmentation of three-dimensional real-time optical images, as these are acquired from the surgical microscope, with volumetric reconstructions from CT/MRI data, ensures that the surgeon is offered optimum visual information about the patient's pathology at the time of surgery, thus allowing him/her to perform highly complicated procedures with a great degree of accuracy.

It is well known that the working conditions in the operating theatre are uncomfortable since the surgeon has to operate by looking through a surgical microscope. The interface looks to improve the way surgeons perceive visual data, by introducing the use of a prototype 3D visual display unit that replaces the viewing end of the surgical microscope, thereby making surgery much less stressful for the surgeon and much less hazardous for the patient.

The design process of the prototype, based on Human Factors Engineering and Ergonomics, is divided into four cycles. In Cycle I, the author focuses on potential users and their needs during microsurgical operations. This is done by creating user physical and behavioural aspect models from in-situ video observations. In Cycle II, an original design for the proposed interface is established, followed by early and continuous usability testing of the prototype under development. The cycle concludes with prototype evaluation and a quantitative analysis of the interface's accuracy and precision. In Cycle III, the designer applies iterative design techniques in order to optimise the interface's effectiveness, particularly the sensation of depth during autostereoscopic viewing. Scientific work during this phase includes psychophysical experiments in stereo vision and a qualitative analysis of effective and pleasing viewing, based on ranking. Finally, in Cycle IV, the designer suggests an integrated design of the interface in the operating theatre, based on successful I/O control of the interface's broadcasting and communicative signal parameters. Research work during this phase comprises the design and implementation of a fibre optics network for real-time stereo imaging.

AKNOWLEDGEMENTS

The author would like to thank his supervisor, Dr. Alfred Linney, whose direction has been vital for the completion of the study. Furthermore, this work would not have been possible without the financial support of the Institute of Laryngology and Otology and the personal commitment of Professor Anthony Wright. The author would also like to extend his gratitude to the members of the Physics, Surgery and Instrumentation Group of University College London, particularly Mr. A C Tan and Mr. Gus Alusi for their insightful comments, as well as Mr. Cliff Ruff from the Medical Physics Department for lending his ear and revising the document.

Finally, the author would like to thank his parents for their support and encouragement toward academic achievements from early childhood until today. And Him, for being an incessant source of strength.

DEDICATIONS

To my sister Yianna, whose bravery and courage have been the inspiring forces behind this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS

Sections of this research work have been published previously:

- **Chios, P.**, Tan, AC, Linney, A. D., Alusi, G. H., Wright, A., Woodgate, G., Ezra, D. (1999). The potential use of an autostereoscopic 3D display in microsurgery. In C. Taylor & A. Colchester (Ed) *Medical Image Computing and Computer-Assisted Intervention MICCAI '99,* vol. 1679 of *Lecture Notes in Computer Science*, pp. 998-1009, Springer, September 1999
- **Chios, P.**, Tan A C (2000). Implementation of a Dedicated Optical Fibre Network for Real Time Stereo Imaging. In A. Hasman *et al* (Ed.) *Medical Informatics in Europe – MIE 2000*, volume 77 of Studies in Health Technology and Informatics, pp. 1151-1155, August 2000
- Lapeer, R. J., **Chios, P.**, Alusi, G. H., Linney, A. D., Davey, M. K., Tan, A C. (2000a). Computer Assisted ENT Surgery using Augmented Reality: Preliminary results on the CAESAR project. In *MICCAI'2000 Conference Proceedings*, October 2000
- Lapeer, R. J., **Chios, P.**, Alusi, G. H., Linney, A. D., Davey, M. K., Tan, A C. (2000b). Augmented reality for ENT surgery: an overview of the CAESAR project. In *ISA/ICC-2000 Conference Proceedings*, December 2000
- Lapeer, R. J., **Chios, P.**, Linney, A. D., Alusi, G. H., Wright, A. (2001). HCI: Chapter XV, The Next Step Towards Optimisation of Computer-Assisted Surgical Planning, Intervention and Training (CASPIT). In Chen (Ed) *Human Computer Interaction: Issues and Challenges*, pp. 232-246, Idea Group Publishing

GLOSSARY OF ABBREVIATIONS

Numbers in brackets indicate the chapters or appendices in which the abbreviation appears.

2AFC	Two-Alternative Forced-Choice (7)
3D	Three-Dimensional (1, 2, 3, 4, 5, 6, 7, 8, 9)
AI	Artificial Intelligence (2)
ANOVA	Analysis of Variance (7, B)
AR	Augmented Reality (1, 4, 8)
AV	Augmented Virtuality (2, 8)
CAESAR	Computer-Assisted ENT Surgery using Augmented Reality (1, 8, 9)
CAMI	Computer-Assisted Medical Interventions (1)
CAS	Computer-Assisted Surgery (1, 4, 5, 6, 7, 8, 9)
CASPIT	computer-assisted surgical planning, intervention and training (1, 4)
CBT	Computer-Based Training (4)
CCD	Charge Coupled Device (3, 4, 6, 7)
CCIR	International Radio Consultative Committee (8, C)
CCU	Camera Central Unit (6)
CIST	Computer Integrated Surgery and Therapy (4)
CMO	Common Main Objective (3, 7)
CT	Computed Tomography (1, 4, 8)
DIF	Design Information Framework (5, 9)
DIVO	DIgital Video Option (8)
DOF	Depth of Field (3, 7, 9, B)
ENT	Ear, Nose and Throat (1, 4, 5, 7, 8, 9, B)
FN	Field Number (3, 7)
FOV	Field of View (1, 3, 4, 7, 9)
HCI	Human Computer Interaction (1, 4, 5, 6, 7, 8, 9)
HIS	Hospital Information Systems (1)
HMD	Head Mounted Displays (3, 4)
ICS	Image Coordinate System (4)
IGS	Image Guided Surgery (1, 4)
IPD	InterPupillary Distance (1, 3)
ITU-R	International Telecommunications Union-Radio (8)
LAN	Local Area Network (8)
LCD	Liquid Crystal Display (1, 3, 6, 9)
LED	Light Emitting Diode (1, 4)
MAGI	Microscope-Assisted Guided Intervention (4)
MCS	Microscope Coordinate System (4)

MRI	Magnetic Resonance Imaging (1, 2, 4, 8)
NA	Numerical Aperture (3, 7)
NTSC	National Television System Committee (3, 4, 6, 8)
OHD	Off-Head Display (3)
OME	Otitis Media with Effusion (7)
PACS	Picture Archiving and Communication Systems (1)
PAL	Phase Alternation Line (3, 6, 8)
RGB	Red, Green, Blue (3, 6, 8, C)
ROI	Region of Interest (1, 3, 4, 7)
SECAM	SEquential Colour Avec Memoire (3, 8)
TFT	Thin-Film Transistors (3)
VCR	Video Cassette Recorder (3)
VIEW	Virtual Interface Environment Workstation (4)
VE	Virtual Environments (2)
VR	Virtual Reality (2, 3, 4, 9)
WCS	World Coordinate System (4)

(this page intentionally left blank)

CHAPTER 1

INTRODUCTION

Let the reader imagine the following futuristic narrative:

"... His first day at the new job had not worried him that much until that morning. Why should it anyway? The Giles Volare Institute might have been the best ENT hospital in the country; nonetheless, his record spoke for itself. It was the operation that was foremost on his mind. The Chameleon was a tricky tumour to deal with. Growing axially behind the trigeminal nerve makes it almost impossible to see, let alone removing it! This one looked even worse. While looking at the pre-op scans on his home PC the night before, he noticed that it was bundled right at the end of the nerve, nearly touching the brainstem. He tried a couple of surgical simulations on that amazing surgical-planning software from Takamitsi, trying different approaches, and then decided to use the posterior fossa. He ran the virtual fly-through, saved it, and uploaded the file to the hospital server.

His PDA bleeped as he was entering the hospital door. The patient was in the operating theatre, asleep. He walked into the room ten minutes later, after discussing a few details with Dr. Moyes, the assistant surgeon. He sat in front of the theatre's computer workstation and loaded his patient's electronic folder. He nodded Dr. Moyes to start the procedure while he began teleconferencing with Professor Rupier, at the University Medical School. The operation was going to be broadcasted live to trainee surgeons and a small introduction was required.

Everything was running smoothly. The skin was cleared and he was now ready to cut the bone. He placed the microscope over the lesion and switched on the computer-guidance system. The lesion appeared on the 3D display, augmented with the virtual reconstructions of the adjacent anatomical structures. He began burring the bone guided by the path line of the fly-through. Forty minutes later, he was there. He could see the virtual tumour residing deep in the image background, highlighted in neon blue colour. His burr, appearing in the foreground, was touching the surface of the display. He proceeded slowly and started removing the tumour bit by bit. At the same time, rendered views of the fly-through appeared on the terminal screen, the tracked position of the tool flashing in red. He smiled! He remembered the old X-ray light boxes his uncle used to have..."

1.1 Computer-assisted surgery

The introduction of computerised systems in medicine started more than a decade ago. The first applications were mainly focused on archiving and general database management of patient records with the aim of building fully integrated Hospital Information Systems (HIS) and fast transfer of data and images (e.g. PACS - Picture Archiving and Communication Systems) between HIS. At the same time, specialised computer systems were built to process and enhance image data from diagnostic modalities, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT). The use of enhanced CT and MRI images led to the birth of Image Guided Surgery (IGS). Other terminology for similar concepts has since been used, e.g. Computer-Assisted Surgery (CAS), Computer Integrated Surgery and Therapy (CIST) and Computer-Assisted Medical Interventions (CAMI) [1]. The term 'assisted' implies that the user, i.e. the surgeon, is still responsible for the outcome and the overall control of the operation. The system is there as an aid, thus the surgeon has to decide as to whether the information provided by the system is reliable or not.

Surgical simulation and planning was developed in parallel with CAS, possibly because of its solid base for testing new technologies which might prove too unreliable for direct use in a real clinical intervention. At the end of the 80's, simple 3D visualisation of the anatomy of the patient, mainly obtained from MRI and CT images, was already a major step forwards in the planning and simulation of surgery. Before that, surgeons had to reconstruct the human anatomy mentally, i.e. by observing a series of 2D images and then trying to mentally visualise the 3D structure. The development of IGS systems allowed displaying a virtual surgical tool into the 3D virtual scene, hence offering more realistic planning and execution of interventions like stereotactic and endoscopic surgery which previously provided the surgeon with only local visual information. Displaying the trajectory of the surgical tool in 2.5/3D was a useful aid to the surgeon even if the accuracy was not much better than a few millimetres. Today, CAS has become popular in a variety of surgical fields including cranio- and maxillo-facial surgery, ENT (ear, nose and throat) surgery, orthopaedic surgery, cardiac surgery, laparoscopy, interventional radiology and several other surgical disciplines [2].

1.1.1 Microsurgery

Microsurgery has always been considered as one of the most difficult and delicate types of surgery. This is mainly due to the limited amount of working space and the high density of anatomical structures that exist in the region of interest (ROI). The use of surgical stereo microscopes is universal, but they limit the surgeon's head movement. However, the instrument offers the surgeon stereoscopic viewing, which is important for the perception of depth. Surgical microscopes use a single objective lens that allows the eyepieces to receive two different views of the ROI, thus producing stereo vision effects. The field of view (FOV) under the objective lens covers a sensing area of 20-120 mm in diameter. Magnification and the physical properties of optical elements derive the depth of field, which is the area above and below the focal plane in the ROI that is still perceived as a sharp image. Further magnification in the instrument's eyepiece enhances the relative depth between two or more points. This viewing arrangement makes the surgical operation stressful, uncomfortable for the eyes and hence more hazardous for the patient.

The use of CAS systems in microsurgery becomes more difficult. This arises from several reasons. The most important one is that the surgical intervention now involves a 'not-natural' viewing interface, an optical instrument (*e.g.* a microscope). When the surgeon looks through a stereo microscope, his/her vision is accommodated to the microscope binoculars; s/he is 'immersed' in the microscopic environment. In order to see the computerised 3D reconstruction of the human anatomy, the surgeon needs to look away from the microscopic scene, at a computer monitor. The transition from microscope 'immersiveness' to computer workstation monitor requires a mental reorientation which relies completely on the surgeon's own skill and knowledge of the variations of the normal anatomy. This becomes more difficult when a surgical tool needs to be held in position.

The concept of **augmented reality** (AR) has become the new challenge in microsurgical disciplines such as ENT- and neurosurgery. Stereoscopic images, as obtained through the surgical microscope, are *augmented* with *virtual* overlays from 3D rendered patient data, e.g. from MRI or CT computer reconstructions. This allows the surgeon to view anatomical structures which are not visible in the original microscopic image, thus providing him/her with optimum visual information about the patient's pathology before, and at the time of surgery.

1.1.2 The CAESAR project

The research group from the Institute of Laryngology and Otology at UCL, London, has been conducting active research in the field of AR in ENT surgery [3]. The aim of its latest project, Computer-Assisted ENT Surgery using Augmented Reality (CAESAR), is to create a tool based on the concept of AR that will allow surgeons to realistically plan, rehearse and execute complex otological and skull-base surgery with a greater degree of accuracy, thus reducing the risk of morbidity, as well as deafness to the patient. The scientific and technical issues relate to creating dynamic 3D augmented images, integrating an accurate tracking technique and creating a suitable human-machine interface for viewing the augmented surgical microscopic scene.

The overall project, shown in Figure 1.1, involves research and development in the following fields:

- Development of an appropriate computer graphics technique for visualisation of 3D reconstructions of pre-operative CT/MRI scans with real time processing, image buffer updating and parallel processing,
- Calibration of the CAESAR components,
- Registration of pre-operative CT/MRI images to the patient's head,
- Tracking objects participating in the surgical procedure (e.g. drilling tools, patient's head, surgical microscope, etc.),
- Design of a 'heads-up' viewing interface,

The design of a 'heads-up' display also aims to tackle a basic constraint of conventional, and potentially computer-assisted, ENT surgery: the stressful viewing conditions that arise from using the microscope. Research into designing a 'natural' viewing interface, looks to improve the way the surgeon perceives the microscopic scene, augmented or not.

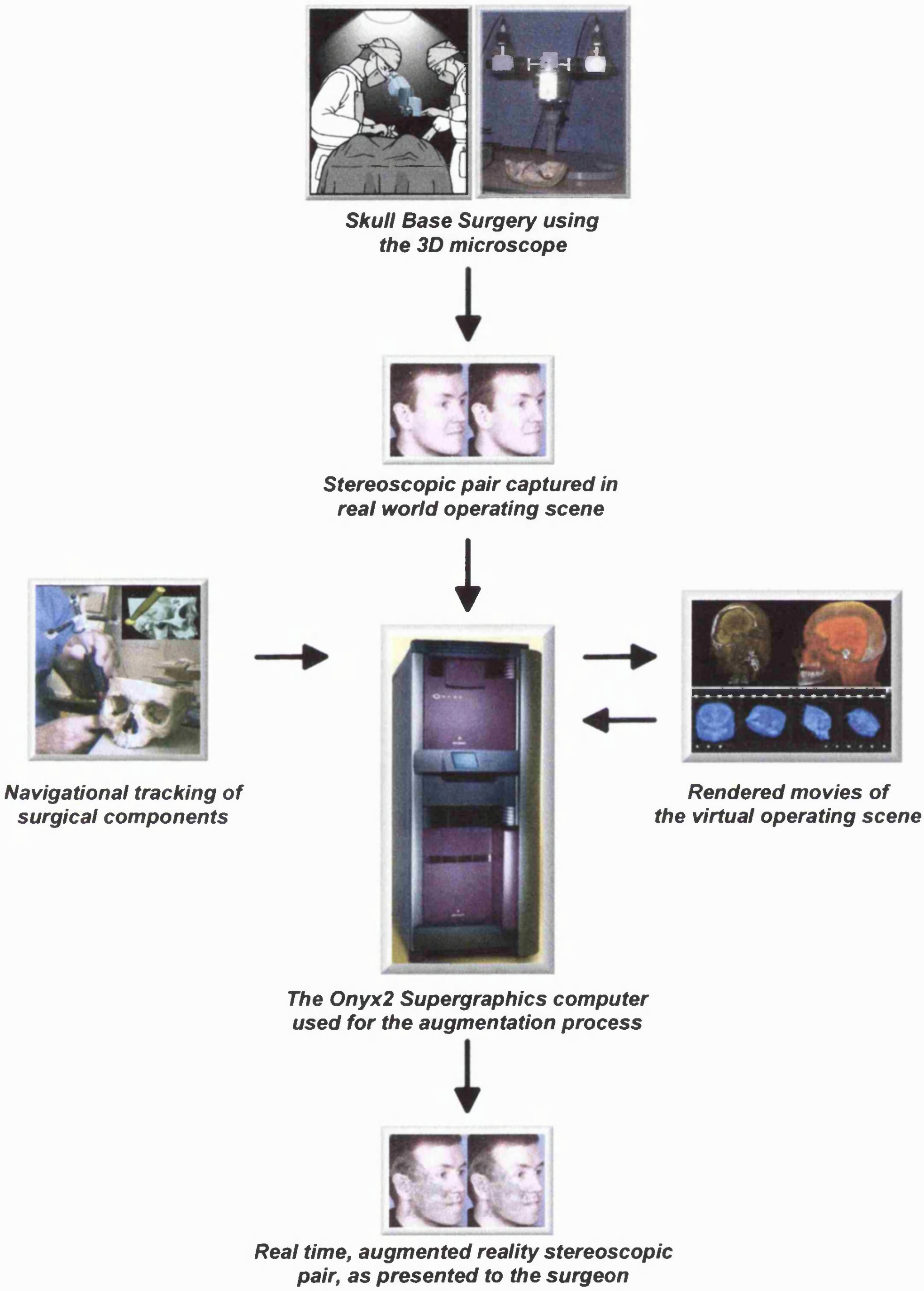


Figure 1.1: Schematic presentation of the CAESAR project.

1.2 Visual display engineering

Spatial vision is based on monocular and binocular depth cues contained in the observed scene. Monocular cues are visible with one eye while stereo vision is based on local displacements (disparities) between the 2D projections of the scene onto the left and right eye retinas. Stereo vision, along with pictorial and motion cues, is responsible for the perception of depth in real and virtual scenes [4].

Two-dimensional displays retrieve depth information from monocular (*e.g.* perspective, occlusion, shading, etc.) and motion cues. On the contrary, stereoscopic displays are based on stereo vision effects. Current electronic 3D display technology can provide true depth by displaying to each eye a disparate view of the visual scene, such that the right eye sees only the right-eye scene and the left eye sees only the left-eye scene. This is done by means of various techniques. In fact, 3D displays are classified by the technique used to channel the left and right images to the appropriate eyes (see Table 1.1) [5]. Some require optical devices close to the observer's eyes (stereoscopic), while others have the eye-addressing techniques integrated into the display itself (autostereoscopic). Adding depth information to a display helps it match the capabilities of the human visual system. As a result, the development of 3D display interfaces has become increasingly popular for use in teleoperation, simulation, entertainment and other multimedia applications [6].

Principle of eye addressing		Effective origin of waves	Number of different views	Head-tracking
Stereoscopic (aided-viewing)	multiplex (colour, time, polarisation, location)	fixed/gaze-controlled image plane,	two	optional (for a single observer)
Autostereoscopic (free viewing)	direction multiplex (diffaction, refraction reflection, occlusion)	fixed/gaze-controlled image plane	two	optional (for a small number of observers)
	volumetric display, electro-holography	depth slices entire space	unlimited	inherent (for a small number of observers)

Table 1.1: Taxonomy of 3D displays.

There is no doubt that if a display system is to be developed for 3D imaging applications like computer-assisted microsurgery, a more natural viewing approach is required. Autostereoscopic displays have the ability to present the content of a 3D scene in the same manner that this would appear under natural viewing conditions. The quality of a 3D scene depends on the following parameters:

- *Field-of-View*: The visual angle subtended by the display surface,
- *Spatial resolution*: The number, angular size and spacing of the pixels,
- *Refresh rate*: The frequency with which the display hardware can draw the image on the display surface,
- *3D content presentation*: Generating binocular disparity effects by displaying separate images for each eye,

Stereo image presentation is characterised by the stereo camera properties: the interpupillary distance (IPD), the distance between the nearest and furthest object shown, the location and orientation of the viewpoints and whether the object appears in front of or behind the screen.

1.3 Human factors

Recent advents in display and computing technology have revolutionised the way humans interact with machines. Now that state-of-art LCD panels are produced at relatively low cost, the time has come to focus on designing systems with maximised capabilities in ways most effective for the user.

Human Computer Interaction (HCI) is a discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use, and with the study of major phenomena surrounding them [7]. When designing a new interface, the importance of analysing not just the observer or the image presented but also the task to be performed is fundamental [8]. This implies that a human-centered design approach based on Human Factors and Ergonomics research is desired for evaluating the quality of a viewing interface.

The same approach needs to be followed when studying the effects of display parameters, such as FOV, number of pixels, pixel size, use of stereo imagery, and frame rate, as these determine the spatial and temporal capabilities of a viewing interface and thus have significant effects on VDS usability [9]. The designer must also assess their interactions to optimise for a given application. An effective viewing interface design results from trading-off the display parameters, in terms of both economics and visual perception [10].

Gould [11] suggests a four-cycle design methodology that provides good HCI in interface design. Each cycle attributes separately to the overall design of the new system.

- Focus on the User: The aim is to understand users' cognitive, behavioural, attitudinal and tasks characteristics.
- Early and Continual User Testing: This phase includes the empirical design of components, prototype development and evaluation.
- Iterative design: The aim is to optimise the design. The prototype must be modified upon the results of psychophysical tests and behavioural analysis.
- Integrated design: This phase involves all aspects of usability, such as help system, training plan and documentation.

Some design phases can and should be run in parallel. For example, interviews and training plans of the new system must be run in parallel in order to create a good user manual. Table 1.2 lists the various methodologies that can be employed in a good design process.

Focus on Users	Early User testing	Iterative design	Integrated design
Talk with users	Protocols	Follow-up studies	All aspects of usability
Visit customer locations	Video scenarios	Design modification	
Observe users working	Simulations	Development work	
Videotape users	Early prototyping	Optimisation	
Participative design	Early demonstrations		
Task analysis	Make videotapes		
Surveys	Formal prototype test		

Table 1.2: Design methodology for good HCI.

1.4 Research aim

The main aim of this thesis is to describe and evaluate the design of an autostereoscopic viewing interface for computer-assisted microsurgery. The interface looks to improve the way ENT surgeons perceive visual data, particularly real-time 3D video images. The proposed design introduces the use of a 13.8", prototype Sharp Micro-optic Twin-LCD, autostereoscopic 3D display[1] which replaces the viewing end of the surgical microscope, thus equipping the viewer with 'electronic eyes', and thereby making surgery much less stressful for the surgeon and much less hazardous for the patient.

The design process aims to:

- Derive physical and behavioural patterns of surgeon-microscope interaction during ENT surgery,
- Design, develop and evaluate a prototype 3D video surgical microscope,
- Optimise and evaluate the prototype's viewing comfort and sensation of depth,
- Design and implement an "in-house" audio/video communications network for real-time CAS intervention at the Royal National ETN Hospital, which allows parallel broadcasting to remote locations.

The process, based on Gould's methodology, is divided into four cycles: During the first design cycle (*Focus on User*) the designer focuses on potential users and their needs during microsurgery by means of in-situ video observations. In the second design cycle (*Early and Continuous Testing*), an original design for the proposed interface is established, followed by objective evaluation of a user task. During the third cycle of the design process (*Interface Optimisation*), the designer applies iterative design techniques in order to optimise the interface's effectiveness, and carries out a subjective evaluation of stereoscopic viewing. In the last design cycle (*Integrated Design*), a computer network for integrated design of the interface and other CAESAR components

[1] Courtesy of Sharp Laboratories Europe for participating in EPSRC grant award "A NEW AUGMENTED REALITY 3D DISPLAY SYSTEM FOR IMAGE GUIDED SURGERY".

in the operating theatre is set up, based on successful I/O control of the CAS system's broadcasting and communicative signal parameters.

1.5 Thesis layout

Following the introductory material in this chapter, Chapter 2 presents a philosophical review in the theory of light, the course of optical technologies and surgery in time, and the concept of mixed reality environments. Chapter 3 describes the basic properties of stereo vision and 3D visual systems. A literature review of CAS systems, including their advantages and disadvantages, is presented in Chapter 4.

Chapters 5, 6, 7, and 8 discuss the author's experimental work. In Chapter 5, physical and behavioural aspect models of ENT surgeons are derived from surgical task analysis of conventional microsurgery. Chapter 6 exhibits the design characteristics of the interface, and results from a quantitative analysis of the interface's accuracy and precision. Chapter 7 presents a field experiment using the prototype, followed by laboratory-based psychophysical experiments in stereo vision and a qualitative analysis of viewing comfort and depth appreciation. In Chapter 8, the author addresses the broadcasting demands of the interface and explains the design and implementation of a dedicated fibre optics network for real-time stereo imaging and CAS microsurgery.

Chapter 9 summarises major contributions of this research and presents areas of further exploitation.

CHAPTER 2

PRELIMINARIES

Γνώση : Knowledge

As stated by Aristotle, the meaning of '*knowledge*' needs special attention. The philosopher never used the word itself. Instead, he used the word '*γιγνώσκειν*' which in English would be interpreted as '*the ability to know*'. He used '*εiδεναι*', a word that finds its root in the noun '*εiδος*' which in turn means '*form*'. Both the two latter Greek words are derived from the root of the verb '*ιδεiν*', which means '*to see*'. In other words, Aristotle thought of vision as perhaps the most important sense in acquiring '*knowledge*'.

In the first sentence of his book '*Metaphysics*', the philosopher attempts to give a picture of a man in words. We are to see him trying to see, as if he is reaching out for a light which would make everything clear. According to Aristotle that is what man is by nature. The more care he uses in his seeing the clearer things, and the distinctions between them, become. The spectator develops his theory based upon his visual experience; and to have theory is the consummation of human life.

So, vision is the consummating sense. Without it, nothing would be seen. Thus, it is the visible world which is really primary in the matter of theory. It is the consummation of perception because *whatever other than visible things is perceived, is in the visible world attached in some way to what is visible.*

2.1 A treatise on light

To see, we need light! The statement alone raises some basic questions. What is light? Where does it come from? How does it produce an image? What happens inside the eye? To answer to these questions, philosophers developed theories about the nature of light.

The oldest theory of light was proposed by Empedocles (450 B.C.) and taken on by Plato. According to them, the eye projects forward a narrow visual ray that feels the object's radiation and returns through the pupil into the sensitive part of the eye, where it creates an image in the mind. The straight visual ray is considered as a long finger projecting from the eye, and sight is a kind of touch[2]. The theory was supported mathematically by Euclid who, through the geometrical principles of straight lines, suggested that a finite number of rays that leave the vibrating eye, sweep out the surface of an object and reflect back to the eye an *eidolon* (model) of the surface. Euclid also derived the law of reflection, which states that when a visual ray strikes a polished surface at an incident angle, the angle of reflection is equal to the angle of incidence.

The Platonic theory was not debated until the early years of 1000 A.D. when Alhazen (965-1038), an Arab philosopher, proposed after empirical observations that light is an accidental quality of an object that is illuminated by a source. When this happens, the object becomes luminous and is itself a source. The new source of illumination spreads outward in every direction and in this way is propagated from point to point until it reaches an opaque object.

The development of astronomy and in particular, astronomical observations led to new theories about the origin and motion of light. Scientific findings were mainly derived from observation and supported by mathematical relations. The new theories on cosmology influenced astronomers and mathematicians to study the properties of light. One of them, the great astronomer Galileo (1564-1642) was already convinced that the Sun is the centre of the world and had extensively studied the planetary motions through his telescopes. At the same time, Johannes Kepler (1571-1630), was the first mathematician who found that these physical motions are governed by mathematical laws. On the motion of light, Kepler proposed that light flows along straight rays in an instant. The rays themselves do not move and are not light. Rather, it is a surface perpendicular to a ray that moves and is considered as light. A generation later, the Dutch mathematician Christiaan Huygens (1629-1695) explained that this surface is a wave which travels in three dimensions.

[2] In ancient times, touch was considered as the most reliable sense as it was easy to understand. Any association of vision with touch allowed people to realise the concept of visual rays.

The birth of experimental philosophy originated new ways of studying the nature of light, mainly through its properties. An example of this comes from Isaac Newton (1642-1727) who, through his Theory of Colour, explained that light consists of pulsating particles that move through space in straight lines. When these particles encounter a reflecting or refracting surface, they set up vibrations in the particles that compose it. The vibrations propagate in the aether, resulting to a final effect that we sense as white colour. Nevertheless, if the vibrations could be separated, the longest one would be sensed as red colour and the shortest as violet. The latter observation suggested that there is secondary motion that follows a wave pattern when light travels in aether (shown in Figure 3.5).

The aim of mathematical physics is to mechanise the universe, to explain natural phenomena by means of a logic that follows the laws of mathematics. The debate between the Newtonian and Keplerian approaches to the motion of light reached its peak with the arrival of Dr. Thomas Young (1773-1829). He, like Huygens, proposed that light should be considered as matter in aether in exactly the same way that sound is matter in air. In 1801, Young gave a lecture at the Royal Society in London, where he explained that light is a travelling wave that follows a fundamental concept, which he called the *principle of interference*[3]. Young's greatest contribution to the nature of light is found in his studies of colour. He realised that the explanation of colour perception is not found in the nature of light but in the physiological characteristics of the human eye (as described in Chapter 3). Young experimented with mixing colours and found that all colours consist of three primary colours –red, green and blue. Hermann von Helmholtz (1821-1894) studied his theory further, deriving what today is known as the *tristimulus model of colour vision.*

The study of electricity and magnetism in the 19th century offered new ground for research to the quality of light. James Clerk Maxwell (1831-1879) expressed the empirical relationship between electromagnetism and light in a few mathematical formulae from which he derived a formula for the speed of light inside a medium. He continued by declaring that light is an electromagnetic disturbance in the form of waves

[3] The principle of interference states that when either two undulations, from different origins, coincide perfectly or very nearly in direction, their joint effect is a combination of the motions belonging to each.

propagated in the aether. Still, the new properties of the aether remained unclear and in contradiction with other theories. The problem was to be solved 25 years later, by Albert Einstein (1879-1955). Einstein rejected the aether hypothesis through his theory of relativity, and later showed that light is always propagated in empty space with a definite speed *c* that is independent of the state of motion of the emitting body.

2.2 Optical technology

The use of optical instruments began in ancient Greece during warfare activities. The study of Euclidian geometry had already uncovered the law of light reflection, and enlistment of mirrors as defence devices can be found in many historical battles. Later, the Greek-Egyptian philosopher Claudius Ptolemaeus (ca. 100-170), made a first scientific approach on the law of refraction. He found through an experiment that when a light ray travels inside water, it bends. He tabulated results of refraction for various incident angles. The phenomenon was not studied further until the beginning of the 13^{th} century, when Robert Grosseteste (ca. 1168-1253) studied how a lens brings light to a focus: from rays refracting at each surface of the lens. The latest theory triggered the production of spectacles.

2.2.1 Telescopes

The first clear lenses were manufactured in Italy, late in the 16^{th} century. Optical experimentation using a combination of convex and concave lenses led to the invention of the first telescope in September 1608 by Hans Lippershey (1587-1619), a Dutch spectacle maker. A foot-long tube, with a concave eyepiece at one end and a weaker convex lens at the other, could make distant things seem closer.

Galileo made the first telescopes for observing planetary motions. He managed to manufacture an objective lens with a magnification power of 30×, which allowed him to observe the texture of the moon's surface. Soon, other scientists started making their own telescopes. However, all of them suffered from three effects: low image brightness, spherical chromatic aberrations, and small magnification power. In search of a solution, researchers designed two types of telescopes: one that used mirrors (**reflecting telescope**) to guide more light to the object, and another that combined a number of large, masked lenses (**refracting telescope**) to increase the field of view. While the former telescopes were suffering from spherical aberrations, the latter suffered from low magnification power. With the development of silver-coated glass mirrors, the production of reflecting telescopes surpassed that of refracting ones. Nowadays, reflecting telescopes are equipped with a thin figured transparent plate mounted at the front end that acts as a refractor and allows the reflector to give a sharp image over the entire field of view.

2.2.2 Microscopes

It is believed that the first microscope was created by Galileo in an attempt to study insects. Nevertheless, the invention of the first compound[4] microscope was attributed to a Dutch spectacle maker, Zacharias Janssen (ca. 1595).

Originally, microscopes were used for the study of insects and bacteria. New, improved instruments were manufactured, offering better image brightness, higher magnification and greater resolution. Anthony van Leeuwenhoek (1632-1723), a Dutch master of the art, made a single lens microscope that had a magnification power of 250×. His instruments were very simple, made out of a single lens from glass or quartz crystals. The lens itself was very small and curved, very finely ground and polished. Unfortunately, the small size of the lens made long viewing tiresome and uncomfortable. Moreover, like telescopes, the quality of the new instrument was limited due to the physical properties of the single lens.

The first drawings of the microscopic world were drafted by Robert Hooke (1635-1703), and presented in his book *Micrographia* (see Figure 2.1). His microscope was based on the Galilean model and consisted of two lenses: a convex front lens formed a magnified image of the object and a second concave, called the eyepiece, magnified the image further. The microscope is illustrated in Figure 2.2.

[4] A system that is composed of two or more lenses.

Figure 2.1: Magnified image of mould growing on the leaf of a rose (Micrographia, 1667).

Figure 2.2: Hooke's microscope.

The instrument began to take its traditional shape at the beginning of the 19^{th} century with the contribution of Carl Zeiss (1816-1888). The German engineer opened the first microscope production workshop in 1846 with the aim of producing high-precision microscope for dissections. Soon, he realised that no matter how precisely his microscopes were made; the only way to make them better would be to establish a computational rather than experimental design for his instruments. He went on to collaborate with Ernst Abbe (1840-1905), a mathematician and physicist at the university of Jena. A few years later, Abbe developed his wave theory for microscopic imaging and for the first time microscopes were assembled according to theoretical considerations. In 1886, Otto Schott (1851-1935), a glass chemist and new member of the Zeiss-Abbe partnership, produced a range of new objective lenses that were free from chromatic aberrations[5].

[5] This type of lens is called *apochromat* as it produces images with no colour distortions.

The full potential of Abbe's mathematical designs was not reached until August Koehler (1866-1948) devised a new illumination method for microphotography in 1893. Three years later, Horatio S. Greenough designed a light microscope, consisting of two objectives and two oculars, which allowed the perception of depth in a microscopic image (see Figure 2.3). An alternative microscope design for depth effects, consisting of a single objective lens and two oculars (CMO), was introduced early in the 20th century. Both the Greenough and the CMO microscope designs are explained further in Chapter 3.

Figure 2.3: Early 19th century, Greenough-type, microscope.

Abbe's designs and Koehler's illumination method form the basis of light microscope development up to this date. Still, the theory of optical instruments has proved that there is a limit to the power of light microscopes. Optical studies have shown that no light microscope can clearly visualise any detail smaller than 200 nm. However, the applications of light microscopy have become almost endless, with microscopes commonly found in all kind of laboratories, industrial sites and hospitals.

2.3 Surgery through time

The practice of surgery dates back to the late days of the Stone Age. Archaeological excavations of skulls and surgical tools in various regions of the planet have confirmed surgical treatment of the head and the lungs. The most common treatment for head injuries was trepanation, where a drill was used to open a hole in the skull. The technique was used in order to relieve the patient from hydrocephaly, severe headaches or, in some supernatural way, bad spirits! Similarly, respiratory diseases were treated by opening a hole in the lungs of the patient and let the *pleuma* pour out of the body. Amazingly enough, some patients survived the procedure and lived on!

In most ancient cultures, surgeons were usually priests who enjoyed special status within the society. In ancient Greece, surgery became a profession, with young candidates from all ends of the state attending the Hippocratic School on the island of Kos. The Romans, who used to assign physicians to their military units, further increased the role of qualified surgeons.

Naturally, the success rate of surgical procedures was limited by the poor knowledge of human anatomy. Nevertheless, improvements came during the Renaissance period. Andreas Vesalius (1514-1564) was the figure that managed to change the course of medicine. Already a surgeon and anatomist, he published the first complete textbook of human anatomy, *De Humanis Corporis Fabrica* in 1543. The book, with illustrations from cadavers' dissections and new scientific terminologies, ended Galenian anatomy and inspired the birth of Modern Medicine.

During the 18^{th} and 19^{th} centuries, surgery moved back to the battlefields. The abundance of surgical scenarios improved surgical skills, but morbidity rates were still high. Pain, infection, haemorrhage and shock were the major obstacles to post-operation survival. However, some of these problems were overcome with advances in inflammation control, sterilisation and pre-surgical planning. The most notable contribution came from the development of anaesthesia in 1845 by William Morton.

The introduction of physical sciences and engineering in medicine came in the early years of the 20^{th} century. The anatomical atlas of the human body was completed through the x-ray images of human organs. Medical imaging allowed less invasive procedures, thus increasing the quality of surgery. With the addition of other medical and technological advances, such as histology and microscopy, surgery became a scientific rather than an experimental discipline. Classification was imminent; cardiothoracic, orthopaedic, head and neck surgery, all became specialities with their own characteristics and approaches. The ability to study the human anatomy and physiology at microscopic level inspired surgeons to perform more complex and delicate operations, particularly in Head and Neck surgery.

Vesalius's vision for respect to the human patient became the driving force behind surgical developments during the last century. The power of medical imaging has been a significant factor for that. Nowadays, new imaging modalities, such as Computerised Tomography (CT) and Magnetic Resonance Imaging (MRI), allow surgeons to plan, perform and teach their task better than ever before.

The dawn of the new century finds surgery facing new, great challenges. Computer graphics-generated images of the human body, new engineering models of human-machine interfaces, and the technological prospects of Artificial Intelligence (AI), now unveil a new era in surgery.

2.4 The inquiry of Reality

At the beginning of this chapter, the author presented Aristotle's view of nature through the eyes of a man. This man acts as a sensory receptor in a natural environment, interpreting it through his vision and feel. The collection of such information initiates a cognitive process from which '*reality*' is formed. Real objects are identified through their form, material and behaviour. They are comprised of tiny particles, called atoms[6], and have motion characteristics.

The influence of Christianity in the medieval period introduced a new essence of reality, where religious symbols construct a reality that is not based on the materialistic nature of things but on their superreal meanings.

The conflict between the Aristotelian and Clerical arguments of what is '*real*' ended in the early days of the Renaissance, when the concept of reality was re-examined by René Descartes (1596-1650). The French philosopher was the first to separate reality into mind and matter, observer and observed, subject and object. This bifurcation is called '*Cartesian dualism*'. The objective part of the duality is described as the external world; it occupies space and it is susceptible to being treated by mathematical sciences and geometry[7]. Other aspects, such as colour and feel, are subjective and belong to the

[6] Aristotle did not innovate the idea of atomic matter. He used the Democritean world model for which every body in the universe descends to molecules, which sequentially consist of smaller units, the atoms.

[7] Descartes invented a new branch in geometry that was later called co-ordinate geometry. He had the idea to measure the position of a point by its distance from two fixed lines, the Cartesian axes.

mental side of reality. Later on, Newton added that the external world is a classical order system, and the mind its epiphenomenon.

The Newtonian model of reality was expanded further during the 18^{th} century by Immanuel Kant (1724-1804). Kant separated reality into Forms of Sensibility and Understanding that follow three physical, *a priori* conditions, Space, Time and Order. These are imposed upon the external world by the nature of our senses so that any object that is contained in it can be located in space while events that are associated with it take time and occur in an ordered temporal sequence.

Quantum physics attributes reality to atomic matter that has internal energy and is related through non-locality and interconnectedness but in general, Space-Time reality as devised by Kant, has formed the foundations of how humans envisage and construct reality. What we appear to experience as real is a result of seeing in four dimensions. The model works out experimentally; hence, the construction of a reality is always carried out through geometrical (x-y-z space) and arithmetic (time) methods. However, as Kant added to Aristotle's metaphysical argument, our knowledge of reality is limited by our finite perceptions.

The application of mathematics to empirical experiment allows physical sciences to observe the external world through calculus. Mathematical models transform the observer from a contemplator to a calculator. Nevertheless, some calculations can be so complex and tiresome that the ability to understand and construct the external world is limited. The latter problem has been largely overcome with the invention of computers.

The computer constructs the external world based on human input; it receives data that comply with a pre-determined logic, performs algorithmic calculations of algebraic and geometric functions, and returns an artificial spatial map of the world onto the computer screen. The screen is the interface between the user and the machine. It acts as the boundary between the world that is perceived through human senses and the 'virtual' one that is constructed from human experience. The concept of *virtual environments* (VE), or *cyberspace*, was first introduced by William Gibson in his science fiction novel *Neuromancer* [12].

The construction of an environment that can be manipulated and modified through Human-Computer Interaction (HCI) has introduced a new form of reality, which is called **virtual reality (VR)**. However, since the advent of this new reality, many questions have been asked regarding its nature, and while some epistemologists consider the term as a human experience, others claim that VR is just a new technology.

In many textbook definitions, VR is defined as the use of computer technology to create the effect of an interactive three-dimensional (3D) world in which the objects have a sense of spatial presence [13]. For others, VR is considered as the stimulation of human perceptual experience to create an impression of something that is not really there [14]. Heim believes that VR consists of three interrelated components: *immersion*, *interactivity* and *information intensity* [15]. Immersion refers to the type of interface used for the human to experience the technology. Based upon this criterion, VR systems can be either **fully immersive** or **desktop-based**. The former use helmets and gloves while the latter use 3D displays and pointers (both systems are analytically explained in Chapter 3). Interactivity relates to the HCI features of the system, the real-time feedback that the user obtains from experiencing the synthetic environment. This includes visual, aural and haptic feedbacks. Information intensity usually refers to the amount of realism that the user enjoys when using the technology.

The realism of a VR system depends on the level of virtuality that the system produces. According to Milgram [16], between the extremes of real life and virtual reality lies the spectrum of mixed reality (see Figure 2.4). In mixed reality, views of the real world are linked with views of the virtual one. It can be divided in two categories, i.e. *augmented reality* (AR) and *augmented virtuality* (AV). In AR, the final view consists primarily of a real scene, with some computer graphics enhancements. In AV, the virtual environment is enhanced by added elements of the real environment.

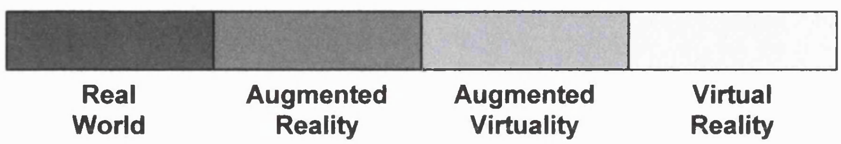


Figure 2.4: Mixed Reality Spectrum.

The aim of using mixed reality systems is to make the virtual elements in the scene less distinguishable from the real ones so that the viewer perceives a world that is familiar and makes sense to him/her [17]. Scientific arguments suggest that because of this, AV systems will fade out as computer graphics capabilities increase, and better interfaces are built.

CHAPTER 3

HUMAN VISION AND VISUAL DISPLAYS

The perception and cognition of the external world are comprehended by means of feedback from the five major human senses: vision, audition, olfaction, touch and gustation. Each of the body organs responsible for one of the senses reacts to a specific stimulus from the environment. Stimuli are processed in parallel, using different percentages of the brain bandwidth (see Figure 3.1). The combination of some sensory stimuli generates secondary senses. For example, visual and audible feedbacks generate the sense of balance, which gives information about the position and movement of objects within the environment.

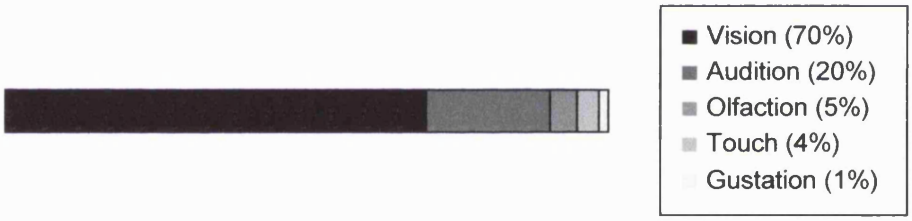


Figure 3.1: Primary human sensory system bandwidth.

Vision is the primary sense in human perception. The brain's interpretation of visual images signalled in the eyes forms the basis of a '*real*' experience, so that any attempt to reconstruct the external world carries the characteristics of the visual feedback obtained by the observer during the experience. The cognitive process of this psychological state of the brain is very complex and not discussed here as it exceeds the scope of this research. This chapter refers to the properties of human vision, with more significance attributed to **stereo vision** and the perception of **depth** in stereoscopic images. Furthermore, it describes the design and development of **stereo microscopes** and concludes with the discussion of advanced **3D visual displays**. The latter have been the subject of intense research for many academic and industrial groups since they constitute a major part of presenting 3D VR models.

3.1 The human eye

The behaviour of the eye as an optical instrument was first established by Descartes who also explained how its mechanisms obey the laws of physics and optics. In general, the characteristics of optical systems are determined from their structure and the physical theory that is used in forming a visual image. In the case of the human eye, the first is defined from the anatomy and physiology of the organ while the second is based on the lens theory of geometrical optics.

3.1.1 Anatomy and physiology

The human eyes are protected by a pair of bony cavities in the skull, called the orbits. Each eye is supported by six extraocular muscles which allow a range of movement of 50° to either the left or right of the resting position and 40° above and 60° below the straight ahead position. The first optical surface of the eye is the cornea; a transparent bulge with no blood supply that obtains its nutrients from the aqueous humour that fills the anterior chamber lying behind it (see Figure 3.2). The latter leads onto the iris, an extremely delicate membrane with a circular opening at the centre, which is called the pupil. The iris dilates or contracts, depending on the amount of light entering the pupil. Behind the iris lies the crystalline lens, which adjusts the focal length of the eye to bring images into focus on the back face of the eyeball, the retina. The lens and retina are separated by a large cavity called the posterior chamber, which is filled with vitreous humour. The latter acts as heat absorbent. The retina is a photosensitive assembly of nerve cells, blood vessels and connective tissue that converts electromagnetic radiation into nerve impulses[8].

[8] The retina comprises of ten layers, as shown in Figure 3.3b. The second layer close to the posterior chamber contains the optic nerve fibre, where the spatio-electrical impulses are generated. In-between layers convert electromagnetic radiation to action potentials. The ninth layer is the photosensitive one while the last layer is the pigment layer.

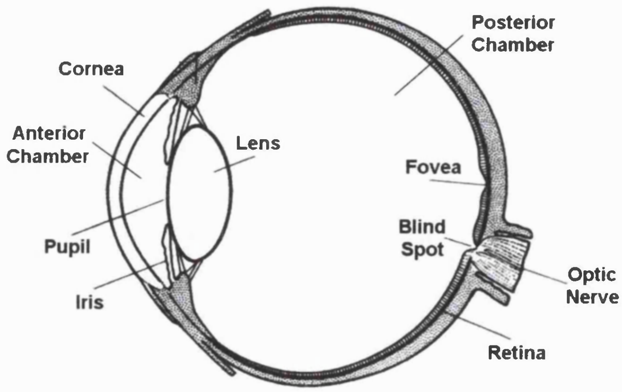


Figure 3.2: Human eye anatomy.

Light rays are initially refracted at the convex corneal surface. The aqueous humour has almost identical refractive index with the cornea; hence, the rays do not bend much when inside the anterior chamber (Figure 3.3a). When there is excess illumination in the chamber (*photopic vision*) the iris contracts. The opposite effect occurs when insufficient amount of light (*scotopic vision*) enters the pupil. The rays are further converged by the crystalline lens before an inverted image of the external world is formed onto the photosensitive layer of the retina. The layer is packed with two groups of photosensitive receptors, rods and cones. Rods are more sensitive to light than cones, but the latter can distinguish different wavelengths of light. Rods are unevenly distributed in the retina. They are absent from the fovea, where the main concentration of cones is found.

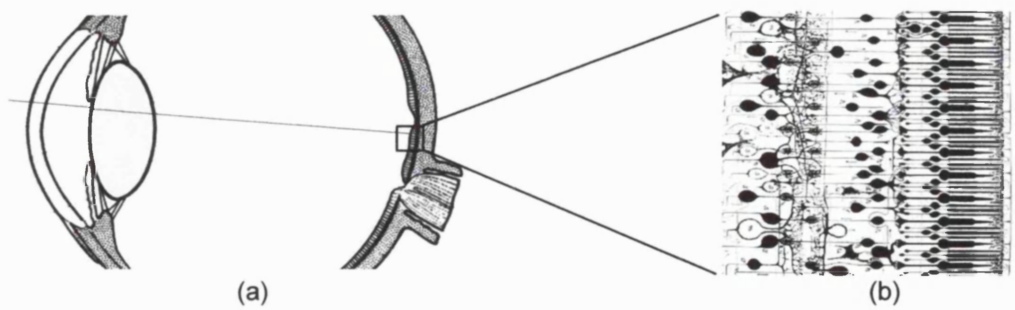


Figure 3.3: The human eye: (a) a positive, double lens arrangement that casts a real image on a light-sensitive surface, (b) schematic presentation of the retina layers in a region near the fovea.

3.1.2 The visual image

The creation of a visual image in the eye depends on the physiological characteristics mentioned above. Incoming rays are affected by two specific mechanisms in the iris and lens: **adaptation** and **accommodation**. Adaptation involves the mechanical opening and closing of the iris that controls the intensity of light entering the pupil. Accommodation is the process by which the lens changes its curvature in order to bring a laminated distant object into focus at the retina. At rest[9], the **objective focal length** of the lens is six metres (see Figure 3.4).

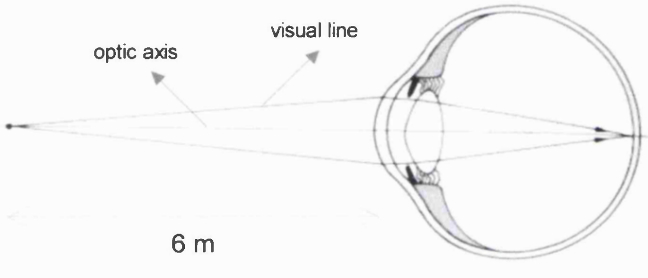


Figure 3.4: Objective focal length of the resting human eye.

The photo-sensing effect has three perceptual characteristics: *brightness*, *hue* and *saturation*. The perception of brightness is directly proportional to the intensity of light entering the pupil. However, rays with the same intensity but with different wavelength do not always have the same brightness. Hue is the attribute that distinguishes the different colours of light. It is described by the **tristimulus model of colour vision**, in which there are three types of photoreceptors, red, green and blue (**RGB**), with differing spectral sensitivities (see Figure 3.5) [18]. Colours are normally described as mixtures of RGB wavelengths. When two rays with the same spectral composition are observed, they will appear to have the same hue. Nevertheless, it is possible for two or more rays with different spectral compositions to be perceived as having the same hue[10]. Saturation describes the "whiteness" of a ray. The less saturated is a colour; the more

[9] The value here assumes that the pupil opening is five mm. Pupil size varies from 2-8 mm, depending on the rays' intensity.

[10] Such colours are called metamers.

whitish it appears to be. The degree of saturation is also known as the colour *chrominance.*

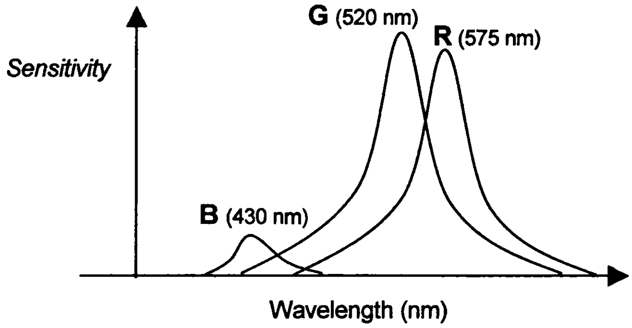


Figure 3.5: Human eye response to radiation of different wavelengths.

In photopic vision, visual sensitivity generates further perceptual characteristics: contrast, flicker, and resolution. The ability of the human eye to discriminate luminance differences in the adjoining fields of photoreceptors depends on the average intensity of the light rays, their size, and wavelength. This **contrast** discrimination is not linear; any noticeable difference is a function of the light intensity. Under normal photopic vision, contrast can be seen in 2% luminance differences, known as the Weber fraction. When the luminance of a scene fluctuates periodically several times in a second an annoying sensation is perceived, known as **flicker**. As the frequency of fluctuation increases, sequential scenes appear disjointed, as frames. The irritation reaches a maximum at around 10 frames per second, but fades out as the frequency rises further. The human eye can detect frames only within the boundaries of photopic vision. The ability to perceive two individually discernible lightwaves within a frame defines the eye's limit of spatial **resolution**.

The study of the human eye as an optical system is necessary in the development of television and video technologies. Knowledge of the physiology and psychophysiology helps in understanding the design parameters, as well as the limitations, of video.

3.2 Stereopsis

Stereopsis is the ability of the brain to extract a stereoscopic percept from two correlated retinal images. The resulting effect is a "solid" perception of the 3D structure of the world when seen by a pair of eyes. Each eye uses different images of the same object to form a solid view in the human visual system. The 3D view is achieved by using cues such as scenes hidden by opaque objects lying closer to the observer, foreshortening of distant objects by perspectives, shadows cast by oblique illumination, shading of the surface luminance, rotation of the object, and stereopsis.

3.2.1 Binocular vision

All animals with two frontal eyes have binocular vision. Individually, each eye possesses a monocular visual field that is defined as the solid angle within which an object is visible to the stationary eye. For humans, the monocular visual field of the stationary eye extends horizontally about 95° in the temporal direction and about 56° in the nasal direction [19]. The total visual field is the solid angle subtended at a point midway between the two eyes by all those points in space visible to either eye or both. It extends laterally about 190° when the eyes are stationary and about 290° if the eyes are allowed to move (see Figure 3.6). The portion of the total field within which an object must lie to be visible to both eyes, for a given position of the eyes, is known as the **binocular visual field** (θ). The binocular field is flanked by two monocular sectors within which objects are visible only to one eye. Each monocular sector extends about 37° laterally from the temporal boundary of the orbital ridge to the boundary of the binocular field at infinity. The left and right boundaries of the binocular visual field are formed by the physical position of the nose, and are about 114° when the eyes converge symmetrically.

On average, the human eyes are placed about 6.5 cm apart, hence each eye sees images from different viewpoints. As a result, the left eye sees more visual lines of the left side of a 3D object while the right eye sees more lines of the object's right side. These side-to-side differences in the positions of the two images generated in the eyes are called **horizontal disparities**. Differences in the up-down positions of similar stereoscopic images are known as **vertical disparities**. Both types of disparity and their effects in stereoscopic vision are discussed analytically in §3.2.3.

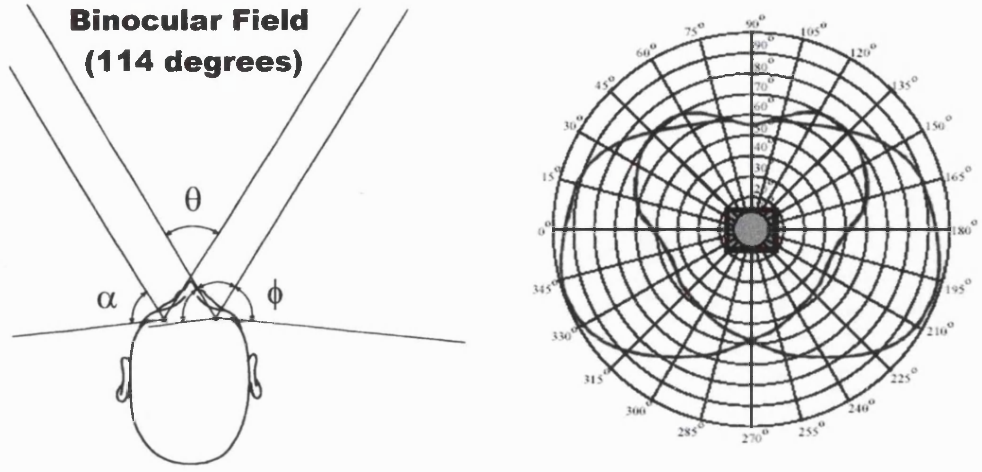


Figure 3.6: The binocular visual field and overlap region. Angles α and φ are the extents of the monocular sectors (contained inside the heavy black lines) at infinity.

For each image point that is formed in the binocular portion of the retina of one eye there is a corresponding point in the retina of the other eye. Image pairs that fall on corresponding points have zero binocular disparity, and produce the sensation of a single object when the two eyes verge on the object's corresponding points. However, disparate images can still produce the same sensation if the level of disparity does not exceed a specific range. The latter is called **Panum's fusional area**. Binocular images that fall outside this area are experienced as double. The resulting effect is double vision, otherwise called **diplopia** (see Figure 3.7). The maximum retinal disparity (also called image separation) for which the impression of a single fused image can be maintained is called diplopia threshold and it varies between subjects [20].

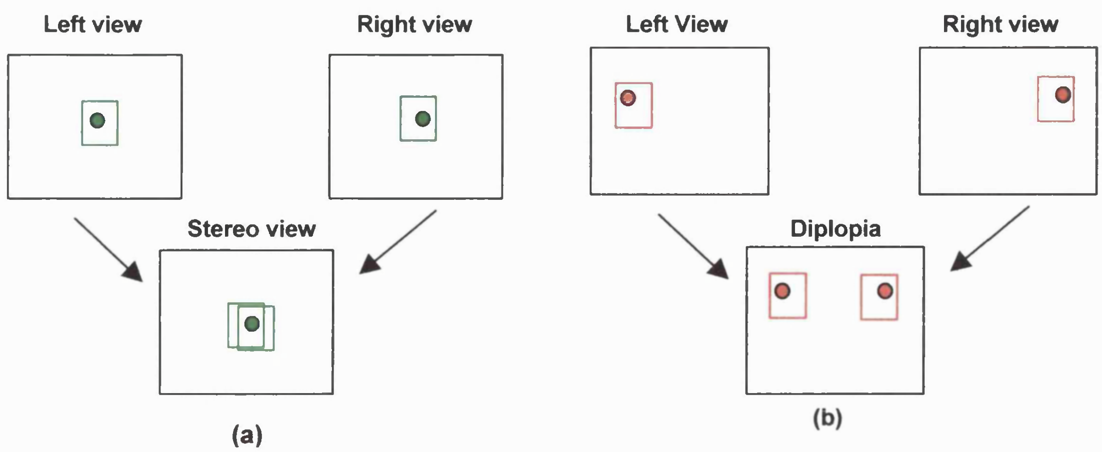


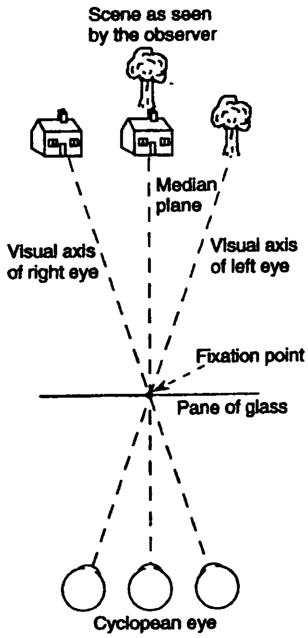
Figure 3.7: Object sensation (a) within and, (b) outside Panum's fusional area.

3.2.2 Depth perception - The cyclopean stimulus

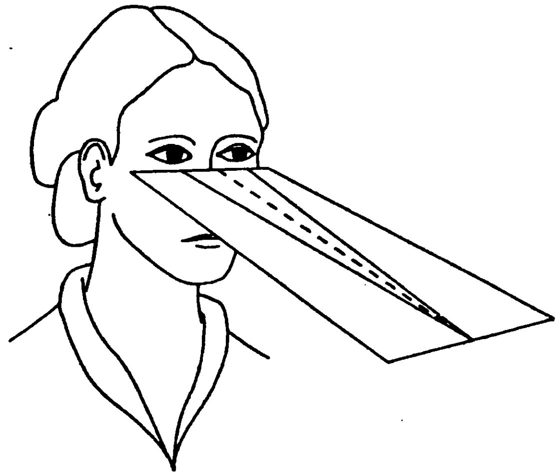
As mentioned earlier, visual stimuli are perceived by the photo-sensing receptors of the eye's retina. A binocular stimulus is one sensed by both retinas simultaneously or in rapid succession. There are two types of stimuli: *dioptic* and *dichoptic*. A dioptic stimulus is a single distal stimulus seen by both eyes. Although the two retinal images differ due to the anatomical position of the eyes, it is assumed that a dioptic stimulus produces identical monocular images. Dichoptic stimuli are distinct distal stimuli, presented separately to each eye. Such stimuli are normally created in a stereoscope (see Section 3.4) or by placing different optical devices such as prisms or filters before the eyes.

When two dichoptic images of the same object are fused by eye conversion or diversion, a binocular visual process is initiated in the brain's primary visual cortex. At this stage, visual processing becomes 'central' since any visual information from the combined retinal stimuli is not evident when the same stimuli are combined in one eye. This type of 'central' stimulation is known as **cyclopean stimulation**, and was first described by Julesz [21].

The concept of the cyclopean eye has been used largely by vision scientists to denote a geometric centre of reference for headcentric directional judgements in order to avoid registration of the eyes' angular position in the head [22]. When both eyes fixate on an object, they point at different directions with respect to the median plane of the head. However, the object appears to have a single direction in space. By using this egocentric term, any directional information from each eye can be referred to the cyclopean point shown in Figure 3.8.



(i)



(ii)

Figure 3.8: Geometry of the cyclopean coordinate system, (i) even though each eye views a different image, the cyclopean 'central' processing mechanism perceives the combined visual information, (ii) schematic diagram of the headcentric space.

The sensation of depth in a 3D image is produced from the disparate images seen by the two eyes. These occur due to the interpupillary distance that separates the two eyes. In general, binocular disparity is the term that refers to the differences between retinal images created by viewing the world from two slightly different vantage positions. Such differences can be specified and measured with respect to the images projected on the retinas or camera planes.

Any point within the binocular field of view of the two eyes can be calculated by the azimuth and elevation angles at each of the two vantage points. It is obvious that particular pairs of azimuth and elevation values will correspond to the same point in space if the two vantage points intersect. Subsequently, if the interpupillary separation of the two eyes is known, the absolute distance of a point from the cyclopean origin can be found (see Figure 3.9). If the elevation or azimuth value of that point from the two eyes differs, it is regarded as disparate. The difference in the azimuth is a measure of the point's **absolute horizontal disparity**, and the difference in its elevation is a measure of its **absolute vertical disparity**. However, if the cyclopean coordinate system is not defined in the occulocentric frame of reference, any patterns of difference describe only **relative disparities**.

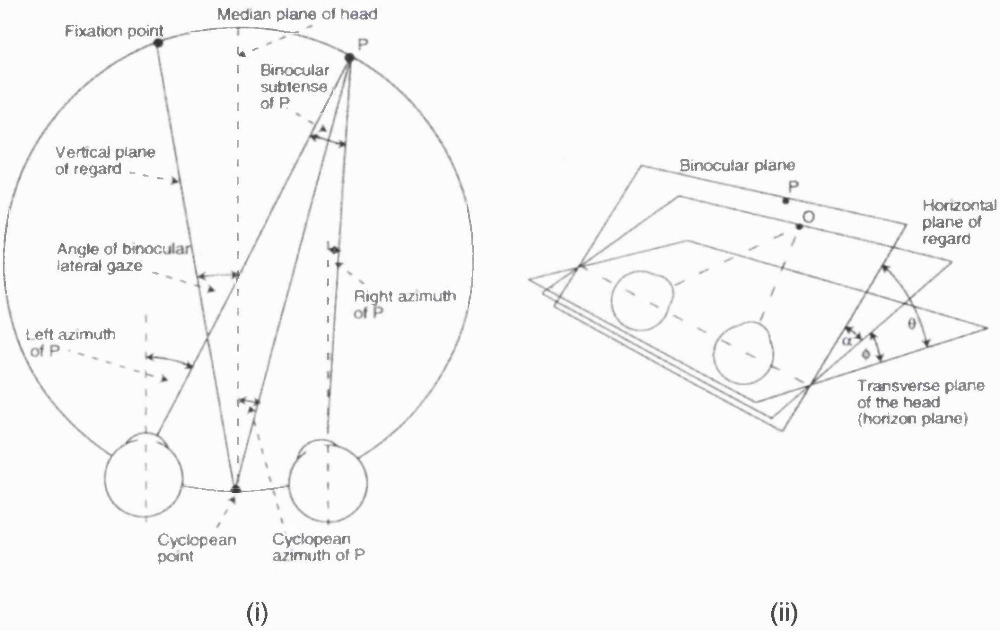
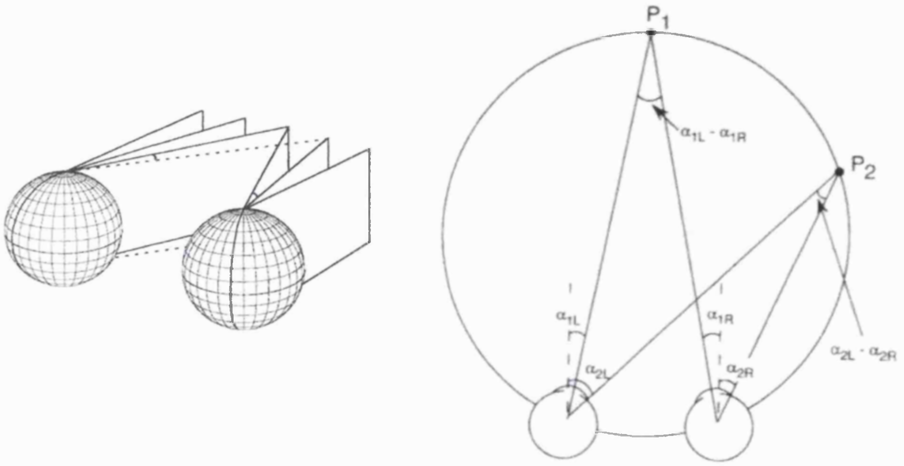


Figure 3.9:Cyclopean coordinates of absolute disparity: (i) the cyclopean azimuth of a point P is the dihedral angle between the median plane of the head and the vertical plane containing the point and the cyclopean point midway between the eyes, (ii) the cyclopean elevation of P in headcentric coordinates is the angle θ between the horizon plane and the horizontal plane of regard containing the point.

Similarly, disparity can be measured from differences in the direction of two or more points in space. This technique is preferred for stereo camera systems whose design aims to create an effective viewpoint that is identical to that of the observer's own eyes [23]. When the optic axes of the eyes converge to a single point in space, a plane of regard, passing through the point and the optical centres of the two eyes, is created (similar to the egocentric plane in Figure 3.8 (ii)). If the eyes are torsionally aligned, the y-axis of the coordinate frame in each eye will be orthogonal to the plane of regard and pass through its optic centre. Under this system, vertical azimuth planes in each of the eyes will radiate out from the y-axes (see Figure 3.10). The vertical plane projecting to the fixation point in each eye is the plane from which the azimuth angle in that eye will be measured. Any binocular difference of the separate angles between the azimuth planes containing the points in each eye is called **horizontal width disparity**. Any difference of the separate angles between the elevation angles is called **vertical height disparity**.



(i) (ii)

Figure 3.10: (i) Azimuth planes: all points which lie in a vertical plane have the same azimuth, (ii) The absolute azimuths of two points P_1 *and* P_2 *are specified with respect to aligned axes in the two eyes: the binocular subtense of each point corresponds to the difference between their absolute azimuths –* $(\alpha_{1L}-\alpha_{1R})$ *and* $(\alpha_{2L}-\alpha_{2R})$*.*

The images of a corresponding fixed point F that fall on the eyes' foveae have zero disparity. When the eyes converge onto a second point P, placed at a distance Δd beyond F, the image of P is ϕ_r degrees of visual angle to the left of the fovea in the right eye and ϕ_l degrees to the right of the fovea in the left eye. If P lies on the cyclopean axis, as shown in Figure 3.11(i), the two visual angles are equal ($|\phi_r|=|\phi_l|$). The circle that passes through the point of fixation and the nodal points of the two eyes is called the *horopter* (see Figure 3.11(ii)). The resulting images of P have **uncrossed disparity** since the visual lines of the two eyes intersect beyond the horopter. Respectively, for any point A that lies inside the horopter, the resulting images of A have **crossed disparity**.

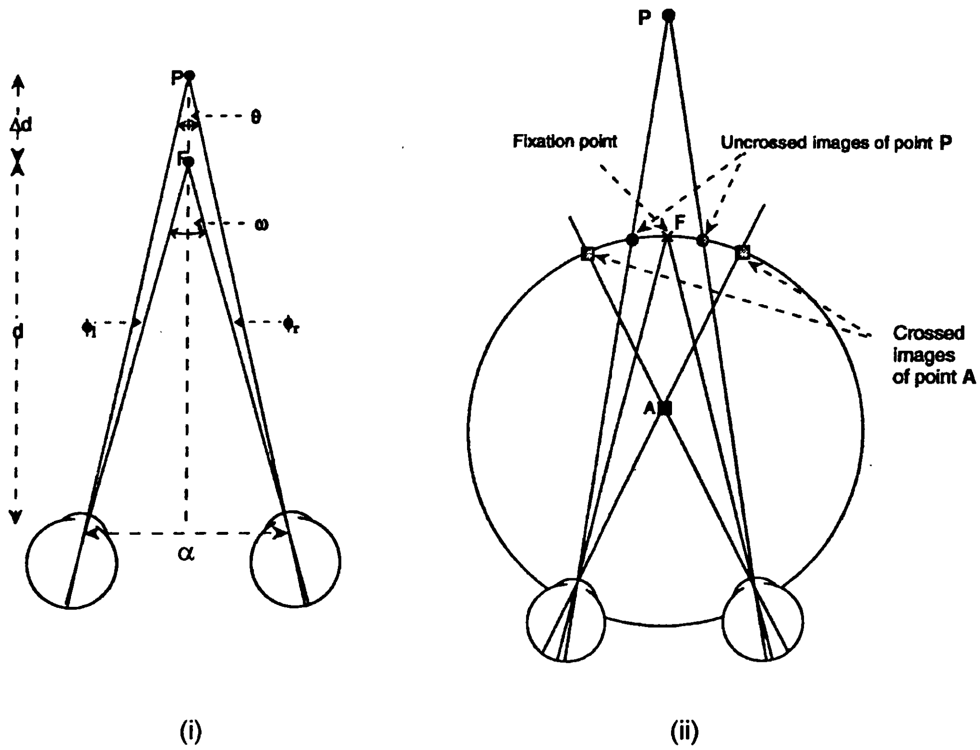


Figure 3.11: Geometric representation of binocular disparity: (i) When the two eyes, separated by distance α, converge to a point P outside the horopter, the apparent image has uncrossed disparity, (ii) When the eyes converge to a point A inside the horopter, the apparent image has crossed disparity.

By using basic axioms of Euclidean geometry in Figure 3.11(i), one can derive the following relationship between the depth percept Δd of a point and the disparity gradient d_F of that point with respect to a fixated point of the horopter:

$$\tan\phi = d_F = \frac{\alpha\Delta d}{d(d+\Delta d)} \qquad \textit{(Equation 3.1)}$$

The geometrical characteristics of binocular disparities form the basis of the computational theory of stereopsis. In the field of imaging, the cyclopean coordinate frame of reference is preferred for its simplicity.

3.2.3 Binocular imaging

Knowing the geometry of the cyclopean coordinate system is very important in machine vision, particularly when generating stereoscopic images for binocular imaging. Such images are normally captured using stereoscopic camera systems. The latter can be video-based or computer-generated, depending on the nature of the captured scene (real or virtual). Usually, a stereoscopic camera system consists of two cameras which capture separate monoscopic images of the same scene, one for each eye. By setting the two cameras' *interposition* (**camera separation**), *direction* and *field of view* (**FOV**) one can determine how the resulting 3D images will be displayed and viewed. Once the position and direction of the two cameras are known with respect to the cyclopean axis, the entire metric structure of the 3D world can be recovered [24].

Several groups, especially in the field of cinematography, have attempted to calculate and tabulate sets of the above parameters for specific visual effects, such as image width and perceived depth [25, 26]. The FOV of a stereoscopic image is mainly controlled by the properties of the telephoto lens used to capture the scene: its focal length (*f*) and magnification level (*M*). However, the direction and camera separation of the camera system is based on the chosen viewing geometry. It has been shown that converging (toed-in) cameras introduce vertical image disparities that destroy the stereoscopic percept [27]. Moreover, as shown in Figure 3.12, a parallel-axes stereoscopic camera system where the camera separation (*a*) matches the interpupillary distance (*p*) of the two eyes is preferred [28]. Still, in some 3D viewing situations it is not possible to impose the proposed restrictions.

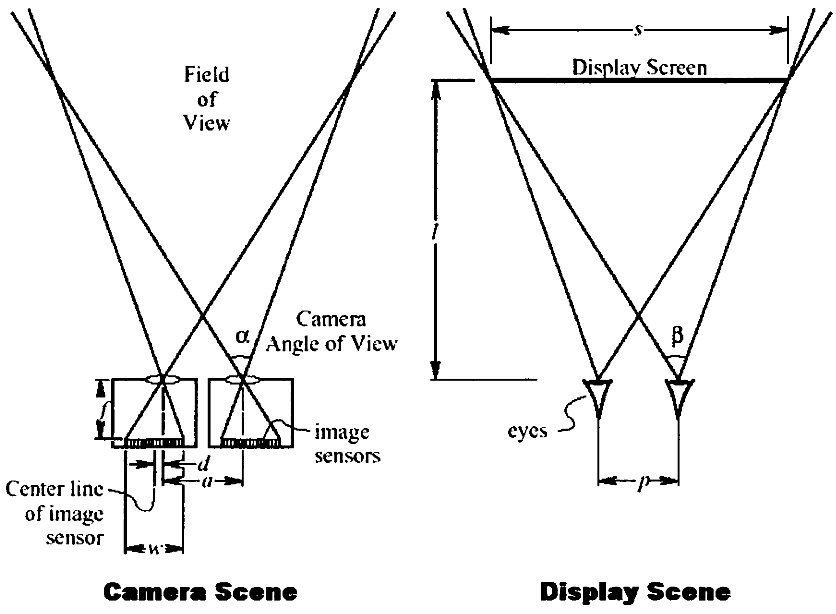


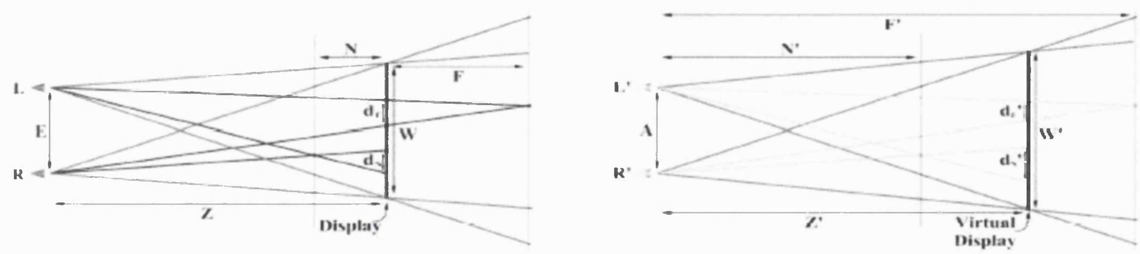
Figure 3.12: Geometry of a parallel-axes stereo camera system for an orthoscopic viewing frustum.

The geometry of the above condition defines a set of equalities:

$$a = p$$
$$\alpha = \beta \qquad \textit{(Equation 3.2)}$$
$$\frac{w}{s} = \frac{d}{p/2} = \frac{f}{l}$$

where the optical centre of each camera lens is placed at an offset position (*d*) from the cyclopean point (*w/2*) of the camera's photo-sensing element. The resulting stereoscopic image is projected on a flat, rectangular surface of size *s* for which the viewer stands at distance *l*.

The binocular imaging geometry described here can be applied only when the object size and depth match the target display size and depth range respectively. Any changes in the nature of the viewing medium will alter the relationships between the viewer, the scene and the camera separation. This need originates from the display parameters of the new viewing model. For example, in stereoscopic 3D displays, where visual information lays both in front and behind the screen plane, the physical position of the two cameras may be offset in order to ensure that corresponding points in the left and right eye are displayed on the screen plane and depth is perceived comfortably by the viewer (see Figure 3.13).



(i) (ii)

Figure 3.13: Geometry of a parallel-axes stereo camera system for a stereoscopic viewing frustum, (i) in the viewer-display space and (ii) in the camera-scene space.

For an observer who is positioned on the cyclopean axis of the display, a different set of equalities arises from the initial display parameters [29].

$$d_{N'} = \frac{A(Z'-N')}{N'}, \; d_{F'} = \frac{A(F'-Z')}{F'}$$

$$Z' = \frac{\frac{d_{N'}}{d_{F'}} + 1}{\frac{1}{N'} + \frac{d_{N'}}{d_{F'}F'}} \qquad \text{(Equation 3.3)}$$

$$W' = \frac{Z'W_f}{f} = \frac{W}{M}$$

where A is the camera separation, $d_{N'}$ and $d_{F'}$ are the disparities of objects N' and F' units away from the cameras, and Z' is the Zero-Disparity-Plane; the optical plane which will appear to be on the display plane when the images are viewed. In Chapter 7, the reader will see how the above set of equalities can be used in order to examine depth effects in video-based photographic images, captured from a surgical stereo microscope.

3.3 The stereo microscope

Microscopic stereoscopic vision is accomplished by means of two compound microscopes, one for each eye, placed at about the ordinary interpupillary angle of convergence for reading. The different angular view between the eyes leads to the stereoscopic effect. Normally, surgical stereo microscopes are designed so that the object in focus is at a fixed distance. For surgical microscopes used in otolaryngology the focal distance varies between 200 and 400 mm from the objective lens. This is due to the need for adequate working space for the surgeon's tools.

There are two kinds of stereo microscopes: the **Greenough** (binobjective-binocular) type, and the **Common Main Objective** (CMO, monobjective-binocular) type. The Greenough type consists of two objectives, mounted at a slight angle to each other on a single stand, very much like a set of binoculars that seamen use. This small angle, as seen in Figure 3.14(A), corresponds to the matching of the binocular parallax and is responsible for the creation of depth. Still, the system is not very adequate as the intermediate images are inclined from the plane of the working organ and tilted toward each other so that only their central portions are under the same focus conditions[11].

In surgery, the most appropriate system to use is the CMO microscope, shown in Figure 3.14(B). The large sensing area of the objective lens allows the eyepieces to receive two different views of the working region, thus producing stereo vision effects. The main advantage of this system is that the intermediate beams run parallel to the plane of the object and without tilting toward each other. Hence, the two images are extracted simultaneously, and with same focus. In this case, the effect is achieved through the minor disparity between the two monoscopic images. When the disparity is increased, depth is exaggerated. If the disparity between the two images exceeds a value for which the two eyes capture quite different images, depth is lost and diplopia occurs.

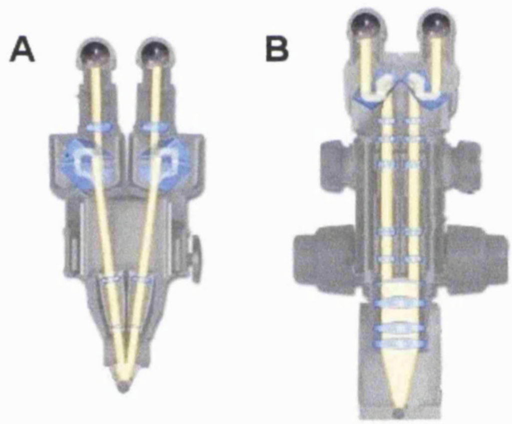


Figure 3.14: Schematic beam path diagram of the Greenough (A) and Common Main Objective (B) stereo microscope systems.

The optical components, coloured in blue in Figure 3.14, are *erecting prisms*. Such instruments were originally engineered by Carl Zeiss in order to produce an erect image of the object under the microscope so that the viewer can see the object, as this would appear without the use of the microscope.

[11] The term discussed here is known as the Keystone effect. When occurring, the peripheral portions of the viewing field are focused either slightly above or below the actual object plane and have very small differences in magnification.

3.3.1 Stereo microscope design

Early stereo microscopes were mainly used for surgical dissections and a few other industrial applications involving small assemblies and miniature components. They were heavy, made from brass, utilized prisms for image erection, and had simple lens systems consisting of one or two lenses. The first modern stereo microscope was manufactured by the American Optical Society in 1957 [30]. Its name, the **Cycloptic**, implies a correlation between the viewing conditions of the instrument and the cyclopean model of vision. Indeed, the design (see in Figure 3.15) is based on the refracting action of a single, large diameter objective lens, through which both the left and right channels view the object. Each channel operates as an independent optical train parallel to the other, and there is collimated light between the individual channels and the objective, so that images are projected to infinity. This arrangement guarantees that with little convergence, the left and right optical axes coincide with the focal point in the specimen plane. Because this parallel axis arrangement is usually extended to include the eyepieces, the left and right images are viewed by the eyes with little or no convergence.

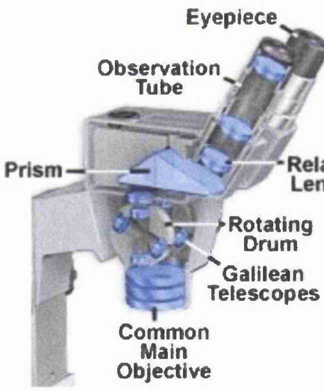


Figure 3.15: Schematic diagram of the Cycloptic® stereo microscope. The CMO lens found at the bottom of the body acts as a cyclopean eye; refracting light rays and projecting parallel images towards the oculars.

A rotating drum placed above the objective lens acts as an internal magnification changer, allowing the observer to increase the objective magnification in discrete steps. The drum is equipped with four pairs of Galilean-style telescopes, with each pair's lenses separated 25 mm apart. Galilean lens systems have the advantage of offering small focal length and very small field diameter while achieving high magnifications.

As the drum rotates, the telescope lenses are used in both forward and reversed orientations (magnifying and minifying), to yield different levels of magnification. Tube lenses are used to sustain the cyclopic effect. At the viewing end, the microscope utilises one-piece glass erecting prisms to produce the 3D image. In stereo microscopy, erect images are of vital importance because they maintain natural perception of the stereoscopic effect.

Images are viewed through the *eyepiece*. The latter is equipped with a diopter adjustment which focuses the two eyes simultaneously, thus enabling the operator to vary the interpupillary distance between the eyepieces over a range of 55 to 75 millimetres. This adjustment is often accomplished by rotating the prism bodies with respect to their optical axes. Since the objectives are fixed in their relationship to the prisms, the adjustment does not alter the stereoscopic effect. Modern stereo microscopes are equipped with standardized widefield eyepieces that are available in magnifications ranging from 5× to 30×, in 5× steps. The overall magnification is the product of the objective and eyepiece magnifications, plus that introduced by other intermediate lenses. Unfortunately, several practitioners often omit the eyepiece magnification when determining the instrument's FOV and overall magnification.

The microscope's FOV, which is the visible and focused area of an object under the instrument's objective lens, is determined by the objective lens magnification and the size of the fixed field diaphragm in the eyepiece. When the magnification is increased, the FOV is decreased if the eyepiece diaphragm diameter is held constant. Conversely, when the magnification is decreased, the FOV is increased if the diaphragm diameter remains fixed. The latter, also known as the *field number* (FN), is measured in millimetres and is expressed by the following ratio:

$$\mathrm{FN} = \frac{\mathrm{FOV}}{M} \qquad \textit{(Equation 3.4)}$$

Contrary to the FOV, the instrument's resolution depends on the degree of illuminating wavelength and the *numerical aperture* (NA) of the objective lens only[12].

[12] The numerical aperture is defined as one-half the angular aperture of the lens multiplied by the refractive index n of the imaging medium, which is usually air.

By dividing the illumination wavelength by the NA, the smallest discernible distance between two points is given by the Equation 3.5, known as the Raleigh Criterion:

$$R = \frac{0.61 \times \lambda}{n \times \sin\theta} \qquad \textit{(Equation 3.5)}$$

in which R is the smallest resolvable distance, λ is the illuminating wavelength, n is the refractive index of the medium between the objective and specimen, and θ is the objective one-half angular aperture.

The resolution mentioned here is considered as the instrument's lateral resolution. Another important aspect to resolving power is the axial (or longitudinal) resolution of an objective, which is measured parallel to the optical axis and is often referred to as *depth of field* (DOF). It is determined by the distance from the nearest object plane in focus to that of the farthest plane also simultaneously in focus, and depends on the NA and magnification of the objective lens. The above is mathematically shown in the following equation:

$$\mathrm{DOF} = \frac{\lambda_o n}{\mathrm{NA}^2} + \frac{n}{M\mathrm{NA}} e \qquad \textit{(Equation 3.6)}$$

in which λ_o is the wavelength of light, n is the refractive index of the lens, and e is the smallest distance that can be resolved by a detector that is placed in the image plane of the microscope objective, whose magnification is M and aperture diameter NA.

3.3.2 Video microscopy

The development of new industrial applications, such as video technologies, has significantly affected the design of the latest microscope systems. For example, the need for documentation of parts of the surgery brought forward the idea of splitting the light beam that comes out of the objective lens of the microscope into two identical beams. The first beam then follows its original optical path and enters the eyepiece, while the other is directed towards a camera, positioned at the focal distance from the *beam splitter* (see Figure 3.16).

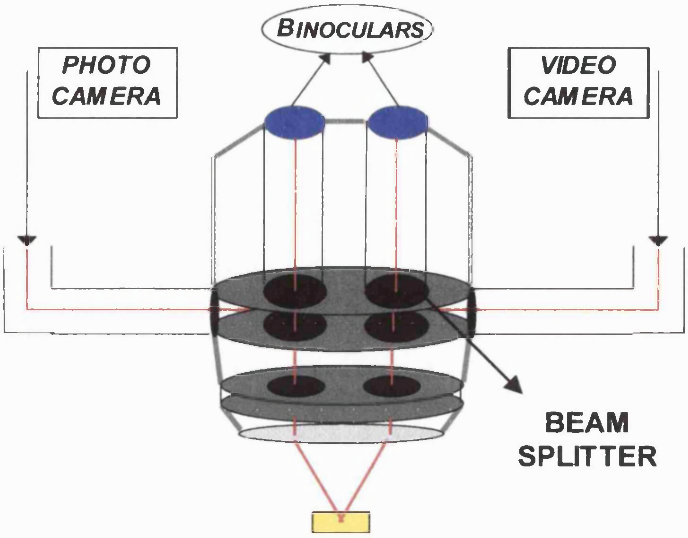


Figure 3.16: Graphical representation of the optical ray path inside the beam splitter.

Quite recently, photo cameras were replaced by video cameras, through which the optical image observed through the microscope can be displayed on a 2D monitor or even recorded on a standard video cassette recorder (VCR). In this way, other staff inside the operating theatre (e.g. consultants, medical students, nurses, etc.) can watch the procedure as though looking through the microscope itself (Figure 3.17). This latest achievement introduced a new field in microscopy, called video microscopy.

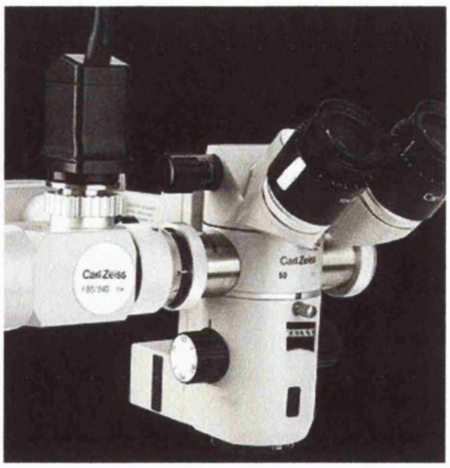


Figure 3.17: One of the latest surgical stereo microscope as it is manufactured by Zeiss.

One of the latest breakthroughs affecting the field of video microscopy is the production of charge-coupled device (CCD) video cameras. The camera is composed of a large array of photodiodes deposited on a silicon substrate. Each sensor element is a silicon photodiode built into the silicon chip but isolated electrically from its neighbours by a channel stop. The interaction of light with the photodiodes leads to the production of electrons, which in turn are attracted to a potential well, called the shift register. In the most common CCD design (three-sensor), each photodiode (pixel) is subdivided into thirds by three potential wells oriented as horizontal rows. The potential well storage capacity determines the dynamic range of a CCD video camera. Modern CCD cameras come in a range of sizes designated by the "inch" notation. The most common ones are the 1/3, 1/2 2/3 and 1 inch, shown in Figure 3.18.

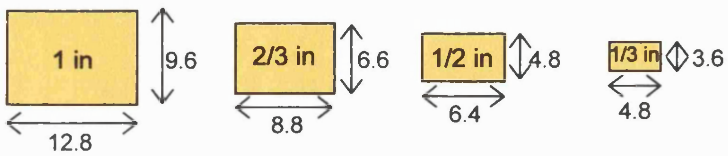


Figure 3.18: Charge-Couple Device sensor formats (all dimensions in mm).

As mentioned in the previous paragraph, the FOV of the microscope is determined by the projected image from the eyepiece, but this is generally larger than the CCD sensor size. For example, for a 1/2 inch CCD the diagonal dimension of the sensor (8 mm) covers only 26% of the image for an 18 mm FOV. This means that the final display on the monitor shows only a proportion of the total FOV, but offers a highly magnified video image. In some cases, the above result is considered as a disadvantage, since it obstructs imaging. The FOV reduces further, when the optical magnification component of the microscope is increased. Still, in cases where a magnified image of the region is required, the use of CCD video cameras becomes essential [31]. The physical size of the image is determined by Equation 3.7:

$$\text{image size}\,(S) = R \times M \qquad \textit{(Equation 3.7)}$$

in which R is the microscope's optical lateral resolution and M is the objective magnification.

Colour is an important tool for good visualisation of images. Hence, colour rather than monochrome video cameras became a necessity. Based on the physical characteristics of the human eye, colour video cameras are manufactured in such way that light of the three primary colours (**R**ed, **G**reen and **B**lue) superimposed in various ratios gives rise to the sensation of colours [32]. The acquisition of both RGB and luminance information from an image is achieved by either three separate CCD sensors or a single sensor filtered and masked to acquire the three simultaneous images. In the three-sensor camera, the incoming optical image is split into three images with a set of beam splitting prisms or mirrors that produce the red, green and blue images (see Figure 3.19). The three images are focused onto the sensing area of the three CCDs at the same time[13]. The resulting effect is the production of separate but synchronised, analogue R, G and B signals [33]. The signals usually have to be carried through four coaxial cables (R, G, B and sync) or three, if the synchronisation pulse is carried together with the green (sync on green).

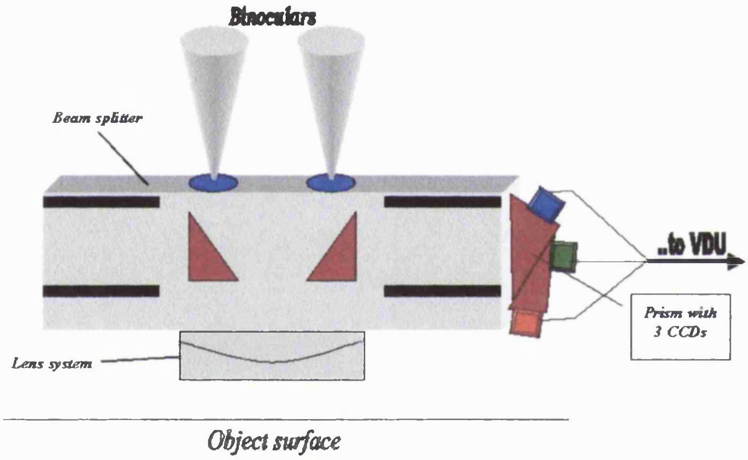


Figure 3.19: Graphical arrangement of a three CCD colour video camera attached onto the side of a microscope's beam splitter.

In the same fashion that the video camera is used to transform an optical image to an RGB signal, the monitor is employed to generate a colour image from the RGB video signal (see Figure 3.20). The monitor accepts the four separate signals (or three if the sync is on green) for its respective channels. The sync pulses are then stripped (split) into horizontal and vertical sync pulses that generate the horizontal and vertical drives of the monitor. As a result, the monitor synchronises with the video camera that

[13] The simultaneous collection of the three colour intensities is achieved using a clock that drives the charged-couple devices.

generates the RGB video signal. The minimum number of pixels required to display the microscopic video image can be calculated by the following equation:

$$\text{min no. of pixels} = (\text{TV lines})^2 \times \text{CCD Aspect ratio}$$
$$\text{where TV lines} = \frac{1}{S} \times \text{Vertical CCD dimension} \times 2$$
(Equation 3.8)

Inside the CRT monitor, the video image is projected onto a glass target that is coated with photoemitters on one side and an electrically conductive layer of pixels on the other. The screen is scanned in a raster by three sharply focused electron beams, emitted by three respective electron guns[14]. When excited by electrons, the photoemitters create an electrical charge on the other side of the glass. Following, bright pixels are discharged until the beam re-approaches and charges them again. Then, a new raster cycle begins as the beam repeatedly scans the screen [34]. When the whole row of pixels is scanned, a TV line is produced.

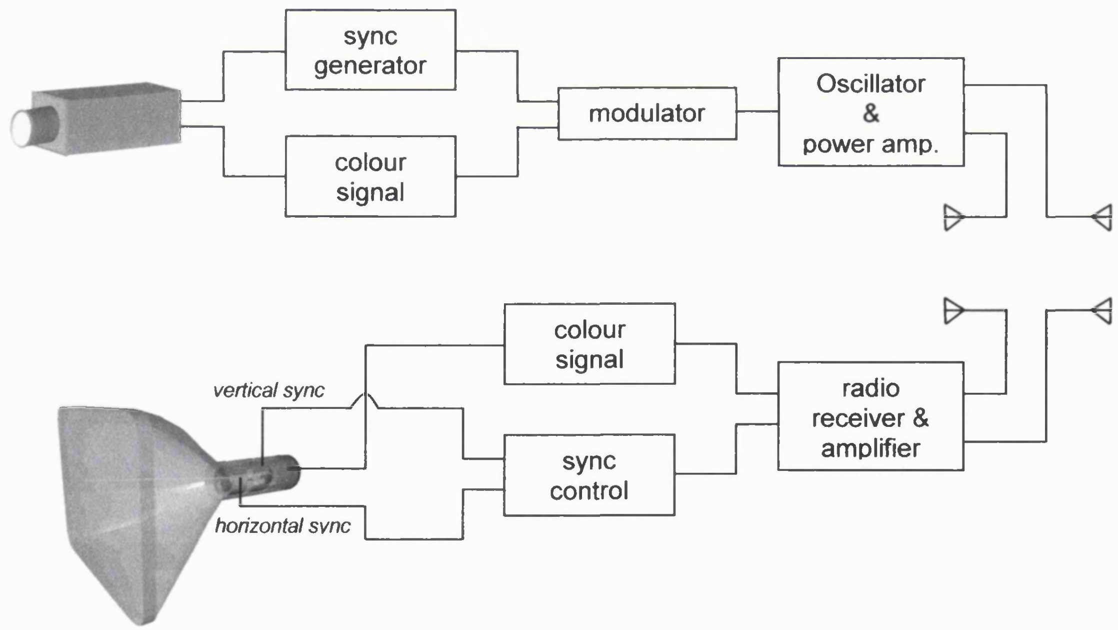


Figure 3.20: I/O signal component diagram of a colour TV system.

14 Raster scanners are devices where two sets of electromagnetic deflection coils from an electron gun are driven in a repeated pattern of horizontal TV lines across a screen. The electron intensity of the gun carries the characteristics of the video signal (e.g. colour and brightness). Colour CRT monitors have three electron guns; one for each primary colour.

The displayed colour format is synthesised by combining the three primary colours, red-green-blue (RGB). Each component presents its own luminance-Y (brightness) and chrominance-C (colour difference). As a result, the image can also be represented through these YC values. Since the human visual system is more sensitive to brightness than colour difference information, the latter can be represented with a lower resolution without significantly affecting the overall image quality [35].

Standards for analogue colour television vary throughout the world but there are three main systems widely used in the present age. The Phase Alternation Line (PAL) system is mostly used in Western Europe while SEquential Colour Avec Memoire (SECAM) system is used in France and in some parts of Eastern and Southern Europe. The United States of America and Japan use the National Television System Committee (NTSC) system. Each system uses three components; luminance Y, blue colour difference U and red colour difference V. The combined signal is called YUV and is used in digital communications (see Chapter 8). In the PAL and SECAM systems, a video frame consists of two fields that are confined in 575 (out of 625) active television lines. The field rate is 50 Hz, hence allowing the transmission of 25 video frames per second[15]. The NTSC system uses a 60 Hz field rate with every video frame being enclosed in 480 (out of 525) active lines, and allows an overall transmission of 30 frames per second.

3.4 Three-Dimensional displays

The creation of natural, stereoscopic displays is an old dream, and has been the subject of active research for over a century now. The first stereoscope was created by Sir Charles Wheatstone (1802-1875). The device consisted of two mirrors at right angles and two vertical picture holders, shown in Figure 3.21. In a later version of the device, the two halves could be rotated to change the angle of convergence. With this arrangement, the effects of binocular variables could be studied quantitatively, providing the experimenter with dichoptic control of the stereo images. The stereoscope also had adjustable arms which allowed the user to vary convergence while keeping the

[15] Since each frame carries two fields, only 25 frames can be transmitted in a second.

disparity constant, and vice versa. With the use of his stereoscope, Wheatstone managed to demonstrate the relationship between binocular disparity and depth perception [36].

Another British physicist of the same period, Sir David Brewster (1781-1868), developed a prototype stereoscope by cutting a convex lens in half and arranging each half with its vertical cut edge on the temporal side of the eye. A partition, placed at the cyclopean point of the two eyes, allowed dichoptic control of the stereo images. The latter, where printed side-by-side about 65 mm apart and viewed through two base-out prisms. In addition, magnifying lenses were incorporated into the prisms. The prism-lens arrangement could offer better, magnified images. Stereoscopes of this type, shown in Figure 3.22, are known as *lenticular* stereoscopes.

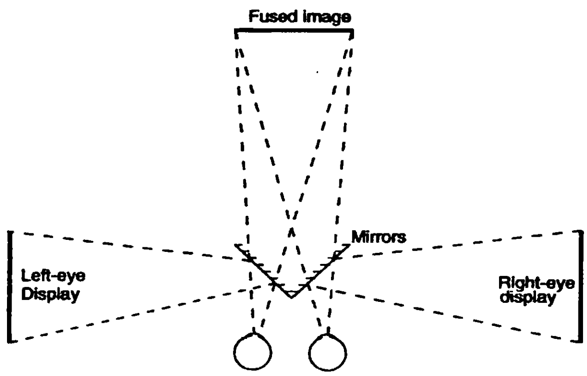


Figure 3.21: Schematic diagram of the mirror stereoscope.

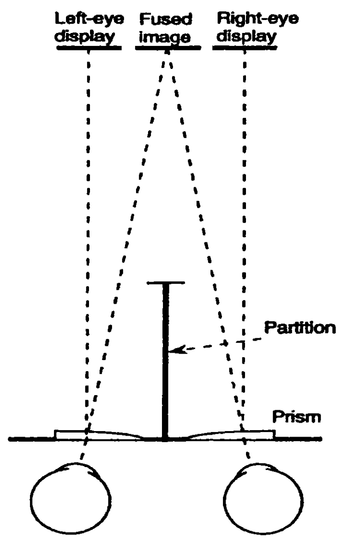


Figure 3.22: Schematic diagram of the lenticular stereoscope.

The success of the stereoscope during the second half of the 1800s was enormous. The new experience of viewing photographs with some depth appreciation was so fascinating that, by the turn of the century, the instrument could be found in almost every household of Western societies. However, the inability to include the fourth dimension into the visual effect limited the realism of the experience. This problem was partially solved later with the inventions of cinematography and television.

Originally, moving pictures were created on 35 mm camera film with the camera remaining still during the movie. Bright, sequential still images were displayed on a screen at a refresh rate of 10-20 frames per second but such display rates introduced flicker (as mentioned in 3.1.2). The invention of television solved the problem of converting an optical image into an electrical signal. The new device could place the viewer in the scene in real time by capturing a video scene miles away and displaying it in front of him/her onto a 2D monitor at a picture update rate of 25 frames per second.

The bulky structure of CRT monitors often becomes a limiting factor in their utilisation. Still, the development of solid-state video monitors helped the invention of laptop computers and flat-panel displays. The most common ones are those based on liquid crystals [37] [38], but recently, technology advancement led to the production of gas-plasma displays [39].

A liquid crystal display (LCD) device is a 2D electro-optical light modulator that is placed in front of a backlight. The light is modulated at each pixel location by applying an electric field to a thin layer of liquid crystal between two crossed polarisers. When charged, the crystal blocks lightwaves from passing through the second polariser, resulting in the appearance of a black pixel on the screen (see Figure 3.23). The screen brightness depends on the intensity of the backlight, while colours are created by placing RGB filters in front of the pixel cells.

Nowadays, the most successful LCDs are based on thin-film transistors (TFT). The latter are arranged in a matrix on a glass substrate. To address a particular pixel, the proper row is switched on, and then a charge is sent down the correct column. Since all of the other rows that the column intersects are turned off, only the TFT at the designated pixel receives a charge. The TFT is able to hold the charge until the next refresh cycle.

Active-matrix LCD technology has been very successful in the TV and video industry. Early LCD resolution was small (*e.g.* 100×100 pixels) but better solid-state electronics have enabled the production of high-resolution displays (*e.g.* 1280×1024).

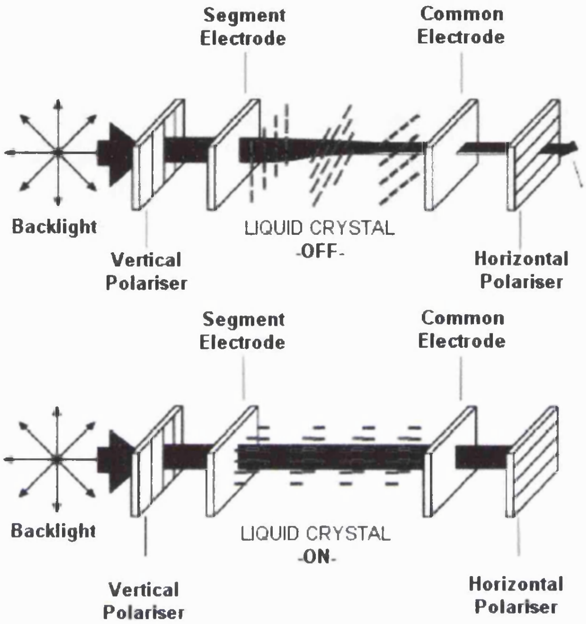


Figure 3.23: Schematic diagram of LCD operation.

3.4.1 Stereoscopic displays

Scientific advancement in stereo vision and better understanding of the physical characteristics of the human eye has initiated the production of 3D stereoscopic displays. Such displays present to the observer a pair of left and right images, which in combination provide the 3D image. The observer can view the final image with the use of goggles.

With this technique, the most common approach for stereo image generation has been through the use of alternate-field video, mediated by LCD glasses [40]. Stereoscopic pairs of images are created by presenting sequential video fields to each eye alternatively, and in rapid succession. Video frames are projected on a display screen while the glasses darken in synchrony with the images so that each eye sees the correct image. However, a different display arrangement can be used. Instead of presenting alternate fields/frames on a single monitor, one can display left and right eye images on two identical monitors, arranged at a right angle. The monitors receive either the first or second field/frame signal. The two pictures are polarised relatively to each other and then combined optically by a beam-splitting mirror. Both the single and double monitor display set-ups are shown in Figure 3.24.

A major disadvantage of both the display set-ups discussed here is that of flicker generation. Unless the displays are driven at a 50 Hz or higher frame rate, the human eye will detect image discontinuity since each image is interrupted 25 times per second.

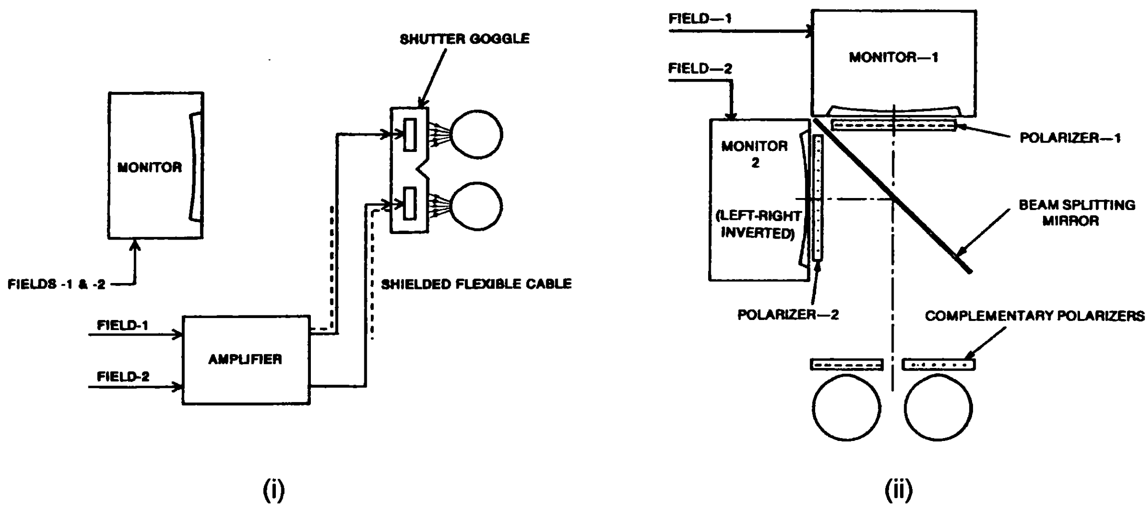
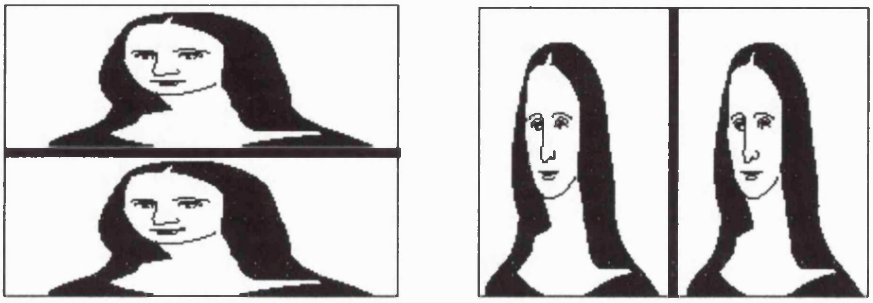


Figure 3.24: Schematic diagrams of: (i) time-multiplexed and (ii) polarisation-multiplexed display, for producing stereo images with video.

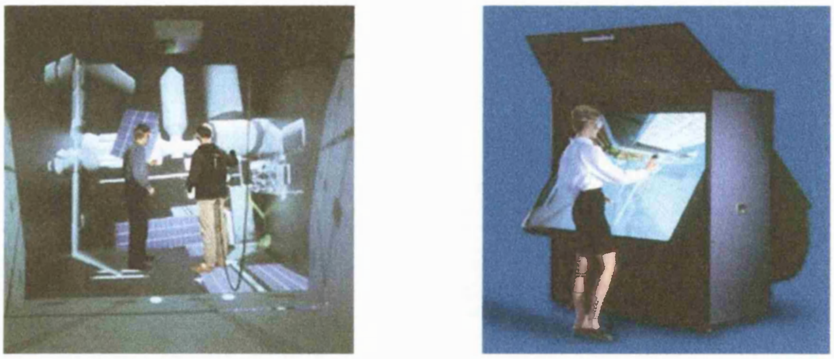
Other LCD systems have also been used for displaying flicker-free, stereoscopic images. At present, the most successful commercial product is *CrystalEyes®*, manufactured by Stereographics Corporation [41]. In this system, stereoscopic pairs are displayed by mapping a vertically compressed left eye sub-field to the top half of the display, and a vertically compressed right eye sub-field to the bottom half (see Figure 3.24(i)). With this video formatting method the refresh rate of image per eye doubles, thus making the display system flicker-free. The above can be applied only to computer monitors that can display interlaced images at high frequencies. However, for PAL or NTSC transmission, such a technique would fail to resolve good images since the number of raster lines would be halved. Instead, for real-time video viewing a side-by-side video format is preferred (see Figure 3.25(ii)) but unfortunately this requires doubling the video bandwidth[16]. As a result, the system is not compatible for real-time viewing and can only act as a playback video interface.



(i) (ii)

Figure 3.25:CrystalEyes® stereo vision formats: (i) above-and-below images: when displayed at 120 Hz/sec a stereo image results, (ii) side-by-side images: left and right images fit into a standard video field.

CrystalEyes® and other off-head display (OHD) systems that use similar display technology may have not been proved very useful in real-time video applications but have had a tremendous impact in computer graphics applications, particularly in virtual reality. Current VR systems that use this technology include the CAVE™ [42] and ImmersaDesk™ [43], shown in Figure 3.26(i) and Figure 3.26(ii) respectively.



(i) (ii)

Figure 3.26: Commercial VR systems using stereoscopic displays: (i) CAVE: a computer-generated environment using a multi-wall display to access high-resolution 3D models and computer graphics, (ii) ImmersaDesk: a fixed-angle display allowing a physical 'heads-on' interaction with the 3D computer environment. Both display systems are commercially available (Fakespace Systems).

[16] The two-fold video bandwidth increase explained here results from capturing a video image using two cameras. Since the PAL or NTSC systems have to be used, the two video signals have to be 'squeezed' into the normal bandwidth.

3.4.2 Autostereoscopic displays

The use of liquid crystal glasses for stereo viewing has been fundamental, especially in VR applications. Nowadays, Head Mounted Displays (HMD) can *immerse* the viewer into a computer-generated 3D scene while continuously tracking any changes in his/her position, orientation or viewpoint [44]. However, there are some serious drawbacks in their use, the most important ones being uncomfortable eyewear, reduced image brightness, image flicker and cross-talk levels of up to 10% [45]. Even in highly sophisticated systems, there are physiological [46] and psychological [47] effects that prohibit long periods of viewing, such as nausea, headache and tired eyes.

Autostereoscopic displays do not require specialised headsets and are generally considered as a more realistic approach to 3D viewing. They offer greater viewing freedom than immersive systems, such as stereo microscope eyepieces and HMDs, and could be useful for multi-view image presentations. There are three main scientific types of 3D autostereoscopic displays: *electro-holographic*, *volumetric* and *direction-multiplexed* displays. Holographic techniques reproduce the properties of light waves (luminance, chrominance and phase difference) almost to perfection, making them a very close approximation of an ideal free viewing world. Volumetric displays project image points to definite locations in a physical volume of space where they appear either on a real surface, or in translucent images forming a stack of distinct depth planes. Direction-multiplexed displays apply optical effects like diffraction, refraction, reflection and occlusion in order to direct light emitted by pixels of different perspective views exclusively to the correct eye. In his paragraph, only **refraction-based** direction-multiplexed techniques are explained as the remaining ones exceed the purposes of this thesis.

Image quality of displays based on optical refraction can be affected by several factors. Cross-talk of the two image channels may occur due to aberrations of the optical system. A limited display bandwidth causes degradation of image quality which results in lower depth perception. Bandwidth constraints are also responsible for the number of views that can be simultaneously displayed.

Head movement problems can be overcome by using observer-tracking displays [48]. The observer's position is normally measured by infrared or video image processing tracking devices. The current trend in autostereoscopic 3D display development is found in the use of TFT monitors. These offer significant advantages such as flatness and thinness, high image resolution, high image contrast and fidelity, good colour quality and acceptable cost.

The design concept of an autostereoscopic 3D display is based on lenticular imaging and field-lens imaging, with recent systems separated into two main categories: **flat panel 3D-LCD** and **Twin-LCD** screens.

A flat panel autostereoscopic 3D display consists of a lenticular array, placed directly above a TFT monitor. With this arrangement, specific display pixels are visible only under specific horizontal viewing angles (see Figure 3.27). If these pixels are charged accordingly with stereo information to match human stereopsis, a 3D viewing window is created in front of the viewer's eyes. The former is provided through spatial multiplexing of the LCD screen[17] while the latter is achieved by aligning each lenticular refractive lens with at least two columns of the LCD pixels [49].

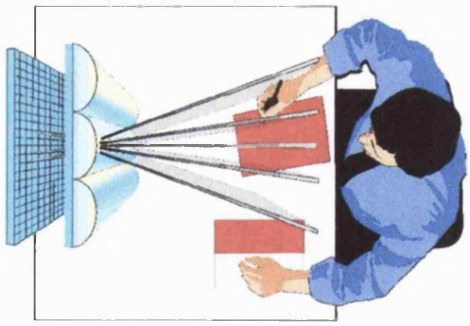


Figure 3.27: Picture projection on conventional 3D-LCD screens. Pixels and pixel boundaries are projected to form discrete viewing zones.

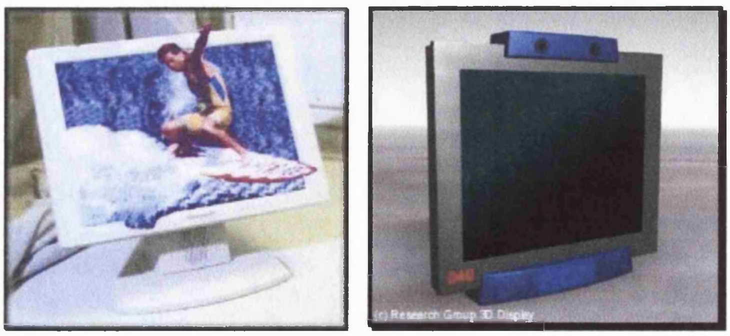
The main side effect of the viewing conditions illustrated in the figure above is that, along with the autostereoscopic effect, the pixel boundaries are also imaged. This leads to the irritating effect of vertical bars through which the viewer has to look at the display. The visual effect of this is that, with sideways movement, Moiré-like fringes appear to run over the screen. A way to overcome this problem is to put the lenticular

[17] The spatial multiplexing process described here is different from the time multiplexing process in stereoscopic displays. In HMD systems the same pixels provide information for different eyes at different times while in autostereoscopic displays left and right eye pixels are spatially separated.

sheet at a slant [50]. Some attempts have been made to create multi-view stereoscopic displays using lenticular flat panels [51], but existing products suffer from poor resolution.

An alternative method of viewing 3D autostereoscopic images is to place a field-lens at the locus of a 2D image in order to collimate the rays of light passing through that image, without affecting its geometrical properties. Such display systems, project fully resolved left-right eye images on two separate LCD screens respectively. The screens, which have a viewing arrangement identical to the one illustrated in Figure 3.23(ii), are optically combined with a beam splitter to create the 3D effect. Twin-LCD micro-optic systems are considered of great prospect as they can offer full resolution of the LCD to each of the observer's eyes [52].

The development of autostereoscopic display systems (see Figure 3.28) runs in parallel with that of VR applications. The ability to display high-resolution, computer-generated images on LCD screens stereoscopically has initiated large-scale objects in various industrial and scientific fields, from military research to industrial assembly units. In medicine, stereoscopic displays look to improve, or even replace, conventional techniques of medical imaging [53], surgical navigation and intervention [54], and education. Their implementation with other VR components and relative applications are discussed in the following chapter.



(i) (ii)

Figure 3.28: Flat-panel 3D autostereoscopic displays from (i) SHARP and (ii) NEC-GWT.

CHAPTER 4

VIRTUAL REALITY FOR HELATHCARE SYSTEMS

Originally, VR technologies emerged as computer tools that could model manufacturing systems and components with effective use of audiovisual and other sensory features. At the time, it was believed that computer simulations would allow partial or total prediction of potential problems in product functionality and manufacturability before the real process began. However, the great asset of the new technology was thought to be found in prototype manufacturing, as that would improve productivity and cost-effectiveness of new products. Other applications were limited due to the lack of computer power, but as new computer hardware and software were introduced during the last decade, VR found new fields of usage in scientific, technological and medical fields.

In this chapter, the reader is presented with a literature review of VR applications, from their beginning up to this date. More emphasis is given in the surgical applications of VR and AR systems, as they form the field of the research work described in this thesis.

4.1 VR applications

The first application of VR was found in military research. The large costs of training jet fighter pilots with complex aviation systems forced the US Navy to invest on simulating flight conditions. One of the first products was built by General Electric in 1972. The Advanced Development Model was used as an instrument to measure the effectiveness of computer-generated images for pilot training. Three independent projection screens surrounding the physical mock-up of the cockpit, provided pilots with rendered images that covered an 180° FOV. The cumbersome size of the simulator and the reduced realism of the flight environment were the main limitations of this early attempt.

In chemistry, scientists attempted to employ VR systems for viewing and manipulating molecular structures. In one of the early projects a VR tool was used for molecular docking [55]. One of the important innovations of the new tool was the creation of haptic feedback mechanisms. A mechanical handgrip that exerted physical force against the operator's hand was simulating the physical feel of docking a molecule.

In the mid '80s, VR technology was used for the first time in architectural design [56]. A 3D model of the structure was created using the original blueprints of the building under construction. Computer graphics enabled views anywhere in the model, with movable handlebars allowing the user to physically walk down hallways.

Rendering computer-generated images was indeed very difficult during the '80s. Some computers had the ability to construct complex geometrical space, but only at very low speeds. As a result, most VR tools suffered from offering a pragmatic interaction. In most cases [55], the models had to be simplified before being used.

In NASA Ames Research Center scientists were familiar with the rise of VR technologies, and in particular flight simulators, as they could offer good training ground for their astronauts. Moreover, they had the computer resources to render complex computer environments. In 1988, they constructed the first ever, fully immersive VR system, called the Virtual Interface Environment Workstation (VIEW). The system comprised of a HMD, 3D sound, voice recognition, voice synthesis and a device for manipulating virtual objects that was named *DataGlove* [57].

In the realm of medicine, VR was firstly used in surgical training simulators [58]. Virtual world models that were created by CT, MRI, ultrasound and other 3D imaging modalities led to the birth of Image Guided Surgery (IGS). The first application was in virtual endoscopy [59], but soon the technology was used in stereotactic neurosurgery. The degree of realism in medical applications is dependent on:

- Fidelity (*e.g.* high-resolution graphics),
- Display of organ properties (*e.g.* kinematics of joints),
- Display of organ reactions (*e.g.* bleeding of arteries),

- Interaction between organs and surgical instruments (*e.g.* bone burring),
- Sensory feedback (*e.g.* tactile and force feedback),

However, realistic viewing of the virtual models is not enough. The system must also ensue simple and easy interaction between the user and the computer [60].

At present, and with the progress of computer hardware and software, VR systems offer enormous potential of exploration in all fields mentioned above. There are hundreds of academic groups exploiting the full potential of the technology, specifically AR systems as they offer a substantial amount of realism. From the spectacular graphics of the latest computer games and other multimedia applications, to tele-robotic surgery that improves the quality of therapy and rehabilitation, AR systems are becoming the archetype of HCI.

In an AR system, overlaying the associated real-world information with information that is stored in the computer enhances the user's interaction with the real world. Such a system can be used in many industrial applications like exterior construction, interior design and plant monitoring. At the start of an exterior construction project, AR can help architects and customers visualise newly designed structures in the context of the already existing environment. In interior design, AR can help architects design, remodel and visualise rooms. A typical application is the augmentation of a live video stream from a room in real-time with virtual furniture from a database which can be interactively moved about (see Figure 4.1). For automotive construction, AR can be used to verify, edit and train assembly/disassembly processes. The latter can also be recorded in AR or come from a virtual process model, thus closing the gap between virtual prototyping and the real world.

Figure 4.1: Mixed Reality living room

4.2 VR systems in medicine

In general, healthcare applications of VR environments and related technologies can be divided in the following categories [61]:

- Surgical procedures (*e.g.* pre-operative planning and simulation, IGS, AR surgery),
- Medical therapy (*e.g.* radiotherapy),
- Preventive medicine and patient education,
- Medical education and training,
- Visualisation of large medical databases,
- Skill enhancement and rehabilitation,
- Architectural design of health care facilities,

In the following pages, the author explains some of the current applications in computer-assisted surgical planning, intervention and training (CASPIT).

4.2.1 Stereotactic neurosurgery

In the past, the task of correlating pre-operative imaging data to the surgical ROI was left to the surgeon's knowledge of human anatomy. Surgical planning and intervention techniques have improved the surgeon's precision in the operating theatre. However, one of the principle problems with using VR images is how to accurately relate them to the patient's position during the surgical procedure. Stereotaxis is a technique that was developed in 1947 for using the concept of a 3D Cartesian coordinate system for the human brain [62]. These coordinates can be defined with respect to both the brain and the radiographic images acquired while the patient's head is fitted with a stereotactic frame, found at the tip of a frame carrier. In essence, the frame itself provides a rigid reference system that establishes a coordinate system relative to the patient, presents identifiable landmarks in the images and allows safe mounting and guide for the tools used during the neurosurgical procedure (see Figure 4.2).

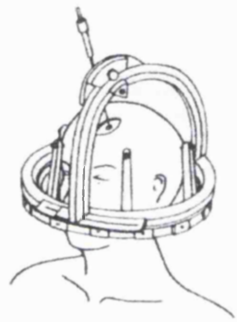
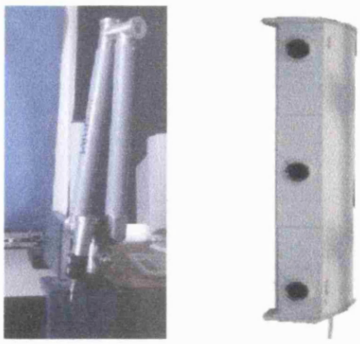


Figure 4.2: Stereotactic frame for image-guided surgery.

The ability to provide 3D visualisation of an intracranial target through the position of the stereotactic frame enables the surgeon to carefully plan the surgical procedure prior to the actual operation. Pre-operatively, small metallic blunt pins (or plastic plates filled with copper sulphate fluid in case of MR imaging) are placed onto the patient's head in six[18] different positions. During the scan, anatomical structures and fiducial markers are localised within a well-defined space, and the coordinates within that space can be read directly. Next, the CT or MRI slices are transformed to 3D images in which structures such as skin surface, skull and tumour sites are constructed as 3D objects. The latter process is called **volume reconstruction**, and is very important in being able to know precisely the patient's head position with respect to a known reference coordinate system.

[18] The number of fiducial markers on a stereotactic frame can vary. However, at least three markers are needed to register a 3D volume.

Stereotactic frames have proved disadvantageous as they are cumbersome and involve a minimally invasive procedure so that they can be fitted onto the patient's head. Alternatively, frameless systems can be used as long as there is a tracking device that can register the patient continuously during pre-operative planning or surgical intervention. At present, there are four types of tracking systems (see Figure 4.3): mechanical, electromagnetic, ultrasonic, and optical, each with six degrees of freedom. Mechanical trackers use a probe that is physically linked by a multijointed arm to the apparatus restraining the patient's head, while the position of the probe is determined by sensing the angles at each joint [63]. Ultrasonic [64] and optical tracking devices use signal transducers to determine the position of the patient's head with respect to the tracker's origin. Each type of tracking system has advantages and disadvantages. Mechanical trackers are always in communication with the computer but are bulky and difficult to use in the limited space of the operating theatre. On the other hand, optical and ultrasonic trackers need a clear line of sight between the transmitting and receiving ends of the tracker. Electromagnetic systems [65] do not suffer from this effect but the presence of metallic components in the vicinity may distort the sensors' reading, thus limiting tracking performance. The accuracy and precision of most today's systems are reported to be better than ±1 mm [66], which is good enough to track the smallest possible movement of surgical tools.



(a) (b)

Figure 4.3: Tracking devices: (a) Mechanical tracking device (the FARO arm, Faro Industries Inc, Florida, USA), (b) Optical tracking Device (Optotrak, Northern Digital Inc, Ontario, Canada).

4.2.2 Endoscopic surgery

Minimally invasive surgical techniques have been developed from the need for better quality of care. Nowadays, keyhole surgery ensures minimal damage to the skin, muscle, surrounding tissue and organs. Endoscopic surgery can be performed in many parts of the human body: the gastrointestinal tract; gastroscopy [67], joints; arthroscopy [68], the abdomen; laparoscopy [69], coronary arteries; arteriography [70], the chest; thoracoscopy [71].

The endoscope is a tubular instrument through which images of the surgical ROI are presented to the surgeon via a video monitor. There are two types of endoscopes: flexible and rigid. The former are designed to extend through non-planar paths, conveying images from the ROI through fibre optics to a video camera at the viewing end. The camera is connected to an eyepiece, and images are displayed onto a 2D video monitor [72]. Rigid endoscopes are preferred for applications where the ROI can be approached easily through the incision.

Virtual endoscopy is a newly developed technique for IGS [73]. Computer graphics of pre-operative CT or MRI scans simulate the view of an endoscope that is placed in a body cavity. In its simplest mode of operation a tracking device tracks the tip of the instrument, while the three orthogonal X-ray slices that intersect the tracked point are displayed onto the monitor. In a more complicated approach, 3D models of the ROI can be visualised through computer rendering of the data volume.

The visualisation process (see Figure 4.4) is completed in five main image-processing steps: *acquisition*, *reconstruction*, *segmentation*, *labelling* and *rendering*. Data acquisition is based on the imaging modality that is used for diagnosis. Modern 3D scanners produce 512×512 -CT-, and 256×256 -MRI- pixel array, axial radiological slices of the patient's anatomy. Then, scanner software (*e.g.* 3D Slicer, Analyse, etc.) reconstructs a 3D volume dataset that is comprised of volume elements, or so called voxels. By thresholding the greyscale value of voxels in CT data, low intensity ones are removed from the 3D volume, thus separating bone structures. As a result, soft tissue, blood vessels, and other structures of interest, are easier to identify. This process is known as **segmentation**. As soon as this step is finished, the structures can be labelled, based on the properties of the observed intensities as well as known anatomical information about normal subjects. The final step of visualisation involves displaying the volume. There are two main types of rendering: **surface** and **volume rendering**. In surface rendering of volume data, polygons (*e.g.* triangles) of the structures are first derived from the voxel values. The former share the same properties, and when interconnected, they form an isosurface of the structure. The latter normally corresponds to the surface of an anatomical structure or to surfaces of equal functional activity. The abstracted isosurface is then isolated from the remaining dataset and can displayed at interactive rates on a PC. In volume rendering, each voxel inside the volume dataset is projected onto a viewing plane with a value related to the physical property represented in the voxel array. This can be done by applying a ray-tracing technique. As a result, each pixel of the projected image is a function of all voxels traversed by each ray. Volume rendering is preferred to surface rendering for fine visualisation of anatomical structures, a very important factor in developing CAS systems. However, the heavy computational process of rendering datasets requires the presence of powerful graphics computers that carry out computations in graphics pipelines rather than computer processors.

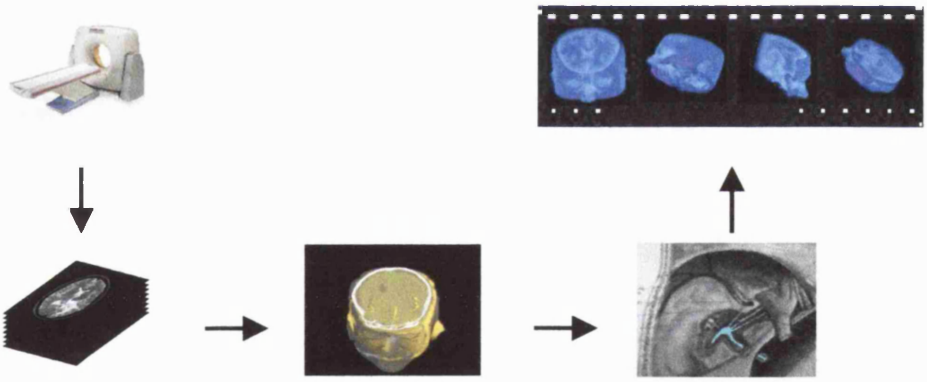


Figure 4.4:Visualisation process of 3D virtual MRI diagnostic data.

4.2.3 Prosthetic surgery

Clinical use of 3D computer visualisation is also found in other types of surgery, such as maxillo-facial [74] and orthopaedic [75] surgery. In cosmetic surgery, a facial anomaly (*e.g.* defective skull) can be treated by prosthetics. Titanium plates that correct the defect are designed from the 3D computer reconstructions of the CT skull dataset. Using VR simulation techniques one can see the predicted outcome of the prosthesis [76]. After the shape and size of the plate are determined, the numerical coordinate model is manufactured on a high-precision CNC milling machine. As seen in Figure 4.5, the technique offers very good quality of treatment.

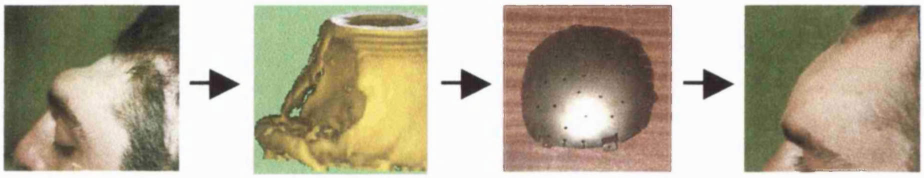


Figure 4.5: Craniofacial anomaly correction technique, based on VR simulation.

4.2.4 Medical Education and Training

Computer-based training (CBT) of medical practitioners is considered as one of the prime applications of VR technology since the user can actively engage in a learning task while experiencing multiple computer-generated sensory stimuli. This is especially true for surgeons, who find it difficult to get 'hands-on' experience of operations; VR simulations offer an alternative method for both educational and training purposes. In most examples, 3D visualisation of the human body enables students to understand better the important physiological principles and anatomy. The 'visible human' project of the US National Institute of Health is the most representative case of VR medical education research [77].

The Institute of Laryngology & Otology, University College London, has developed a multimedia-training course in interactive rhinology [78]. The software is implemented using HTML and Java and thus, only requires an Internet browser to run. The operating platform is Windows-based and has the structure of a web page with buttons and images to click on, areas that show more information when the mouse cursor is placed upon them, as well as on-line help and tutorials. It also includes videos of surgical interventions and movies of virtual 3D anatomical structures, reconstructed from CT and MRI scans (see Figure 4.6).

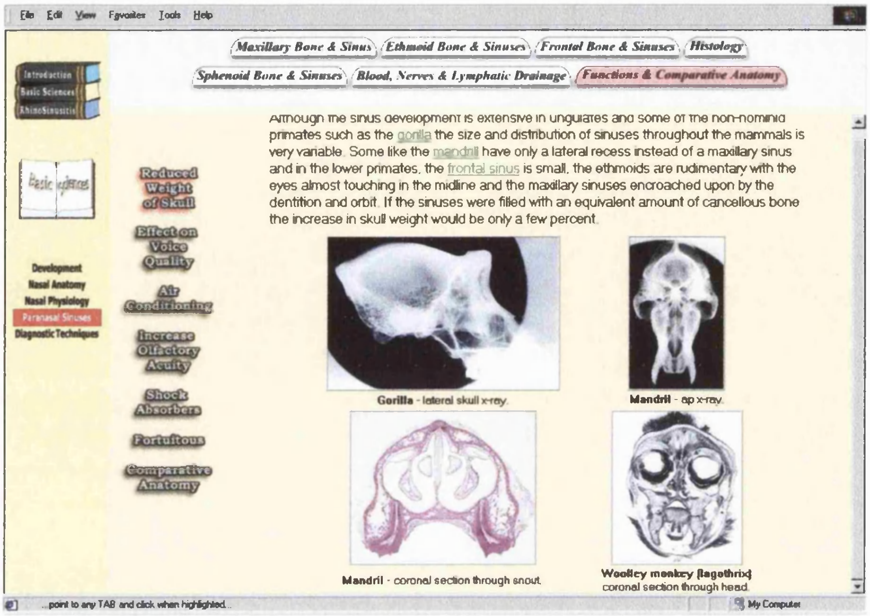
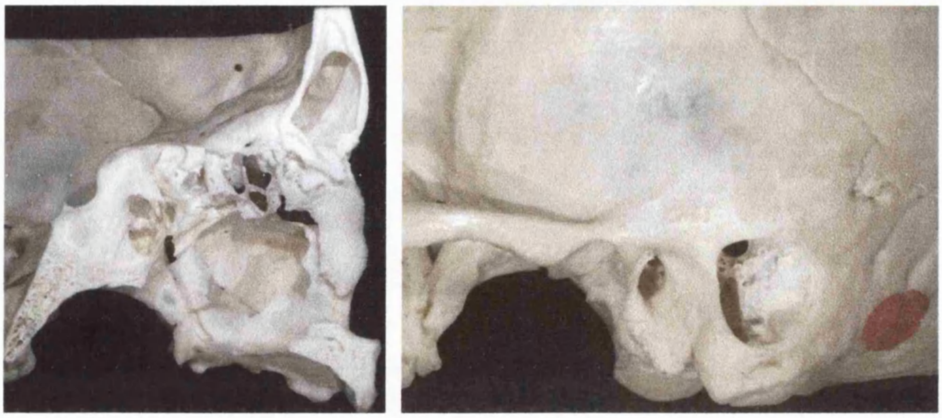


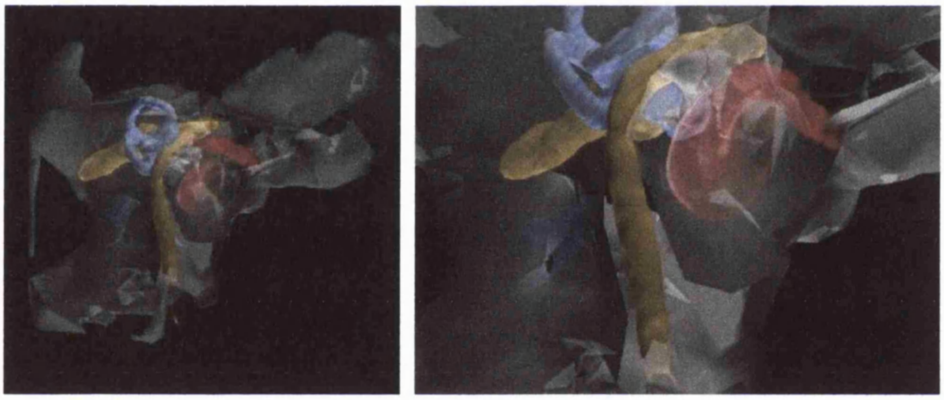
Figure 4.6: The interaction panel from the 'Interactive Rhinology' CBT course as developed at the Institute of Laryngology & Otology, University College London – with permission.

Otology is known as the branch of medicine that deals with the pathology and treatment of ear diseases. Skull base surgery is a very difficult type of surgery partly because of the complicated shape and miniature size of the anatomical structures that are present in this region of the human body, but also due to the skeletal characteristics of the **temporal bone** (see Figure 4.7), the part of the skull that protects it. The latter, makes minimally invasive techniques necessary but also difficult to apply. Surgical training is normally carried out on cadavers which lack some of the characteristics of the real experience (*e.g.* bleeding). Moreover, the anatomy can be studied only at microscopic level while the position and orientation of structures cannot be clearly described [79]. As a result, there is limiting teaching of trainee surgeons that is often acquired by conventional tuition inside the operating theatre. Researchers at the University of Illinois, Chicago, have developed a VR interactive environment for studying the anatomy of the human ear [80]. By using a projection-based viewing system (ImmersaDesk), viewers acquire an inside perspective of the temporal bone and the physiology of the middle ear (see Figure 4.8).



(a) (b)

Figure 4.7: Axial (a) and side (b) views of a cadaver's temporal bone. The petrous part of the bone occupies only a small volume of the skull (about 7 cm x 5cm x 3 cm).



(a) (b)

Figure 4.8:The human temporal bone in its entirety (a). Here the bone has been made transparent to reveal the delicate anatomic structures imbedded within bone: A close-up in the region (b) reveals the organs of hearing and balance, nerves, blood vessels, muscles, the eardrum and ossicles (courtesy of University of Illinois, Chicago).

4.3 Augmented Reality surgical systems

In surgery, AR systems fuse VR images with real time images or superimpose them on the body using video technology. The purpose of this technique is to provide surgeons with enhanced viewing of selected ROIs and, by superimposing 3D reconstructions of pre-operative diagnostic scans allow them to intervene with greater success. In practice, anatomical structures such as skin, bones, tumours, etc. can be viewed together with virtual images of the same morphology. As a result, the surgeon is provided with 'X-ray' vision of lesions that might otherwise not be visible if they reside deeply in the human body.

The success of an AR surgical system depends on the following parameters:

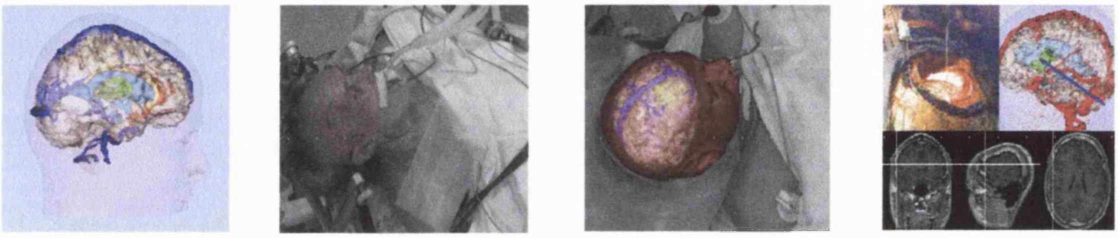
- High-fidelity registration,
- Tracking accuracy and precision of surgical components,
- Real-time rendering and display,
- Taxonomy and ergonomy,
- Cost-effectiveness,

Currently, the use of AR surgical systems is considered for several types of surgery, including neurosurgery, endoscopic, ENT and orthopaedic surgery. In the following pages, four representative systems are presented together with their advantages, disadvantages and limitations. An extensive literature review of VR and AR systems in CASPIT can be found in Moline's survey of VR systems for healthcare applications [81].

4.3.1 AR Neurosurgery

Surgeons at the Brigham and Women's hospital in Boston, Massachusetts, developVR models of the brain in order to assist operations in real-time. By using AR techniques, acting surgeons can visualize internal structures through an automated overlay of 3D reconstructions of the internal anatomy on top of live video views of the patient.

Patient **registration** is achieved through laser scanning of the patient's skull as the latter is positioned onto the operating table. A registration algorithm is then used to match the MRI skin with the laser scanned 3D surface data of the skull, before a second algorithm transforms the MRI coordinate frame to the operating theatre coordinate system. At the same time, the coordinates of a video camera, which transmits real-time video images of the ROI, are also transformed to the theatre's coordinate system. Subsequently, mixed images of the real and virtual environments from the camera's viewpoint are displayed onto a 2D monitor, thus allowing the surgeon to identify the exact position of lesions underneath the skull surface. An optical tracking device, consisting of light emitting diodes (LEDs) and CCD camera sensors, reports the position of surgical instruments in and around the ROI [82]. Figure 4.9 shows the four main steps of the CAS approach.



(a) (b) (c) (d)

Figure 4.9: CAS process using AR at the Brigham & Women's hospital, Boston. (a): 3D computer reconstruction of internal anatomy; (b): Laser-scanning of skull surface; (c): aumgentation of real and virtual environments; (d): software display window of live video, and trajectory of computer-assisted surgical guidance.

The CAS system described above was further developed in order to perform intra-operative imaging of patients undergoing surgery. The reason for developing such a tool is because of the fact that during neurosurgery soft tissue deforms, producing positional shifts of up to 20 mm [83]. To accurately predict the exact location of lesions interventional MRI techniques have been used to reconstruct the anatomy in-situ, during the operation [84].

The CAS system developed by the Massachusetts-based research group has been proved very valuable in particular neurosurgical cases, where apart from the increased success rates of operations, the length of surgical procedures is found to be reduced. This is mainly achieved through the system's high positional accuracy (<1 mm) and little interference with the remaining operating theatre equipment. Moreover, the structural design of the system is easily implemented in current working conditions. A drawback may exist if the system is to be used for cases that need real-time, rigorous rendering of large volume datasets due to the limiting performance of the current computing platform. Overall, the greatest disadvantage of the system is that the surgeon must still look away from the operating scene in order to view the guidance information. Furthermore, it lacks the ability to display stereo images of the operating ROI, but such interfaces are not really needed for this type of surgery.

4.3.2 AR Orthopaedic surgery

AR technology has also been used in orthopaedic surgery for total hip replacement, bone fragment recovery, etc., for assisting the surgeon to place prosthetic implants accurately. Data visualisation of pre-operative CT scans is more accurate when comparing with the neurological example mentioned earlier, as the structures involved here have rigid anatomy and do not swell during incisions.

One particular system that has been developed by researchers at Carnegie Mellon University, Pittsburgh, uses an 'Image Overlay' display technique where AR images are transformed in real time so they appear to the user as an integral part of the surrounding environment [85]. The system consists of four basic components: a computer graphics workstation (Silicon Graphics Indigo-II R10K), a position tracking system (Northern Digital OptoTrak 3020), a registration software (HipNav), and a semi-transparent stereoscopic display.

The tracking device used by the group determines the viewer's eyes positions, the patient, the display itself, and the surgical instruments used in the surgical procedure. Registration is achieved by collecting points on the anatomical region and matching them with the pre-operative scans. The system possesses most of the AR characteristics of other CAS systems, with the major difference that augmented images are displayed stereoscopically to the user. This implies more computer hardware platform power since the two monoscopic views must be rendered separately at 30 frames per second per eye (for NTSC video format). On the display end, the viewer looks at the surgical scene directly through a beam splitter glass, while simultaneously seeing a reflection of the virtual image from the glass (see Figure 4.10). This image appears to float below the glass and at exactly the same distance that the patient is below it. This is very important because overlays that are placed too far above or below the ROI confuse the surgeon's focus on the patient.

The greatest advantage of this CAS system is that the surgeon does not need to look away from the surgical field of view to see the virtual enhancement – the latter appears within the patient. In addition, the stereoscopic view enhances depth cues, making the implantation process more accurate than conventional methods. However, there are numerous disadvantages, the most significant being the poor stereoscopic perception of the 3D environment. This effect occurs mainly because the user needs the CrystalEyes® glasses to see the images stereoscopically. By tracking the position of the glasses and making assumptions that the user's eyes lie behind the glasses and look straight forward, the system introduces objective and subjective inaccuracies.

Figure 4.10: Orthopaedic surgeon using the 'Image Overlay' CAS AR system.

4.3.3 AR microsurgery

The integration of microscopic images with those acquired from pre-operative CT and MRI scans is the most obvious CAS approach in AR microsurgery. The augmented images can be integrated electronically in a digital or analogue fashion on a computer monitor or video display. Alternatively, the images can be projected into the visual field of the microscope (see Figure 4.11). Another option is to incorporate the images into HMDs where the position of the surgeon wearing the helmet and the images corresponding to his/her viewpoint are projected onto his/her FOV [86].

One can easily understand that the application of AR in microsurgery is an extremely difficult affair, not only due to the microscopic operating environment but also because of the microscope's optical properties. Moreover, in some applications such as ENT surgery, fusion of CT and MRI scans maybe needed if lesions lie very close to bone structures while at the same time are attached to soft tissue.

A clinical system, called Microscope-Assisted Guided Interventions (MAGI), has been developed at UMDS Guy's hospital, London [87]. The system injects computer-generated stereo overlays into the eyepieces of a surgical stereo microscope (see Figure 4.11), thus allowing perception of depth. The resultant AR stereo effect allows surgeons to view hidden critical structures such as tumours and arteries as if they were `beneath' the surface of the surgical scene.

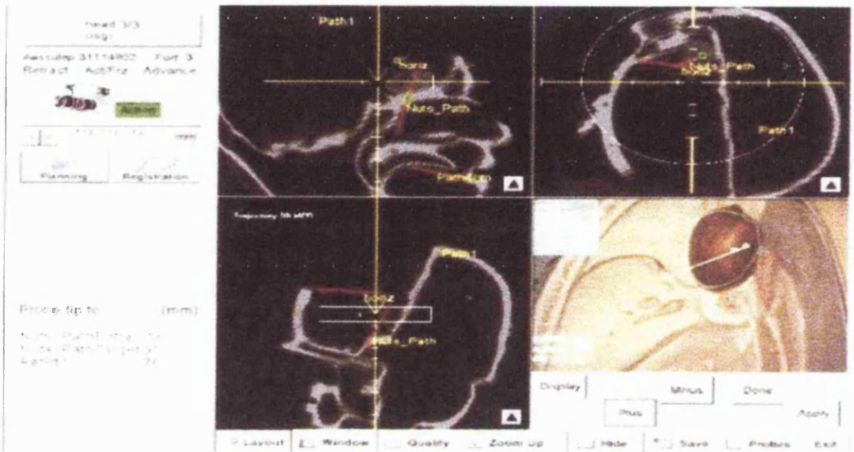


Figure 4.11: Merged data from a tracked microscope and computer model of tumour derived from the MRI and CT scans of a phantom study.

To overlay stereo perspective projections of image data into the microscope's FOV in the correct position, the system employs a **calibration** procedure that consists of four steps (see Figure 4.12). The first step relates to calibrating the microscope optics[19]. A calibration object, with LEDs attached onto it, is placed within the microscope's binocular FOV. Another frame of LEDs, which are mounted onto the microscope's body and at a standard offset distance from the instrument's objective lens, records the position and pose of the microscope. An algebraic matrix ($R_{m\leftarrow o}$) then calculates the calibration object's LED locations to the microscope coordinate system (MCS). The same locations are also marked with a cursor in the 2D image of the microscope eyepiece, hence finding the exact (u, v) pixel position of LEDs in the image coordinate system (ICS). A second matrix (P_m), which projects any point in microscope coordinates onto the position that it appears in the overlaid display, is calculated so that it aligns the 3D MCS with the 2D ICS. Since viewing is stereoscopic, separate projection matrices are obtained for each microscope camera.

[19] Microscope calibration involves the extraction of intrinsic and extrinsic parameters of the video camera(s) used to capture the FOV of the object under the instrument. The extrinsic parameters are derived from the properties of Euclidean geometry, i.e. translation and rotation while the intrinsic parameters are determined form the camera profile, i.e. focal length and radial lens distortion.

In the second calibration step, the 3D VR model of the clinical data is reconstructed, and the important anatomical structures are segmented. Registration is done manually by choosing specific anatomical landmarks and a transform matrix ($R_{p \leftarrow VR}$) determines the actual position of the VR data with respect to patient's head.

The third calibration step establishes the position of the patient relative to the operating theatre. The coordinate system of the latter is called the World Coordinate System (WCS). An LED pointer is used to mark the anatomical landmarks on the skin surface, as in the case of the VR model. A transformation matrix ($R_{p \leftarrow t}$) then calculates the position of the landmarks with respect to the WCS.

During the final step of the calibration procedure, a transformation matrix ($R_{m \leftarrow t}$) is used to calculate the microscope's LED positions with respect to the origin of the WCS. The latter is normally the stationary origin of the tracking device employed by the CAS system.

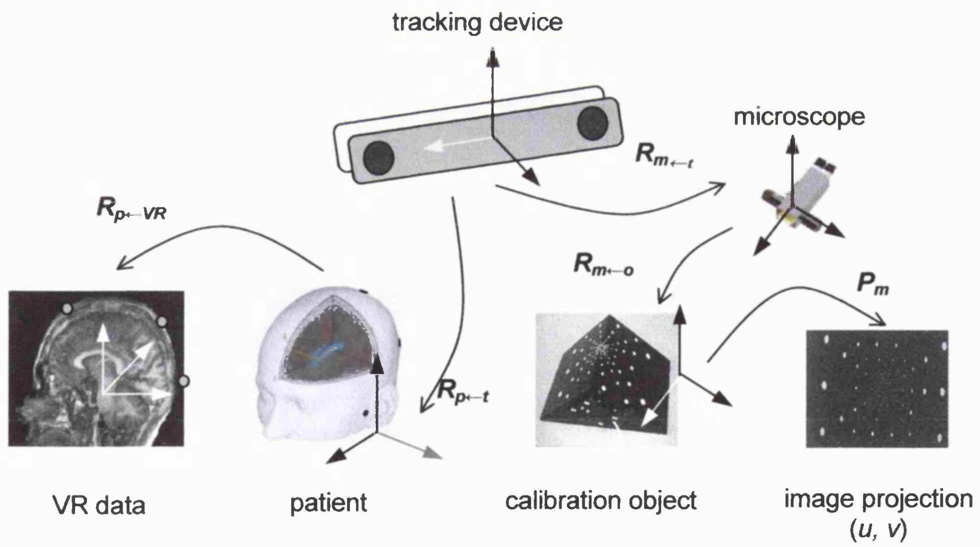


Figure 4.12: Mathematical transforms and projections during the calibration procedure. Several matrices must be computed to enable microscope tracking and relate its position to the patient's image: $\mathbf{R_{m \leftarrow t}}$ *represents the transform between the WCS and the microscope LEDs.* $\mathbf{R_{m \leftarrow o}}$ *relates these LEDs to the object under the microscope.* $\mathbf{R_{p \leftarrow t}}$ *is the transform between the patient's head and the WCS, while* $\mathbf{R_{p \leftarrow VR}}$ *is the transform that aligns the VR image with the real position of the patient's head.* $\mathbf{P_m}$ *is the projection matrix that aligns the 3D MCS with the 2D ICS.*

The quality of the calibration procedure determines to a large degree the overall accuracy of a CAS system. The many calibration steps involved here introduce errors. Their sources in the system can be broken down into three categories: error in automatically localising the calibration points; error in localising the landmarks used for registration in the pre-operative images; and tracking error. Typical values for each of these sources are estimated to be 0.5 pixels for calibration marks, 0.4mm for imaging markers, and 0.2mm for tracking. The error statistics of interest are: calibration error (0.26 mm at the focal plane), 3D alignment error (1.61 mm error when aligning the 3D VR head with the real one), and the final 2D overlay error (0.91 mm at the focal plane), accumulating to an overall positional accuracy of 2-5 mm [88].

The significant overall positional accuracy error limits the system's reliability in CASPIT applications, as this needs to be less than one millimetre, which is the surgeon's minimum hand movement. However, stereoscopic viewing allows the perception of depth which, as mentioned before, is vital in such applications. Moreover, the stereo AR overlays are projected into the microscope binoculars, allowing the surgeon to operate in a familiar environment. Still, the choice of this viewing interface does not reduce the stress factor involved when operating through the surgical microscope.

CHAPTER 5

DESIGN CYCLE I: FOCUS ON USER

The principle of the design process of a CAS system is to aid surgeons without distracting them from their most important tasks, such as bone burring and tumour removal. The whole objective of ergonomics in a working environment is to maximise the operator's safety, efficiency and reliability of performance, to make a task easier, and to increase the feelings of comfort and satisfaction [89]. The concept of ergonomics is very important to the surgeon. A well-designed CAS system should not only preserve the state of ergonomics of the conventional procedure but also aim to improve it.

One of the important fields in ergonomics is the analysis of the human behaviour in dynamic systems where people interact with machines. This implies observation of events such as tasks, postures, movements, gestures or episodes of communication or co-operation. The latter procedure is not straightforward as, for specific applications, some of the events run in parallel and/or show overlapping. In this case, task logging becomes quite difficult and cannot be carried out with a standard Personal Digital Assistant. Instead, the observer needs to design a **mental model** which he/she derives after being introduced to the environment of the HCI and the total number of events at the time of operation [90].

Mental models are either analogous representations or a combination of analogous and propositional representations. They are distinct from, but related to images. A mental model represents the relative position of a set of objects in an analogous manner that parallels the structure of the state of objects in the world [91]. An alternative definition explains a mental model as the model people have of themselves, others, the environment, and the things with which they interact [92]. People form mental models through experience, observation, training and instruction. Important human factors that are studied through the mental model comprise the fundamental **design information framework** (DIF) of a product [93]. The latter can be used to record and organise various design elements when a designer collects data on users and conventional systems from user studies. By focusing on the user, developers aim to understand users' cognitive, behavioural, attitudinal and tasks characteristics.

A DIF may consist of two different organisations [94]. One is the organisation of *design information elements* which define the nature of information. The other is the organisation of *design information primitives* which are the basic descriptors of information about users, objects, behaviours, etc.

For the design of a new 3D viewing interface for CAS using AR techniques the author has created a DIF that is based on four design information elements: **motive**, **activity**, **context**, and **need**. In general, motive is the basic need to satisfy users themselves. Activity consists of events or chains of events [95]. The latter can be separated into internal and external events. External events normally refer to observed physical actions while internal events refer to unobserved cognitive actions. Context is described by the natural setting. It involves the relative position of objects, user, etc [96]. Need is how the user would like the interface to function. For the specific application the four design elements are described as follows:

- Motive: perform surgery better and more comfortably,
- Activity: execute surgical tasks,
- Context: the current set-up of the operating theatre,
- Need: an alternative viewing technology,

Design information primitives are critical factors in design activities and for building a design knowledge base [94]. Listed below, are the ten design primitives considered in this study:

1. Entities: all elements that include agents, users, and objects,
2. Agents: active performers of interactions (*e.g.* the stereo microscope),
3. Users: the active human performers (*e.g.* surgeon, nurse, consultant, etc.),
4. Objects: targets of or receptors of behaviour (*e.g.* patient),

5. Spaces: the 3D physical space involved in the domain of concern (*e.g.* the ROI),
6. Behaviours: routines performed by users and agents (*e.g.* surgical tasks),
7. Locations: any geometrical positions involved in the domain of concern,
8. Attributes: description of structurally fixed properties of entities by means of states and time [97],
9. States: certain values of attributes in entities,
10. Time: period of activity,

Design information elements and primitives were established in the early stages of the design process. Information was collected through user studies.

A possible way of using DIFs is by creating aspect models of HCI. For example, a certain use of an interface can be modelled through a location-based view or through a behaviour-based view. The integration of the two models can reveal enhanced information by showing meaningful relationships among the aspects [98].

5.1 Interface aspect models

5.1.1 The physical model

The research work explained in this thesis attempts to implement a new 3D viewing interface for computer-assisted ENT surgery by creating a design process that guarantees a good HCI. To ensure this, the design of the interface must be consistent with its usage criteria at the time of AR microsurgery. The microscope, together with the remaining instruments used in the operating theatre must be in a safe and ergonomically convenient position around the operating table. Since the surgeon will operate whilst looking directly at the 3D autostereoscopic display, it is necessary that the latter be placed at the appropriate location in the operating theatre. With the existing arrangement of equipment (see Figure 5.1) around the table, this is difficult because the microscope is mounted at the lower end of a stand (see Figure 5.2(i)), thus affecting the surgeon's line of sight. The problem is shown more emphatically in Figure 5.2(ii) where the surgeon moves his head downward in order to look through the microscope. The main effect of this posture is neck strain, a symptom that many surgeons complain about after long hours of practice.

Another major issue in operating theatre space is the strict regulations for any additional electrical and electronic components. This is mainly caused by the packed presence of other medical equipment (*e.g.* physiological monitoring, burr and sucker control units), as well as clinical staff, around the operating table. As a result, the new interface modules need to be placed away from the microscope.

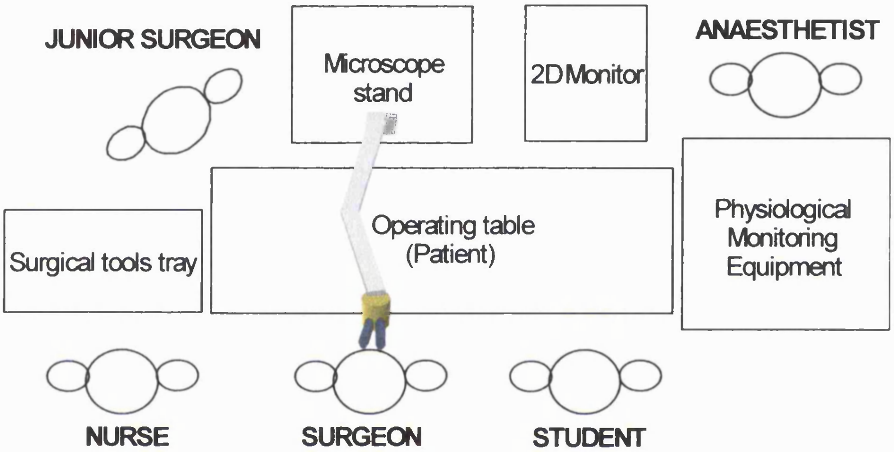
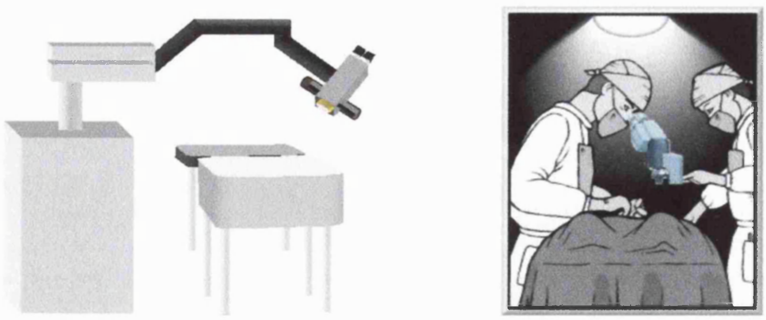


Figure 5.1: Typical arrangement of staff and equipment inside the operating theatre.



(i) (ii)

Figure 5.2: (i) In ENT surgery, the microscope's stand is normally placed opposite to the surgeon, with the microscope arm extends over the patient's body. (ii) In this arrangement, the instrument is placed in the surgeon's line of sight, making the surgeon to 'look down'.

5.1.2 The behavioural model

A clear interface to operate any CAS system is of major importance, not only because of the general non-expertise of a majority of surgeons to operate computerised systems, but also because the CAS system should not distract users from the more important tasks to be performed in the operating theatre. To satisfy this condition, research into the complex relationship between the human-machine interfaces in the operating theatre must be conducted.

Apart from any technological limitations, the success of an interface is mainly determined from human performance during the operation. Human performance inside the operating theatre is mainly affected by organisational, communication and environmental factors [99]. It is desired that, when the new interface is implemented in the existing theatre set-up, these factors will improve or at least remain the same. In addition, it is expected that the surgeon would not have to modify his/her mental model while performing surgical tasks using the new technology. In order to achieve these conditions, the designer needs to create a behavioural model of the surgeon, microscope, and other members of staff during the surgical procedure.

5.2 User studies

Understanding users and their tasks is very important. The best method to accomplish this is by performing user studies. The latter provide a practical way to capture the physical and behavioural conditions of users in the current context. When conducting a user study it is important to choose representative tasks that provide reasonably complete coverage of a system's intended functionality. This can be done by selecting scenarios of microscopic surgery where the microscope is heavily used.

Two user studies were conducted during the "focus on user" cycle of the design process described in this thesis. The first study was *interview-based*, while the second was *video-based*. Types and duration of tasks were logged in both studies. Some information elements (motive and need) were identified before the first study, through informal interviewing of ENT specialists. The remaining elements and information primitives were created after the completion of the first study, and were used for observations in the second study. Aspect models and the surgeon's mental model were created after the completion of the second study, through a task analysis.

5.2.1 Interview-based study

Informal interviews were held with leading ENT surgeons before the first study. Speaking about the upcoming development proved to be most useful. Not only did it help the author to understand the real need for a new tool and the qualities it should bring to surgery, but also to envisage the operating space after the addition of the new tool. However, the only way to observe users in action is to attend microsurgical operations. Being inside the operating theatre allows the designer to realise the working conditions of the user, i.e. temperature, noise levels, physical posture, human psychology, etc. This gives the designer good information about the DIF context.

An activity study was then conducted at the Royal National ENT Hospital, Kings Cross, in order to assess the surgeon's tasks during the surgical procedure (pre-existing *mastoidectomy*). Observations were gathered over a four-hour session using designer field notes. A question and answer session was also conducted with the surgeon at the end of the surgical procedure to ensure good choice of design information primitives. Seven independent external events that occur during the critical stages of surgery were identified. These events and their weight[20] of occurrence during surgery are listed in Table 5.1.

Event list	Time
Performing surgery using the microscope	68%
Arranging the microscope's position	4%
Arranging the patient's position	3.5%
Observing the monitor or physiological monitoring equipment	4.5%
Communicating with other consultant or anaesthetist	4%
Teaching medical students	5%
Fixing surgical tools, communicating with nurse	11%

Table 5.1:Event recording at time of operation.

It was also noticed that events run in parallel with inputs ending at different time in the overall scheme (see Figure 5.3). Seven possible combinations of such parallel events are:

[20] The weight of a task represents its importance during the surgical procedure. This includes the amount of time spent on the task as well as the level of surgical intervention.

1. Burring or sucking, and viewing.
2. Communicating with consultant and viewing.
3. Teaching or training, and viewing.
4. Observing the 2D monitor, and teaching.
5. Arranging the patient's position, and teaching.
6. Moving the microscope and arranging the patient's position.
7. Fixing the surgical tools, and teaching.

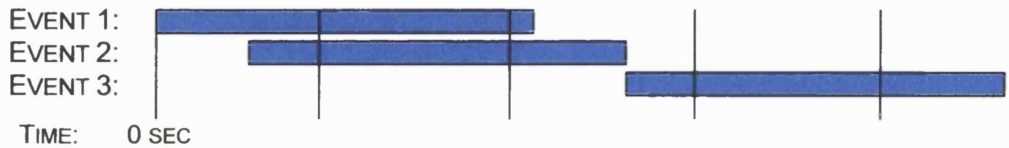


Figure 5.3: Parallel event recording.

5.2.2 In-situ video observations

Preliminary studies in task activity [78] suggested that observations needed to be video taped due to the large number of tasks taking place during the procedure, especially with some of them occurring in parallel. Moreover, video observations present the designer with the advantage of reviewing the procedure and make time measurements easier.

Two specific user scenarios were chosen for this second study. Each scenario reflects a surgical case where the usage of the new tool would be most advantageous: *cranio-facial resection* and *acoustic neuroma removal*. The two scenarios do not differ very much as far as microscope usage is concerned. They are both very stressful and lengthy operations. The first scenario was conducted at the Royal National ENT Hospital, Kings Cross, and the second at the Royal NHS Hospital, Belsize Park. The two operations were performed by different surgeons. Both users are top leading specialists in the field of ENT surgery, with over 20 years of surgical practice.

A total of seven hours of surgery was video taped. Three and half hours (12600 seconds) of the recording exhibit continuous surgery using the microscope (9000 seconds of acoustic neuroma removal and 3600 seconds of cranio-facial resection). The remaining time includes surgery without the microscope and idle-time[21]. A portable video camera was used for the recording sessions. The camera was placed at three metres away from the workspace. From that distance, all components shown in Figure 5.1 were contained within the camera's field of view. The Video Edit Information of both scenarios is shown below:

1. **Acoustic Neuroma removal**
   - Duration: 4 hrs
   - Total Operating Time: 3-½ hrs
   - Total Operating Time using Microscope 2-½ hrs (9000 seconds)
2. **Cranio-facial resection**
   - Total Duration: 5 hrs
   - Total Operating Time: 3-½ hrs
   - Total Operating Time using Microscope: 1 hr (3600 seconds)

Before observations were logged, some parameters that could make the task analysis more efficient without affecting the pattern of the results were defined. From the preliminary study, it was realised that events of very small duration are very difficult to log over long hours of recording. Consequently, the minimum recording time of an event was set to three seconds. In addition, extra attention was paid to events during the early stages of burring since it is one of the most delicate points of the operation and thought to be of particular importance in the way the procedure is managed[22].

5.3 Task analysis

An agent-centred task analysis was taken on to observe and understand real and representative tasks pertaining the use of the surgical ENT microscope (see Figure 5.4)

[21] Idle-time represents the time during which no tasks are performed.

[22] The introduction of an AR surgical system is most useful at times of burring since it offers the surgeon augmented stereoscopic views of the ROI. The virtual data are obtained from pre-operative CT scans of the skull region under burring. Burring effects are shown on both the real and virtual parts of the image.

in microscopic surgery. Two key objectives were identified: 1) understand the way the instrument is used during the operation; and 2) create design aspect models. Observations from both procedures identified nine main events, listed in Table 5.2. Some of the events were occurring in parallel, producing 14 main combined events. The most complex events combined up to four serial events in a single task, one of them always involving some use of the microscope. Combined events gathered from both studies are summarised in Table 5.3. Time measurements of serial events were made in the same way as in the preliminary study. Combined events were recorded as in Figure 5.3, with measurements beginning at the start of the first event and finishing at the end of the last event.

Figure 5.4: The Zeiss OPMI® ORL surgical microscope used in both video scenarios.

# Event number #	Event	Task
1	Repositioning	*Improving physical posture*
2	Teaching	*Transferring knowledge*
3	Communicating with junior surgeon	*Explaining steps of treatment*
4	Assisted by junior surgeon	*Operating in complex situation*
5	Assisted my nurse	*Changing operational tool*
6	Viewing microscope	*Obtaining visual feedback from ROI*
7	Arranging microscope	*Changing FOV*
8	Using surgical tools	*Performing treatment*
9	Fixing surgical tools	*Preparing next operational technique*

Table 5.2: Serial events and associated tasks.

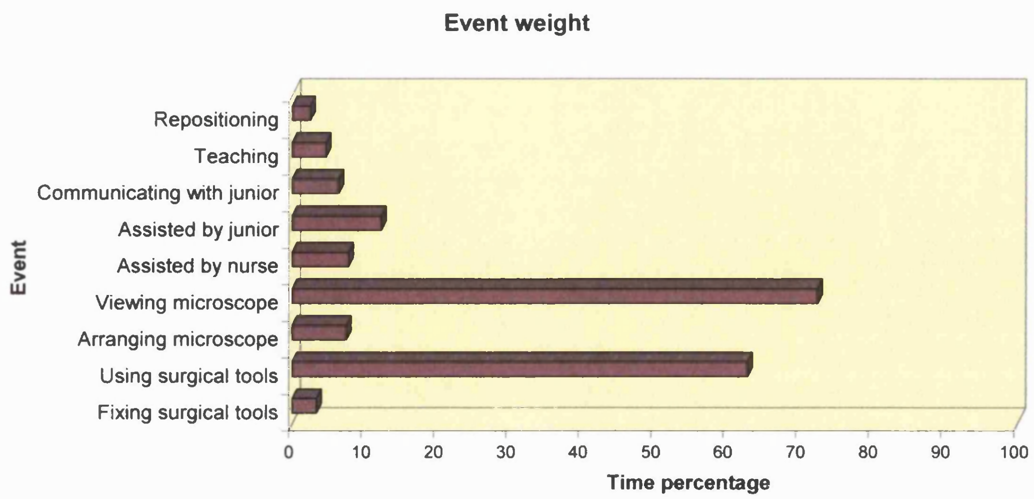
# Event letter #	Event
A	Communicating with junior-Teaching
B	Viewing microscope-Communicating with junior
C	Viewing microscope-Using tools
D	Viewing microscope-Assisted by nurse
E	Viewing microscope-Arranging microscope
F	Fixing tools-Communicating with junior
G	Fixing tools-Assisted by nurse
H	Viewing microscope-Using tools-Repositioning
I	Viewing microscope-Using tools-Teaching
J	Viewing microscope-Using tools-Assisted by junior
K	Viewing microscope-Using tools-Communicating with junior
L	Viewing microscope-Using tools-Communicating with junior-Teaching
M	Viewing microscope-Using tools-Assisted by junior-Teaching
N	Viewing microscope-Using tools-Assisted by junior-Communicating with the junior

Table 5.3: List of main combined events.

5.4 Results

The findings of the task analysis are presented in the following graphs. The graph in Figure 5.5 shows the percentage of time that an event uses up during the procedure, irrespective of whether it occurs in series or parallel. For example, the surgeon may be assisted by the nurse while looking through the microscope or after s/he has turned away from it. In this example, the "assisted by nurse" event occurs in both cases, and time measurements should account for its whole duration. By doing this, the designer can establish the weight of tasks during the surgical procedure. Figure 5.6 displays a time distribution diagram of the most frequent (atomic and combined) events during the same procedure[23]. Both graphs refer to 9000 seconds of acoustic neuroma removal operation. A separate graph (see Figure 5.7), similar to the one in Figure 5.6, shows how the pattern of frequent events is affected during the burring process of the procedure.

[23] In this analysis, only events that occur more than five times during the operation are considered as frequent.



	Fixing surgical tools	Using surgical tools	Arranging microscope	Viewing microscope	Assisted by nurse	Assisted by junior	Communicating with junior	Teaching	Repositioning
percentage	3.32	62.87	7.44	72.50	7.72	12.23	6.35	4.58	2.46

Figure 5.5: Percentage duration of serial events during acoustic neuroma removal.

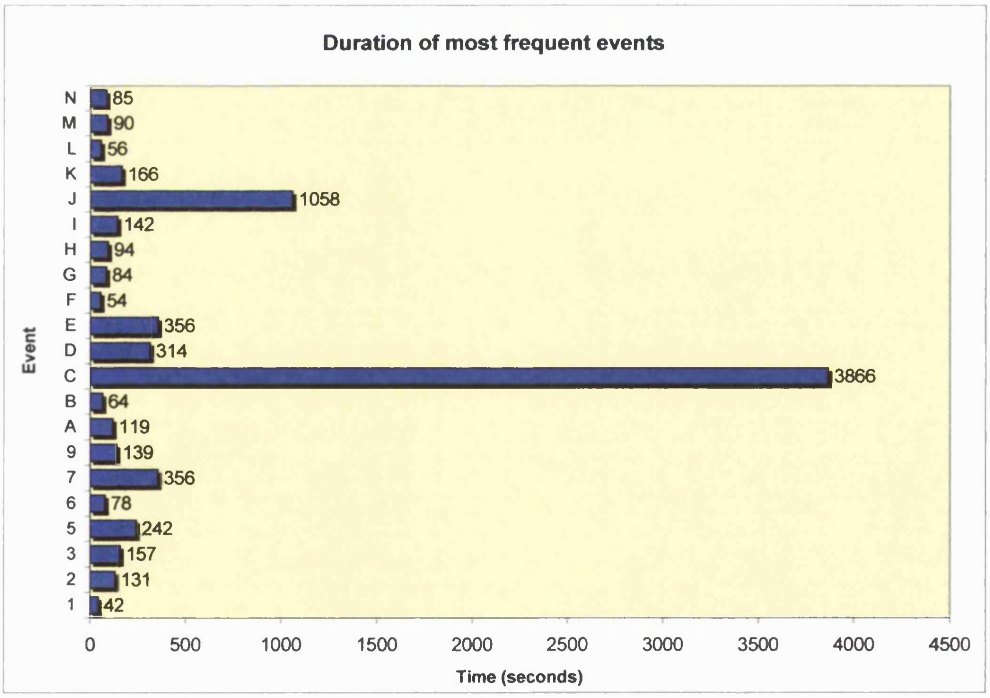


Figure 5.6: Time distribution diagram of most frequent events during 9000 seconds of acoustic neuroma removal.

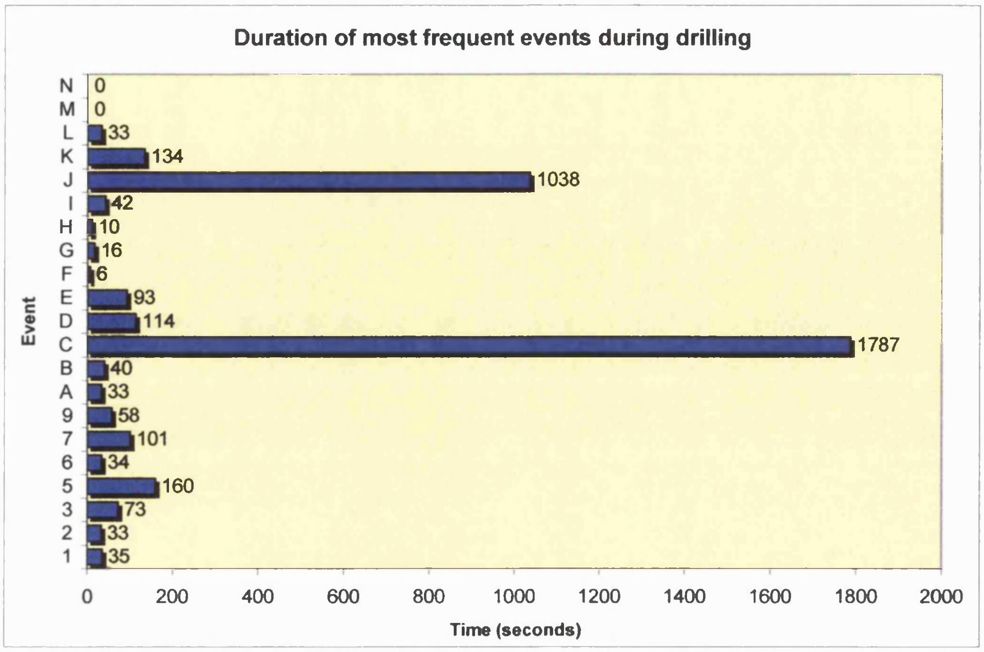


Figure 5.7: Time distribution diagram of most frequent events during the burring period of acoustic neuroma removal.

Results from the second scenario are presented in the following figures. Figure 5.8, like Figure 5.5, exhibits the weight of events during cranio-facial resection. Figure 5.9 is an activity-sampling diagram of the most frequent events that occur in the procedure. This is done by logging a count every time an event takes place during the microscopic surgery. By doing this, it is possible to display a spatial rather than temporal relationship between events and workspace in the domain of concern.

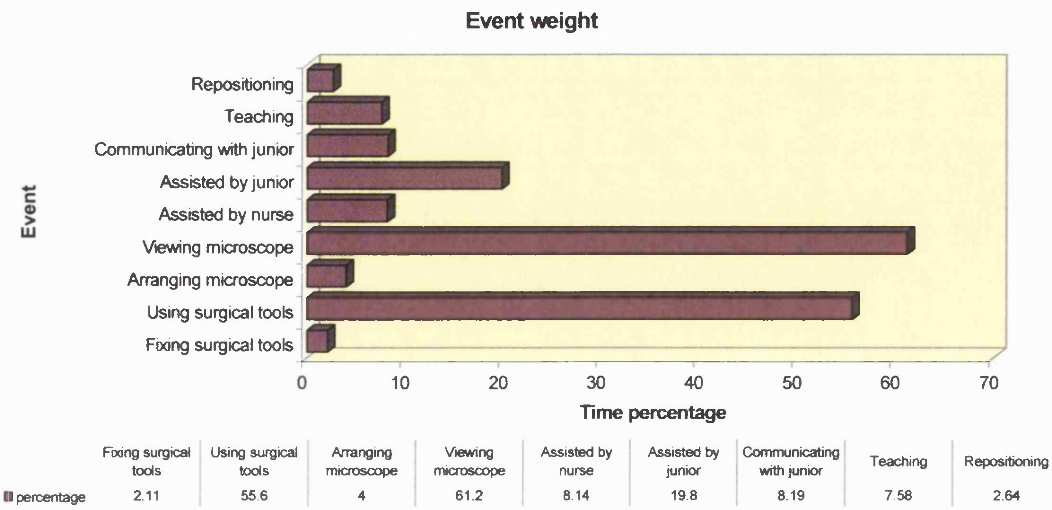


Figure 5.8: Percentage duration of serial events during cranio-facial resection.

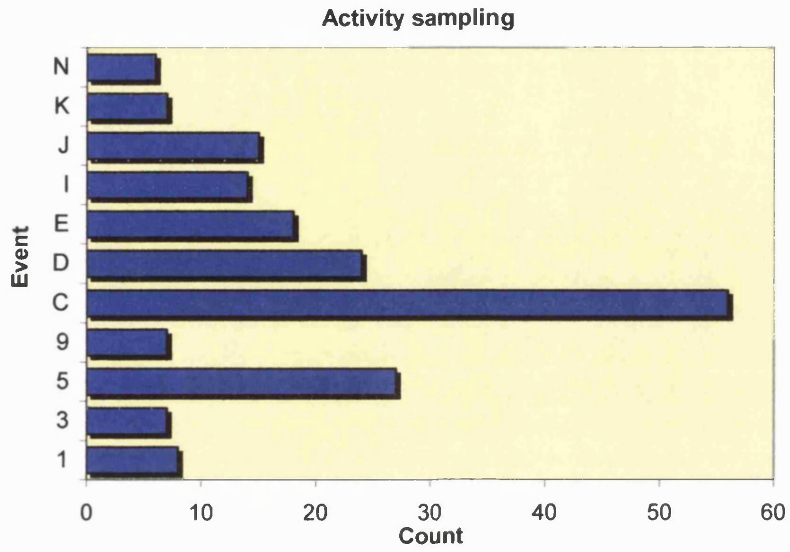


Figure 5.9: Frequency diagram of most common events during 3600 seconds of cranio-facial resection.

5.5 Discussion

While analysing the tasks, some general observations were made about the study. For example, it was found that the junior surgeon has a very active role during surgery. He has a 3D view that allows him to see the ROI while being immersed in the microscopic scene for long periods. It was also noticed that there is no significant patient head movement during the surgical procedure. This observation may prove to be very useful in the overall usage of the AR microsurgical system[24].

The two surgical scenarios studied in the "focus on user" cycle of the design process present tasks with similar weights. This can be seen clearly from direct comparison of the two event weight graphs (Figure 5.5 and Figure 5.8 respectively). Therefore, it was decided that one study should be used to derive the physical aspect model of the human-microscope interface and the other the behavioural model. Consequently, data from the cranio-facial resection were used to create the physical model while observations from surgical tasks during the burring period of acoustic neuroma removal were devise a behavioural aspect of the interface.

[24] Patient head movement was looked at in particular in the study, because the patient head apart from being at the centre of the domain of concern, is also associated with other technical issues of the CAESAR microsurgical system, *e.g.* choosing a tracking device to register the patient's position in the 3D virtual scene.

5.5.1 The physical aspect model

A physical aspect model of the surgeon-microscope interface was created by analysing observation data from a cranio-facial resection surgical intervention. The choice of this surgical scenario was because the particular operation presented very complex context: increased presence of clinical staff (clinical photographer) and a large number of students.The constructed model is a 3D working space, derived from design information primitives. By describing structurally fixed properties of entities with respect to the microscope during particular states of the agent, i.e. viewing through the microscope, one can define the main theatre locations of advanced human-machine interaction. Since the surgeon needs to maintain the current kinesiology of surgical tasks, the new viewing interface has to be designed in a manner which ensures that activity locations remain unaffected.

Activity locations of the operation can be derived from the findings presented in Figure 5.9. By classifying the most common events into *microscope-affected* (lettered) and *neutral* (numbered), specific locations of microscope-affected activities can be identified by the number of counts. The counts can be used to record the physical location of events in the domain of concern. The result is a plot of lines, areas and numbers which define the instrument's working space. Seven microscope-affected events while in the state of viewing through the instrument's eyepieces were identified: using tools (56 counts), assisted by nurse (24 counts), arranging microscope (18 counts), using tools-teaching (14 counts), using tools-assisted by junior (15 counts), using tools-communicating with junior (7 counts) and using tools-assisted by junior-communicating with junior (6 counts). The plot diagram of the physical space is shown in Figure 5.10.

Particular importance was paid to the movement characteristics of the microscope during the operation. From observing "arranging microscope" events, it was found that the instrument moves with four degrees of freedom, listed below.

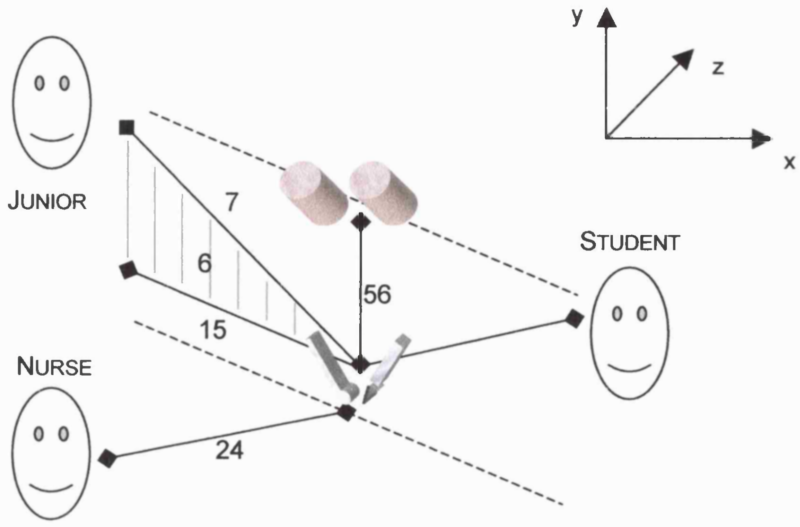


Figure 5.10: Event plot diagram of most microscope-affected events in cranio-facial resection.

- MICROSCOPE MOVEMENT CHARACTERISTICS -

- Translation left and right, from surgeon's face ±10 cm,
- Translation up and down, ±5 to ±10 cm,
- Translation in and out, ±5 cm,
- Rotation of eyepiece around the x-axis, ±30 to ±45 degrees,

Microscope movement characteristics, along with the 3D working space described in Figure 5.10, provide physical attributes of the conventional interface in the current context of microscopic surgery. The values of these attributes provide essential information elements when designing the new viewing interface.

5.5.2 The behavioural aspect model

The behavioural model of the surgeon-microscope interface was built from observations during the burring process of the acoustic neuroma removal intervention. The importance of this process has been described earlier in this chapter (see §5.1.2).

Observations showed that most of the junior surgeon's assistance occurs during the burring stages of the operation. The junior is situated very close to the centre of the domain of concern, actively participating in the procedure. The usually long period of burring (up to 340 seconds) is followed by immediate assistance from the nurse (3-10 seconds), and does not always re-start before the microscope is re-arranged and re-focused. The latter event, which is most important in implementing the new viewing technology, takes on average eight seconds. Sometimes, in between the steps of the above period, other events take place such as communicating or repositioning. This happens only for a few seconds and seems to be mostly associated with the surgeon's psychology. The above can be verified from Figure 5.11 when someone adds the duration of the seven longest events. The figure also reveals a task periodicity, at the following period:

Using tools-Assisted by the surgeon / Assisted by the nurse / Arranging microscope

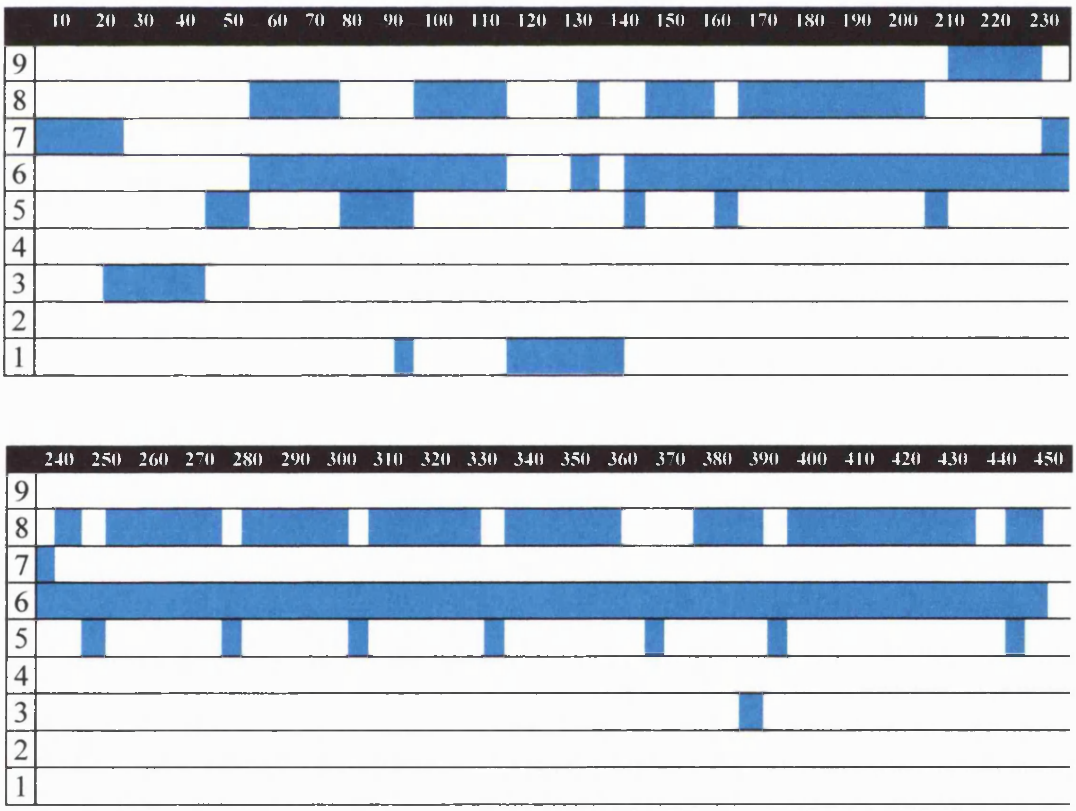


Figure 5.11: Sequence of serial events during the early stages of burring (total time: 450 seconds).

The parallel event recording results in the figure above display the user's behavioural pattern during the delicate process of temporal bone burring. Task periodicity provides another characteristic of this information primitive that can be used in the design process of the new interface.

CHAPTER 6

DESIGN CYCLE II: EARLY AND CONTINUOUS TESTING

One of the fundamental concepts of good HCI in systems development is **usability**. Usability is concerned with making systems easy to use and can be divided into four components [100]: *learnability*; the time and effort required to reach a specified level of user performance, *throughput*; task accomplishment by experienced users, speed of task execution and errors made, *flexibility*; the extent to which the system can accommodate changes to the tasks and environments beyond those specified in the "focus-on-user" cycle and, *attitude*; the positive feedback engendered from users. The above components were studied carefully during the second cycle of the design process described in this thesis.

Early and continuous testing of the new 3D viewing technology was divided in two parts: prototype development and evaluation. During the first part, scientific research was conducted in stereo vision and microscope optics, and more specifically in the extraction of stereoscopic, optical images from the surgical microscope and their efficient display on a prototype autostereoscopic 3D display (courtesy of Sharp Laboratories Europe Ltd.). The main research issues examined by the author included choice of video camera system for binocular imaging, design and manufacturing of special microscope-coupling components and selection of optical devices for usage with the surgical microscope. In the second part, the usability of the prototype 3D microscope as an alternative to the eyepieces of the binocular operating stereo microscope was tested through a task analysis evaluation [101].

6.1 Prototype development

The conventional design of the surgical microscope uses a beam splitter for diverting the optical ray to the binoculars and cameras. Still, the beam exiting the beam splitter is not in focus. In reality, zoomed rays are focused at infinity. In this case, the total focal length is the distance between the lens and the camera plane when the object is at infinity. For photographic lenses, the objects are usually far away; so all images are formed in nearly the same plane.

To attach a video camera at large distances would be unacceptable, as it would increase the size of the microscope dramatically. Therefore, the first aim was to obtain a lens system that would set the focal point to a maximum accepted distance of 60-120 mm (see Figure 6.1). When trying to obtain the shortest possible focal point one has to be very careful in doing it without causing uneven colour parts on the screen, *e.g.* white shading. This effect occurs because the exit pupil is not sufficiently large[25].

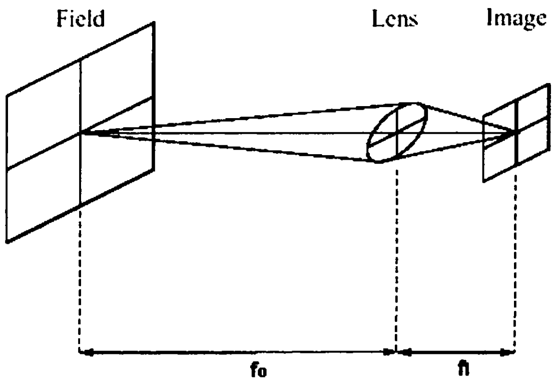


Figure 6.1: Basic components of an optical system. Light is reflected from the field to a lens at a distance f_o away. This light is then converged to focus on the image plane at a distance of f_i from the lens. The field plane is the slice of the real world that is to be imaged. The lens array is used to increase or decrease the magnification of the scene. The image plane represents the point of focus of the magnified scene and the location of the CCD camera.

Another constraint was the fact that the lens could not be greater than 30 mm in diameter as this is the standard Zeiss microscope beam splitter diameter used in most surgical microscopes.

In the pursuit of such an optical system, the author had to take into consideration other limiting factors such as the compatibility of the lens with the CCDs. Sometimes, the adaptation of a lens to CCD is possible only after the use of IR filters. However, the use of IR filters and protective glasses decrease spherical aberrations and increase image sharpness.

An alternative solution to the above problem was the use of a video objective lens (also called TV objective). The TV objective replaces the camera's lens. When attached onto the side arm of the beam splitter, the objective acts as a coupling medium between the light beam in the splitter and the camera's CCD sensing area. The exit pupil of the lens is far in front of the entrance pupil (Figure 6.2). In this way, a long focal length lens can fit into a short package.

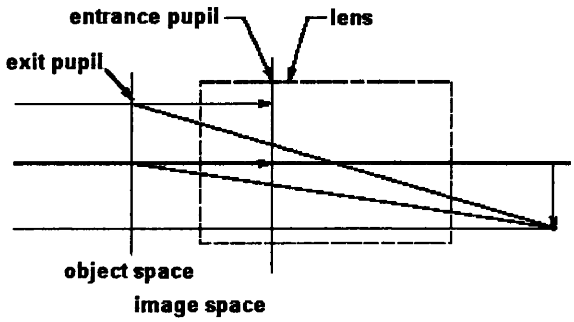


Figure 6.2: Optical geometry of telephoto lenses.

6.1.1 Design Characteristics

The system under development seeks to improve the use of the binocular operating microscope during surgery. In the envisaged set-up the ROI is captured through the microscope by two industrial camera modules which view the same images as the microscope's eyepieces, thus producing a stereo pair of left and right images of the patient. The stereo pair is displayed in a 3D display that replaces the binocular operating microscope eyepieces.

A commercial Video Objective lens, with a focal length of 60 mm, was used. The lens has manual fine focus tuning and a bayonet connection at the camera end. The chosen video objective was not built for the application of mounting a commercial video camera. Even though commercial couplers that can fit on top of a bayonet video objective lens with a C-mount[26] camera attached at the other end exist, the author

[25] The exit pupil refers to the virtual image of the diaphragm formed by the lens. When the exit pupil is at infinity, all principal rays are parallel to the optical axis.

[26] The mechanical definition of a C-mount is a 1-inch hole with 32 TPI threading (turns per inch), female on the camera side, male on the microscope side. The optical definition of a C-mount is that the image reaches the focal plane of the CCD at 17.5 mm beyond the shoulder of mounting ring.

needed a coupler with adjustable focusing positions (see Figure 6.3) and with the ability to offer a small degree of movement on the translation plane of the camera.

Normally, the focusing of a video camera attached onto a TV objective is done by viewing the video image from the camera, and adjusting the microscope focus until the image appears in focus. Framing and exposure metering are also done by viewing the video image. The views taken are full frame, rectangular images, and not just bright circular images with dark corners. Furthermore, even after focusing the cameras individually, it is impossible to find two cameras of the same make and type that will have the sensing area of the CCDs at exactly the same position. Nevertheless, a slight difference of one millimetre in position may lead to a distortion of 20 mm between the two final images [31]. This effect makes the matching of the two monoscopic images very difficult.

Finally, the correct alignment of the two cameras on the autostereoscopic 3D display prototype is not a case of just placing the cameras sideways of the microscope and requires prior calibration (see §6.2.1.2).

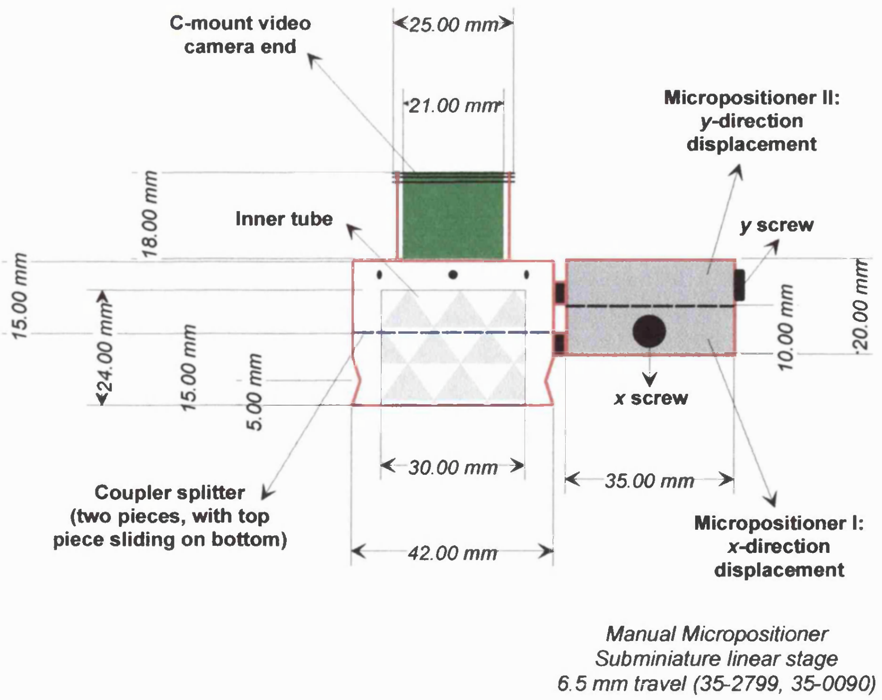


Figure 6.3: Side view of coupler's original design using a commercial micro-positioning stage.

In an attempt to find a universal solution to all the above problems a prototype coupler was designed and manufactured in such form that could give a free *x-y* and *z* direction travelling distance of up to ±5 mm and ±25 mm respectively. The *x-y* translation would solve the fine adjustment of the two monoscopic images on the screen, while the *z* direction would accurately find the best zooming position of the two cameras before or after the focal point[27]. The prototype could meet all these demands, but could not offer a complete locking of the whole system at the desired position due to the material choice of locking screws.

The author designed and manufactured a new coupler that would have the ability to lock at any desired position within the 10 mm gap (as illustrated in Figure 6.4). The new product also presented the feature of holding its zooming position steady for camera weights of less than 500 gr. The coupler itself has negligible weight and does not affect the microscope's overall shape and weight. The overall look of the set-up is shown in Figure 6.5. This design allows alignment of the monoscopic views from the two video cameras on the image plane of the 3D display, a concern throughout the experimentation.

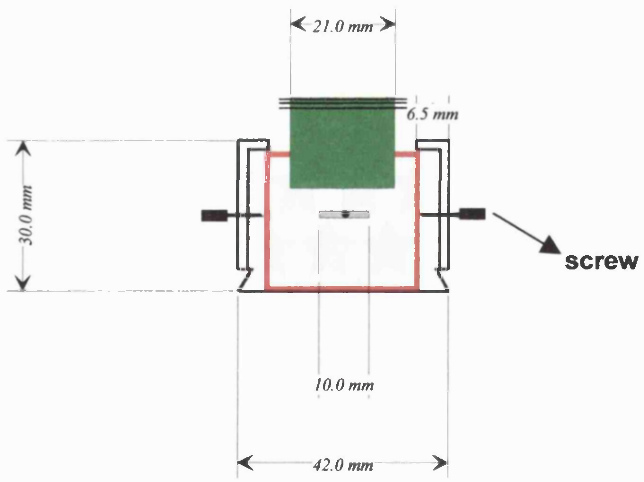


Figure 6.4: Original design of the video objective lens-camera coupler using three screws, placed 120^{O} apart.

[27] The image formed after the focal point is an inverted mirror image of the source. The one before the focal point is not mirrored and of smaller scale.

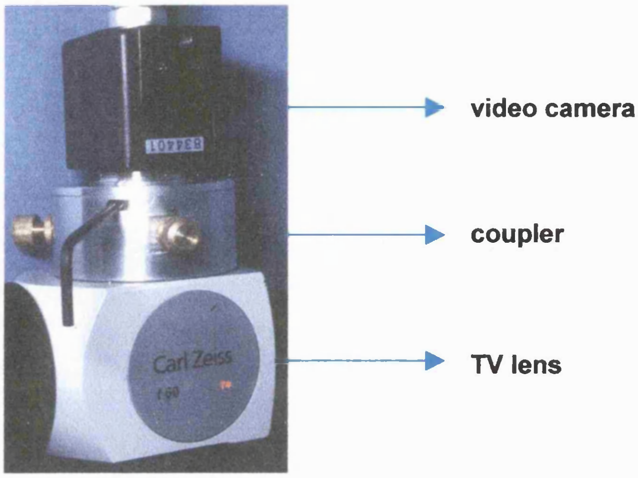


Figure 6.5: Prototype coupler mounted on a Zeiss video objective adapter.

6.1.2 Video camera selection

The choice of an industrial camera for the purposes of the system was not a simple task. The basic parameters were *size*, *weight*, *image resolution*, *colour* presentation and *cost-effectiveness*. Since the global market offers a substantial number of products, the first priority was to obtain information for all brands and models, examine the above parameters and compare them. The first step was to run a survey on cameras that use CCD technology. In particular, the main interest was based on three-chip cameras because of their separate Red, Green and Blue sensing capabilities. This implied better colour quality (as explained in §3.2.2). For the next part of the survey, cameras that would give a serial digital output were found. This was also quite important since the computer's video card only accepts serial digital video signals. The last phase of the market survey involved aspects like size and weight. The primary concern was to find a light and compact camera, preferably one with a remote head in order to have the optics separated from the electronic circuits. A comparison between cameras was undertaken. Compared models are listed in the Table 6.1.

Some demonstrations were arranged for a number of these products. The actual demonstrations took place in the laboratory with the products being evaluated on the prototype 3D microscope. During the demonstration, the author attempted to evaluate factors such as image quality and resolution. This was done by comparing pictures from the artificial skull shown in Figure 6.6, captured by different video cameras.

Camera Model	CCD Type	Horizontal Resolution	Colour, Video Output	Dimensions (W)x(H)x(D) mm	Weight (gr.)	Price (£)*
DAGE DC-330	3 x 1/3" Remote head	570 lines RGB	NTSC/PAL RGB, Y/C	Camera Head 32 x 41 x 40	130	6,245
Toshiba IK-TU40A	3 x 1/3" Remote head	570 lines RGB	NTSC/PAL RGB, Y/C	Camera Head 32 x 40 x 40	60	4,800
Panasonic GP-US532	3 x 1/3" Remote head	572 lines RGB	NTSC/PAL RGB, Y/C	Camera Head 34 x 44 x 52	110	2,800
Panasonic GP-US522	3 x 1/2" Remote head	572 lines RGB	NTSC/PAL RGB, Y/C	Camera Head 34 x 44 x 52	110	3,250
Sony DXC760M	3 x 2/3" Remote head	570 lines RGB	NTSC RGB, Y/C	Camera Head 70 x 75 x 113.5	600	7,700
Sony XC-003P	3 x 1/3"	570 lines RGB	NTSC/PAL RGB, Y/C	50 x 56 x 128	440	2,835
JAI M90	3 x 1/3"	570 lines RGB	NTSC/PAL RGB, Y/C	50 x 60 x 130	480	1,650
Hitachi HV-22	3 x 1/2"	580 lines RGB	NTSC/PAL RGB, Y/C	65 x 65 x 130	600	2,300
JVC KY-F55BE	3 x 1/3"	580 lines RGB	NTSC/PAL RGB, Y/C	64 x 70 x 133	490	2,595

* AS PRICED IN MARCH 2000, (EXC. VAT)

Table 6.1: Survey results on choice of industrial video camera modules.

Figure 6.6: Optical microscopic image acquired from the Panasonic GP-US522 video camera.

At the end of the survey, only two cameras, the Panasonic GP-US522 and the DAGE DC-330, satisfied all selection parameters. Both products have just been released in the market and they represent the best advances of technology in the field of video equipment for industrial applications. Considering the project's budget and the cost of each camera model it was decided to purchase the **Panasonic GP-US522**.

One of the limiting factors in selecting a camera according to the CAS system demands was the fact that there is no commercial digital video camera for medical applications that uses 3-CCD technology. Still, the chosen camera presents the ability of offering an internal 10-bit digital image processing, prior to any software image processing. This is done through the electronic circuitry inside the Camera Central Unit (CCU).

6.2 Evaluation of the 3D-microscope display system

To test the effectiveness of the prototype Sharp Micro-Optic Twin-LCD monitor as an alternative to the eyepieces of the binocular operating stereo microscope, a task analysis evaluation test was devised. To conduct the experiment, a passive robot arm [102] was used to identify and mark three identifiable points of a miniature model placed under the microscope (as shown in Figure 6.8). Seven test subjects (including the author, the rest being all ENT surgeons) were chosen to carry out the experiment. The choice of examiners was made on criteria such as ease of microscope use and understanding of the visual information as it appears on the 3D autostereoscopic display.

A second series of experiments looked into the timed response of the observers on the task of passing a thread through the eye of a needle. In both experiments, the microscopic image was viewed through both the prototype viewing interface and the eyepieces. All experiments were conducted under a scale 10× and 16× magnification factor. The experimental apparatus is displayed in Figure 6.7.

Figure 6.7: Experimental arrangement of the equipment.

6.2.1 Experimental set-up

6.2.1.1 Use of 3D mechanical arm for digitising positions

The passive robot arm used in the experiment is a multiple axis-articulated arm with a spherical working diameter of 2.4 metres. It has a 0.25-inch ball probe attached to its working end and a 1-inch reference sphere placed at its base. The mechanical arm is counterbalanced and temperature compensated, with six degrees of freedom. It has optical encoders placed at each of six joints which, when combined, provide complete point position (x-y-z) and orientation (i-j-k). Three-dimensional digitised measurements are taken between the distance of the ball probe and the origin of the reference sphere. It operates using a standard PC and can be used as a stand-alone 3D-measurement system. The 3D mechanical arm has an accuracy of ±0.070 mm and has a calibration procedure to maintain correct operation.

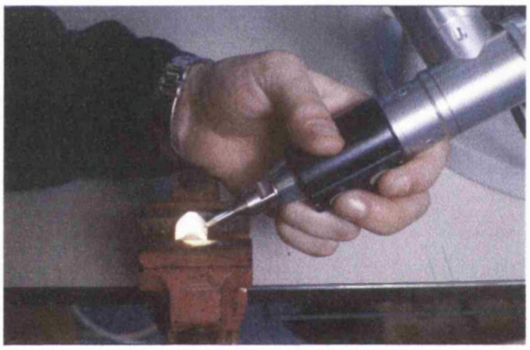


Figure 6.8: Miniature model measurements using the passive robot arm.

6.2.1.2 Calibrating the system

Calibration of the 3D microscope-display-arm system is carried out manually before the start of the experiment.

Initially, the focusing level of each camera is adjusted by placing a calibration grid (see Figure 6.9) under the microscope. The cameras are alternatively focused to the grid's surface. Therefore, both cameras have the same focusing levels. Fine focusing can be achieved by increasing the microscope's magnification factor (as explained in Chapter 3).

The next calibration step involves the alignment of the two monoscopic images on the autostereoscopic 3D display. This is done with the use of a cross-hair grid that is situated at the focal plane of the microscope's objective lens. The grid has a diameter of 35 mm. This ensures that the region illuminated by the microscope's light source is covered when the CCD capturing area exceeds the one viewed through the binoculars (as shown in Figure 7.3a). Next, the left and right displayed images are centred by superimposing the grid. A coarse alignment is accomplished at low magnification factors (6×) using x-y translation of the camera coupler and then fine alignment is achieved at higher magnification factors (16×, 25× and 40×). Correct alignment of the two monoscopic images is achieved only when they overlay each other in the display plane. When alignment is complete, the two couplers are locked at this fixed position.

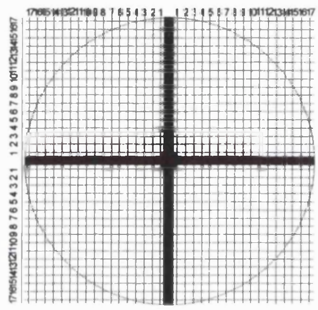


Figure 6.9: Calibration grid (numbers denote mm distance from the grid's origin).

Finally, the passive robot arm is calibrated. This is done by measuring 27 points on the circumference of the 1-inch sphere with the arm's probe. The x-y-z co-ordinates of the probe location are displayed on the PC screen [103].

6.2.2 Experimental protocol

Each operator starts the experiment by trying to identify three marked points on the model under the microscope using the autostereoscopic 3D display. He then attempts to bring the arm's probe in the focusing region of the microscope. The probe is not expected to make contact with the designated points on the object. This ensures that the positioning of the probe is driven by visual feedback and not by a motor reflex due to touching, known as haptic feedback. To correctly register a 'hit', the x-y-z position is recorded by pressing a button on the probe's handle. In addition, each time the operator registers a point he then must remove the probe from the microscope's viewing region before carrying on with the next one. This prevents the user from getting to same position via 'memory' effects. This procedure is repeated five times for each of the designated points. The whole exercise is then repeated again using the microscope eyepieces only (as demonstrated in Figure 6.10). Finally, without the aid of any magnification device, the exercise is done by means of free viewing of the object. During free viewing attempts, operators are allowed to use a large proportion of haptic feedback. The latter ensures that operators use additional sensory information about the point's location in the 3D space. Free viewing values are considered as '*golden standard*' values throughout the experiment.

(i) (ii)

Figure 6.10: Operator performing the experiment using (i) the Sharp 3D autostereoscopic display and (ii) the microscope eyepieces.

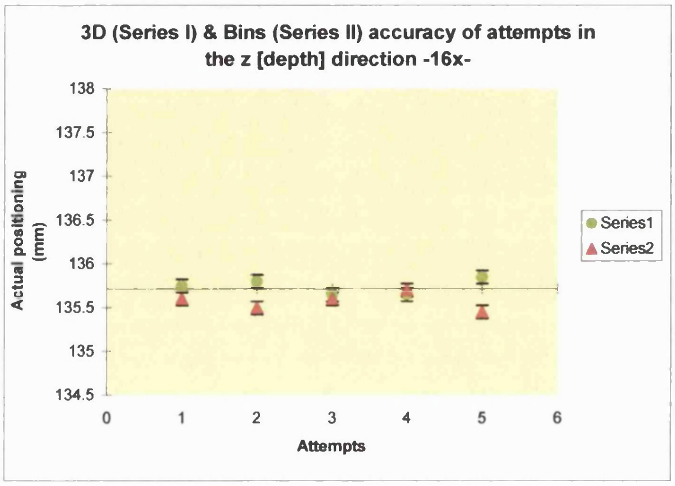
The second series of experiments evaluates the time response of each operator when asked to carry out the simple task of passing a thread through the eye of a needle while looking through the eyepieces, with the magnification factor being 16×. This typical experiment takes place at the early stages of microsurgical training. The same experiment is repeated with the examiner looking through the autostereoscopic 3D display and finally, a standard monitor (i.e. 2D display by switching off one of the image channels).

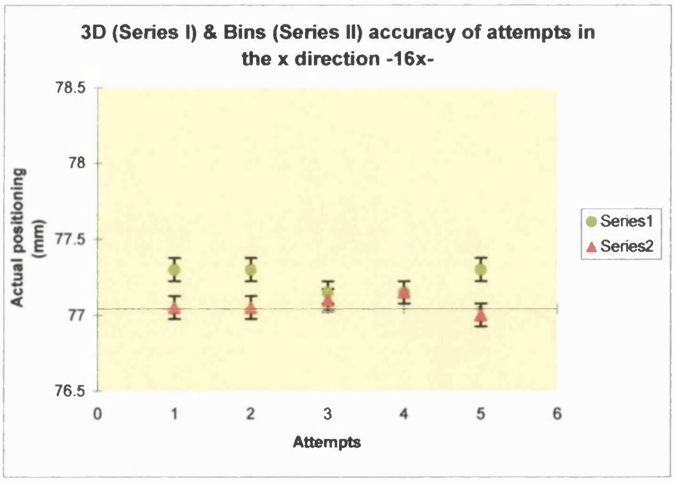
6.2.3 Results

The results showed that there is a correlation between the display and the eyepieces in the x, y and z directions for all operators[28]. The following graphs (see Figure 6.11 and Figure 6.12) display this correlation for one operator attempting to identify the second target point on the model, when using a magnification factor of -16×- and -10×-. Other

[28] By using the term correlation here, the author means that there is a constant tendency for each operator to deviate from the '*golden standard*' value in the same direction when he uses either the 3D display or the microscope's eyepieces.

graphs for the remaining operators, which are not shown here, exhibit very similar patterns. Raw data are shown in Appendix A.





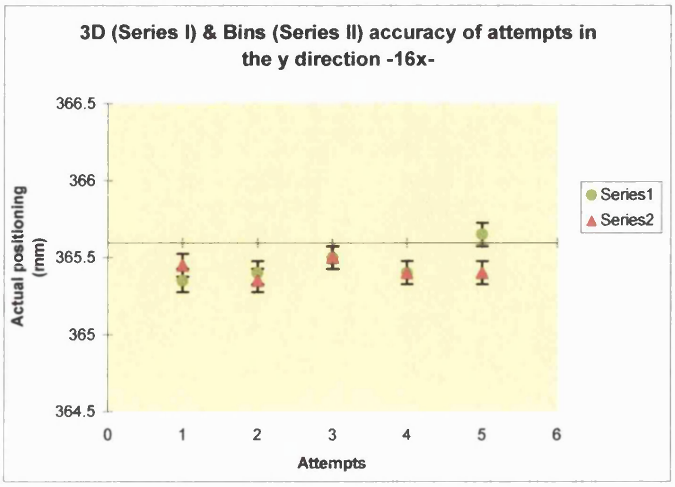
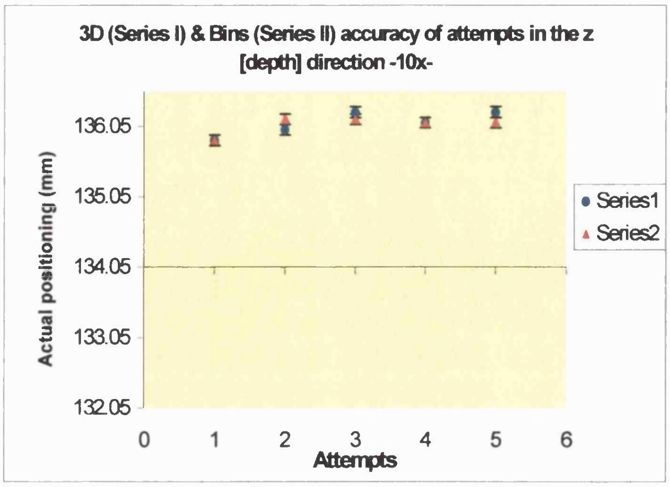
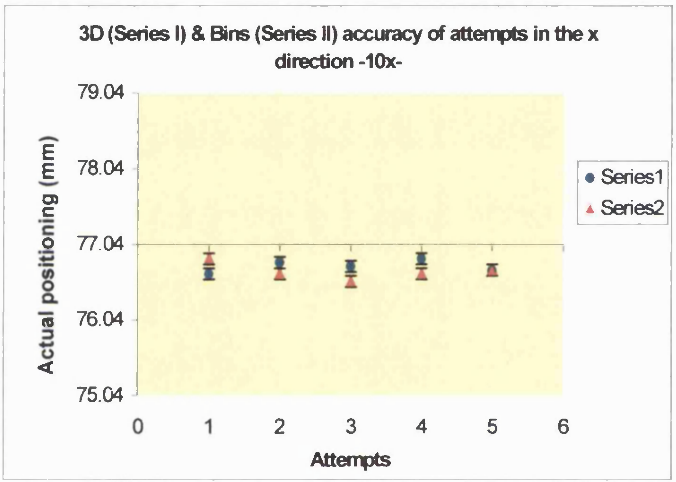


*Figure 6.11: Axial deviations from the actual positioning of the second point for a single operator when using the display (Series 1) and the binoculars (Series 2). Magnification factor is **16**x.*





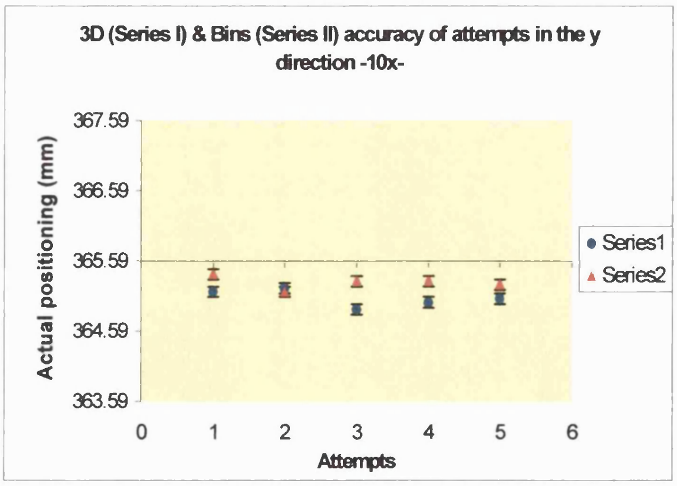


*Figure 6.12: Axial deviations from the actual positioning of the second point for a single operator when using the display (Series 1) and the binoculars (Series 2). Magnification factor is **10**×.*

The experiment uses the free viewing value as the 'gold standard' from which the other two measuring techniques are compared. In the charts for x direction above, 77.041 mm is the gold standard while Series 1 represents the deviation using the autostereoscopic display and Series 2 represents the deviation under the binocular eyepieces of the microscope. This '*gold standard*' value is used as a reference value throughout the experiment. Likewise, similar gold values are used for the charts in the y and z directions.

The above graphs show little differences between the 3D display and eyepieces measurements for the individual's attempts in identifying the target point. The following graph (see Figure 6.13) shows the overall correlation between each pair of results for all operators. This suggests that there is a matching trend in visual perception of the Sharp autostereoscopic 3D display to that achieved by the stereo microscope eyepieces. Overall, performance was estimated through averaging of all operators' attempts. The overall x, y and z errors for all operators hitting the second marked point at the magnification level of **16**× are -0.061 mm, 0.022 mm and 0.050 mm respectively. The overall errors for **10**× are –0.015 mm, -0.062 mm and –0.082 mm respectively. This implies good accuracy in readings.

Moreover, from Figure 6.13, one can see that not only are there little differences between the 3D display and eyepieces measurements for the operators' attempts in 'hitting' the target point, but also that the 'hits' are marginally different in between them. This is demonstrated through the fact that different operators may identify a point at different positions –as in attempts 17 and 32–, but every operator tends to identify the point always at the same distance from the '*gold standard*' –as in attempts 1 to 5 and 6 to 10– when using both display media. The latter implies high precision in readings.

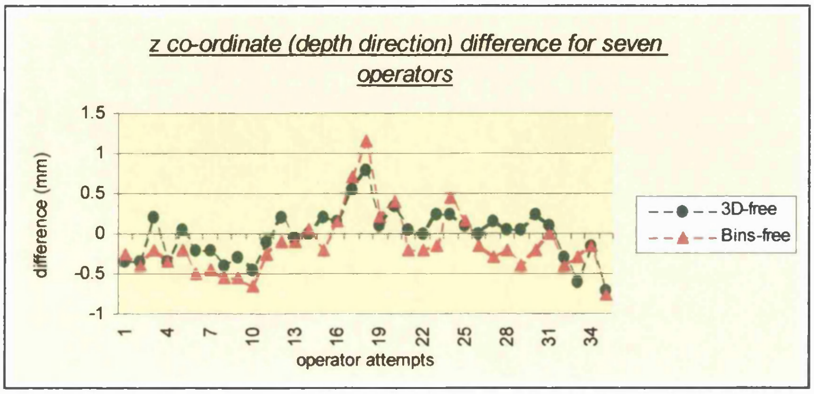


*Figure 6.13: Correlation pattern between the autostereoscopic 3D display and the eyepieces of the microscope. Magnification factor is **16**×.*

The time-response measurements show clearly the importance of depth information when the operator works in 3D space. Additionally, it is noticed that the viewers response when using the autostereoscopic 3D display is very similar to when they use the microscope eyepieces. Finally, by looking at the following graph (Figure 6.14), one can see that the task is always easier (or as easy) when the operator uses the binoculars. It is believed that this results from the fact that most of the operators have been previously trained to use the microscope.

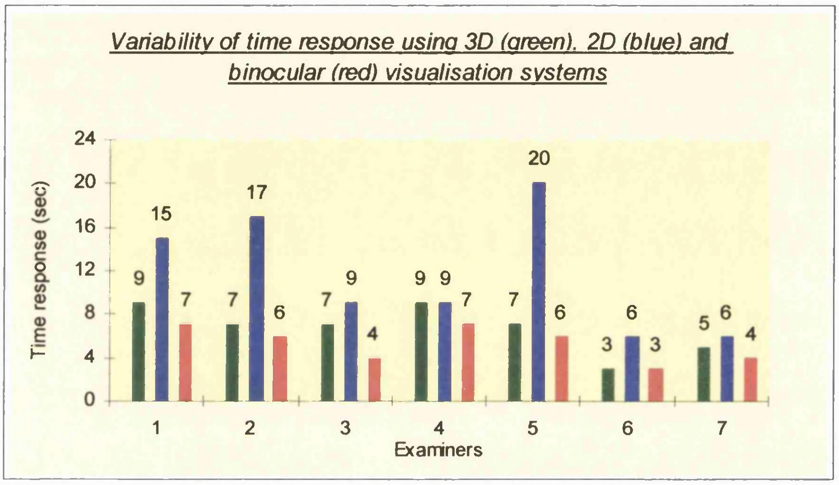
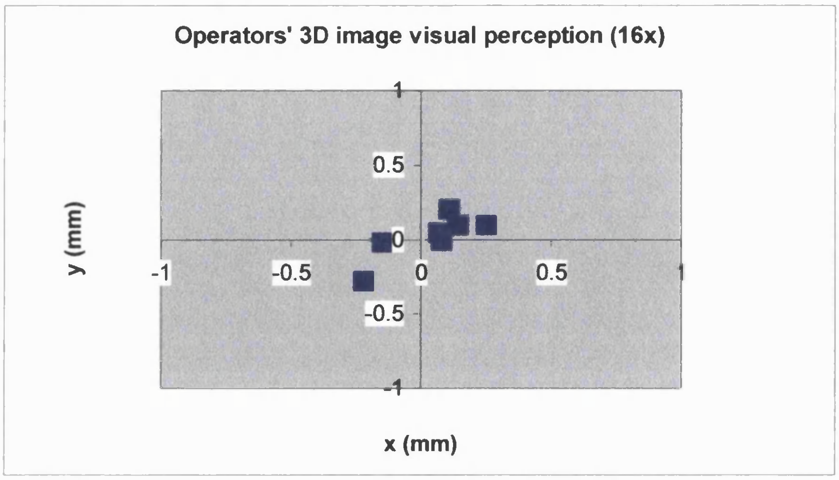


Figure 6.14: Graphical representation of results from the time study.

6.2.4 Error analysis

6.2.4.1 Statistical error analysis

Statistical errors are due to sample selection. The author examined the image perception variance between operators when they use the 3D display or the microscope eyepieces. Figure 6.15 shows the mean values of the difference in binoculars and 3D free viewing (y-axis) plotted against the difference in autostereoscopic 3D display and 3D free viewing (x-axis). Each square plotted refers to a different observer. Ideally, all squares should lie very close to the origin.



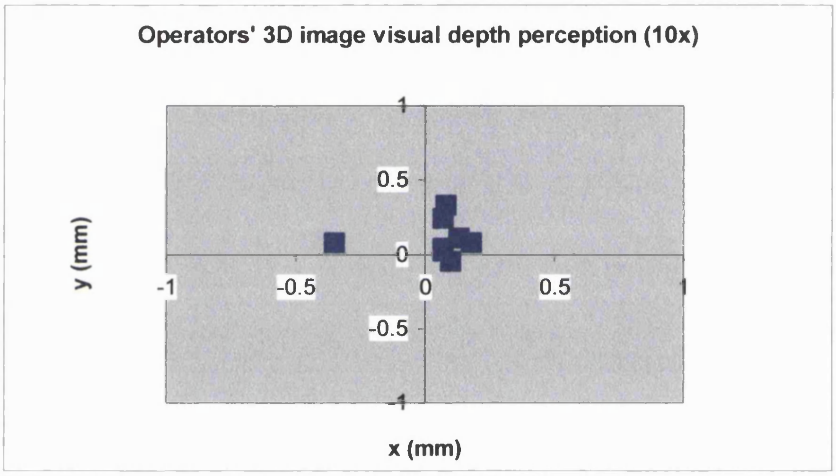


Figure 6.15: Depth perception variation for all operators.

Figure 6.15 clearly illustrates the way all operators view depth information from both visualisation media. Note that position mean differences are very small with all values lying very close to zero, thus implying only small discrepancies between both the Sharp 3D autostereoscopic display and the binocular stereo microscope with respect to the actual position for each operator. The line along which the points are clustered demonstrates that there is a very small difference indeed between the 3D autostereoscopic display and the binoculars depth perception for each operator in spite of variations in operator bias. Ideally, all points should lie at the origin. The data pattern confirms the fact that different surgeons tend to interpret the image differently. This finding suggests future research into the quality of depth percept when using the autostereoscopic display.

6.2.4.2 Systematic error analysis

The active window of our prototype Micro-optic twin-LCD monitor was placed at a distance of 270 mm from the back LCD panel. Every operator was positioned at 270 mm from the window plane (see Figure 6.16). At this position, the specific prototype offers optimum three-dimensional visual information. The average interpupillary distance of the operators is assumed to be at the normal average of 60 mm. The display's depth (or longitudinal) resolution can be determined from the binocular viewing geometry using the divergences of the LCD light rays through the active window. For the display system discussed in this thesis, this was found to be 0.16 mm. This error is comparable to the 0.12 mm error in the operator-response graph of depth perception.

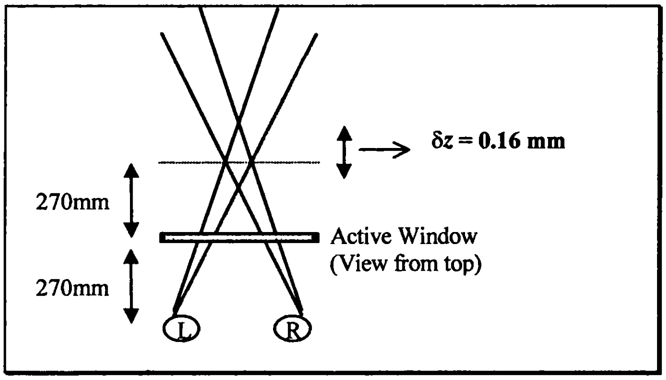


Figure 6.16: Depth variations in the viewer-display space.

6.3 Discussion

6.3.1 Early testing

The early testing period of the second cycle in the proposed design process involved the development of a prototype stereo camera system that can extract microscopic stereo video images from a conventional surgical microscope and display them in an autostereoscopic 3D display. That was done by designing and manufacturing a mechanical coupling component that mounts a video camera to the body of commercially available video objective adapter with bayonet mounting. The optical arrangement of the TV objective-coupler (mounted directly into each side of the Zeiss beam splitter, as shown in Figure 6.17) favours a **parallel-axes** stereo video camera system of binocular imaging. Furthermore, this allows the user to use the binocular eyepieces if there is a need to switch back to the conventional viewing technique. This architecture ensures that the prototype has very good *flexibility*, and can be easily integrated with any standard surgical microscope apparatus.

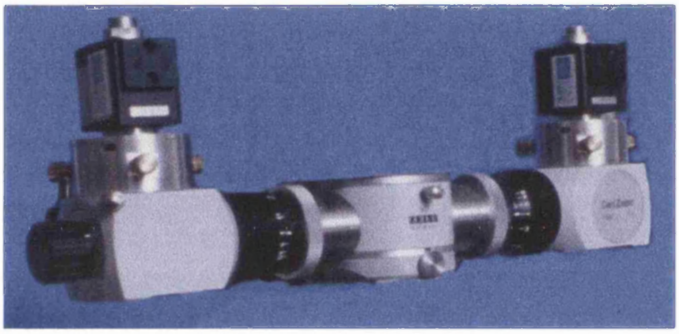


Figure 6.17: The stereo camera system can be installed quickly at any Zeiss ENT microscope. For other commercial ENT microscopes, a different beam splitter must be used.

The ability of the coupler to move the camera head in the x-y plane as well as in the z direction allows good alignment and focusing of stereoscopic images from different cameras. Camera translation could also be useful for controlling the amount of depth in stereoscopic image pairs – for example, through image cropping – a technique that is explained during the third cycle of the design process.

6.3.2 Continuous testing

The evaluation results showed that there's a correlation between the viewing properties of the Sharp Micro-optic twin-LCD autostereoscopic display and the eyepieces of the surgical microscope. Great emphasis is given to the perception of depth, as this is very important to the acting ENT surgeon while s/he performs a surgical operation. Equally, the results establish the accuracy and precision in viewing 3D objects using the new 'heads up' display. A paired t-test sample confirmed that there was no significant difference between the expected and mean measured values for both 10× and 16× magnification factors (see Table 4.2). Depth recognition measurements of a microscopic model result in deviations not more than 1 mm, the average surgeon's hand movement precision.

Conditional Parameters for a sample of n=35	*Magnification –x10- (mm)*	*Magnification –x16- (mm)*
x-direction actual error	-0.015	-0.061
y-direction actual error	-0.082	0.050
z-direction actual error	-0.062	0.022
Statistical t for x-direction	0.40	0.95
Statistical t for y-direction	1.04	0.68
Statistical t for z-direction	0.59	0.27

*Table 6.2: Variation of statistical results for magnification factors **10**x & **16**x.*

The combined effect of good accuracy and precision was also noticed during the time study, where no significant differences were noticed between the execution of tasks using the 3D display and the conventional microscope. This implies good *throughput*. Moreover, the study revealed a significant difference between each of the two methods and the 2D display alone. This effect was expected as depth perception was eliminated, but its demonstration verifies the necessity of stereoscopic viewing in the practice of microsurgery.

The *attitude* of the clinical subjects who tried the experiment was positive to the use of this new 3D viewing technology. The device was found to be easily adaptable, accommodating to the eyes while offering natural viewing conditions and substantial freedom of movement of the user's head. As a result, users were able to maintain a general sense of their position within the surgical workspace through their peripheral vision, as opposed to the immersive viewing conditions of conventional microscopes. The latter was proved a great obstacle in the microsurgical operations studied during the first cycle of the design process, and was described analytically in §5.1.1.

Similar results are also noted by other research groups as in laparoscopic [104] and endoscopic [105] surgery. It is envisaged that autostereoscopic 3D displays offer opportunities for lengthy surgeries. The additional benefit of the 'electronic eye' is that the image can also be routed to other display monitors located outside the operating theatre. This would introduce educational benefits for training purposes, as it will be explained in Chapter 8.

CHAPTER 7

DESIGN CYCLE III: INTERFACE OPTIMISATION

The design of an interface module involves the consideration of safety-critical components to ensure safe and effective functionality. In CAS systems, the safety of the patient is dependent on the system's reliability. Poor reliability may not only result from computational and technological limitations but also from a badly designed interface. In general, the goal of HCI and the human factors methodology is to optimise system performance while maximizing human safety and operational **effectiveness**. For the CAS application described in this thesis, this implies that the viewing prototype needs to generate a stereoscopic effect that will be equally pleasing to the user as one produced by the microscope binoculars. Furthermore, it must be able to maintain this effect when users with different physiological characteristics (*e.g.* interpupillary distance) interact with it. The latter consideration is very important in designing user-friendly interfaces.

The effectiveness of the HCI can be tested through field and respondent design methodologies. The former consist of field studies and experiments while the latter comprise judgement studies and sample surveys [106]. Field experiments are conducted in real settings under natural working conditions; hence, the designer can observe the impact of the new interface within the natural context of operation. Alternatively, judgement studies are laboratory experiments in which the designer collects responses from a relatively small set of "judges" about a set of stimuli. Sample surveys are typically the same, only that they solicit a broader class of respondents. The findings of the aforementioned research can be used to optimise system effectiveness.

In this chapter, the author explains an Interface Optimisation design cycle which ensures that the prototype design is user-friendly, satisfying to use, and preferred by users. The cycle consists of a series of higher order design tradeoffs intended to optimise overall interface effectiveness. These tradeoffs are conducted to finalise the design of the prototype. The methodology used in this design cycle is based on a **field**

experiment and a **judgment study**, intermediated by **iterative design** prototype modifications. The field experiment involves the use of the prototype inside the operating theatre, during a minor ENT surgical intervention. In contrast, the judgment study includes a laboratory-based psychophysical experiment in stereo vision and a qualitative analysis of effective and pleasing viewing, based on user questionnaires.

7.1 Field experiment

Otitis media with effusion (OME) is the most common cause of hearing impairment in young children and is thought to have long-term adverse consequences on children's development through associated hearing impairment [107]. The illness, also known as 'glue ear', is caused by blockage of the Eustachian tube due to allergies, swollen adenoids or other infections of the nose and throat. If the Eustachian tube is blocked, the pressure in the middle ear cannot be equalised with the pressure in the outer ear, and so it begins to decrease. This results to the production and accumulation of fluid in the middle ear, which dampens down the sound being conducted through the middle ear, and thus causes deafness. The most common method of treatment is long-course antibiotic therapy while the second in line is the surgical intervention by grommet insertion. The grommet is a small tube of about 1.5 mm length (see Figure 7.1), which is inserted into the eardrum after a small incision is made. The intervention, also known as *myringotomy*, allows ventilation of the middle ear via the ear canal and prevents fluid building-up. Grommet insertion is a good example of practical surgical skills in ENT surgical training [108].

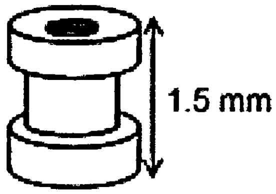


Figure 7.1: Ear grommets are polymer tubes with length of about 1.5 mm.

7.1.1 Human factors considerations

The purpose of this experiment was to study the effectiveness of the 3D prototype as a stand-alone unit in real working conditions. Obviously, in its broader application, the device is designed to be used for ENT AR surgery. However, testing the effectiveness of the prototype as a stand-alone unit before the interface is integrated with the remaining AR components can be advantageous in the design process.

Two main research issues were examined throughout the experiment: the surgeon's physical posture when looking through the prototype, and the quality of the 3D display's stereoscopic effect. More specifically, the experiment looked to identify the effects of the display's physical location in the operating theatre with respect to the physical posture of the user while he was using it, as well as the quality of 3D viewing while the display was placed in that position. Other human visual factor issues involved in the experiment included the field size of the instantaneous FOV and magnification factor (M), and their subsequent effects in determining the overall resolution of the autostereoscopic interface.

The reason for examining closely the above fields emerged from the results of the field study described in Chapter 5, and the need for stereoscopic systems to generate comfortable 3D scenes [109]. For example, in §3.2.3, it was mentioned that the image width and depth of a 3D scene depend on the camera-scene and display-viewer geometrical space relationship. However, when describing the design of CMO stereo microscopes in §3.3.1, it was explained that due to the parallel lens axis arrangement, the left and right images are projected to the eyes with little or no convergence. Additionally, the fact that the lens separation in surgical microscopes is only 25 mm, which is much less than the average eye separation, suggests further depth compression. Finally, the display-viewer geometrical space of desktop stereoscopic displays does not maintain constant perceived depth over a range of typical viewing distances [110].

Physical posture and the quality of 3D viewing were examined from direct observations and on-site interviewing of the user while he was performing the surgical task of grommet insertion. Some questions regarded the comfort of task execution and ease of prototype use while others referred to image resolution and the level of perceived depth information from the 3D scene. Observations were analysed through a **heuristic evaluation**. The latter relies on the experience and judgement of evaluators to identify usability and effectiveness problems. Given limited time and resources, heuristic evaluation can be a reliable method for improving the usability of a user interface. Claimed success rates for heuristic evaluation can be as high as 60%, depending on evaluator experience and domain knowledge [111].

7.1.2 Grommet insertion

The surgical task of grommet insertion, using the 3D viewing prototype under development, was performed by an ENT consultant during Day Surgery at the Royal National TNE Hospital, London. In its conventional form, the task involves four main behavioural and mechanical steps. Firstly, the surgeon makes a small incision in the eardrum while looking through a stereo ENT microscope. Then, he turns away from the microscope and picks up the grommet with the use of surgical tweezers. He looks back at the operating site through the microscope eyepiece and tries to bring the grommet within the instrument's FOV. Finally, he places the grommet into the eardrum opening while immersed in the microscopic environment.

During the experiment described in this paragraph, the conventional form of the surgical task was altered by the introduction of the 3D prototype. The number of steps for the execution of the task remained the same, the only changed condition being the microscope's viewing interface. A grommet was inserted in the middle ear of a single volunteer and each time the 3D prototype was placed at a selected location around the operating table. The surgeon's ability to view stereo images from the specific prototype was verified prior to the beginning of the surgical task, as he was one of the participants in the evaluation experiment described in Chapter 6.

The procedure did not require the use of physiological monitoring equipment or the presence of other clinical staff, which are normally found in the operating theatre set-up during major ENT surgery, so the 3D display, placed on a surgical trolley, could be positioned in three different locations around the operating table, as shown in Figure 7.2. In *Position 1*, it was positioned to the left of the surgeon, by the near edge of the operating table. The surgeon could view through the 3D display by tilting his head 40-60 degrees to the left from the upright position. In this set-up position, the 3D display Active Window (shown in Figure 6.16) was placed at 400 mm from the viewer. In *Position 2*, the display was positioned at the far left edge of the operating table, at a location where the junior surgeon is normally placed during ENT surgery (as shown in Figure 5.1). The Active Window was placed 720 mm away from the surgeon's face. In

Position 3, the 3D display was positioned across the table from the surgeon, in a position normally reserved for the microscope stand. Subsequently, the stand was shifted to the right of the operating table. The Active Window was placed at 600 mm from the viewer's face.

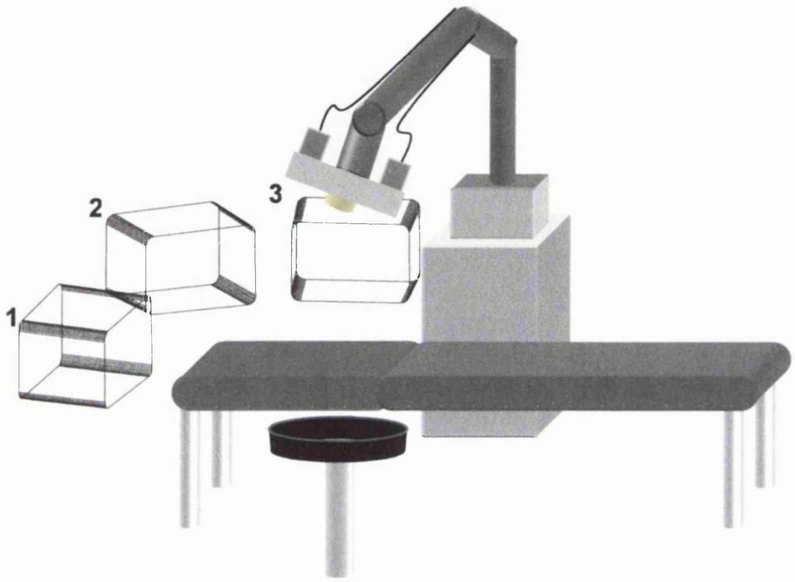


Figure 7.2: Schematic diagram of prototype set-up during grommet insertion. The numbered positions –1-2-3– show the three different locations of the 3D autostereoscopic display.

7.1.3 Observational results

Several observations were made during the experiment. They considered the surgeon's physical posture, task duration, image size, resolution and perceived depth. They were registered in a laboratory logbook, along with the user's personal comments about the viewing conditions during the procedure[29], and are listed below.

Position 1: The large viewing angle suggested a slightly uncomfortable physical posture. This was because the relative orientation of the surgeon to the display was significantly different from the orientation of the microscope to the volunteer's ear. Image brightness and size were adequate to present the surgeon with a highly resolved image of the

[29] The user was familiar with the aim of the study and, given his expertise in the application domain, was encouraged to make comments during the experiment.

operating site. Depth appreciation of the 3D content was somewhat limited. Still, there was substantial depth discrimination between various regions of the 3D scene and the surgeon was confident enough to perform the surgical task without prolonging the duration of the procedure.

Position 2: Reduction of the viewing angle improved the surgeon's physical posture. However, the long viewing distance had negative effects on image brightness and size. Furthermore, the perception of depth had vanished. The surgeon tried to compromise by moving his head closer to the screen, but was still unable to perceive a good image with depth. Hence, he was unable to perform the task.

Position 3: This position provided the most natural viewing orientation, with the surgeon looking at the display from the same viewpoint as when using the microscope eyepiece, but without any head tilt. Again, as with Position 2, image brightness, size and depth were severely affected by the long viewing distance. Nevertheless, after the surgeon moved his head about 100 mm closer to the screen he could perceive the stereo image but with very little depth and low brightness. The biggest obstacle in performing the task was the microscope's position, as the instrument was obstructing the viewer's line of sight. Consequently, the surgeon had to alter his original posture to a new one. The latter decreased the level of confidence in executing the task and prolonged its duration.

Additional comments from the surgeon's experience of testing the prototype revealed some useful information about the prototype's FOV and the overall image magnification. These were determined by the camera-scene geometry, and remained the same for all display positions. During the experiment, it was proved that the prototype's reduced FOV (see page 40) did not affect the surgical approach, as the surgeon never failed to place the grommet within the instrument's FOV. On the contrary, personal comments suggested that limited image magnification was adding some psychological pressure to the user when executing the task.

7.1.4 Heuristic evaluation

In general, the user gave positive feedback about the potential use of the new viewing interface in the operating theatre, for AR microsurgery as well as a stand-alone unit. Alongside positive comments about the natural viewing conditions and the ease to quickly establish stereo vision, a significant freedom of head movement (±50 mm from the cyclopean axis) was also reported by the surgeon. The additional load from new optical components -pairs of Zeiss TV objectives, optical couplers and Panasonic cameras- did not affect the microscope's movement, and its physical arrangement (changing FOV, see page 81) did not change at all from the conventional way. A physical constraint during the procedure had been the bulky size of the autostereoscopic 3D display. With the display placed on a surgical trolley, the closest viewing distance to the observer that could be achieved was that in *Position 1* (400 mm). In *Position 3*, the viewing distance was quite large (600 mm) and it was not clear whether the surgeon executed the task using 2D or 3D vision, even after moving his head closer to the screen. Finally, in *Position 2*, the viewing distance was so large that diplopic vision was reported and the task could not be executed.

Task execution was also affected by the interface's viewing quality. When compared to the microscope binoculars, task performance was influenced by two factors: low image magnification and depth compression. In the eyepiece, the image could be further magnified up to 30 times so that an enlarged picture of the ROI appeared to each eye. Furthermore, the user was immersed in an environment where the FOV of the operational site covered the whole visual field, thus enhancing his depth percept of the 3D image. When viewing was swapped to the 3D display, the user's total visual field also included the display's physical surroundings. As a result, the DOF of the 3D image inside the screen became comparable to the actual spatial position of the display and other objects inside the operating theatre.

7.2 Iterative design

A fundamental concept of the design process described in this thesis is that a prototype under development must be modified, based upon objective and subjective evaluation of usability and effectiveness testing of the prototype. This process of design, testing, feedback, evaluation and modification must be repeated iteratively to improve the viewing interface.

Following the findings of the field experiment described in the previous paragraph, an iterative design technique was developed in order to improve the interface's 3D viewing quality. The aim of the technique was to establish the trade-offs between FOV, magnification and DOF, and optimise viewing performance. Image size trade-offs were examined by controlling the focal length (i.e. f_i, see Figure 6.1) of the video objective adapter. The longer the focal length, the higher the magnification and the smaller the field visible on the 3D display. On the contrary, the shorter the focal length, the lower the magnification and the larger the field visible on the monitor. The size of the image on the 3D display depends on the focal length of the video objective lens and on the size of the CCD.

A possible way to improve image magnification of a video objective lens with the current commercially available optical technology is to use extenders. The latter are short relay optics that fit between the lens and the camera and increase the effective lens focal length. They are usually available at 1.5× and 2× power.

The author surveyed three different Zeiss adapters, with a focal length of 60, 85 and 105 mm respectively. Images of a target area under the microscope were captured from the two 3-CCD, half-inch, Panasonic GP-US522 video cameras for each video objective adapter selection. For the 85 and 105 mm video objective adapters, the monitor's FOV showed only a percentage of the FOV that was seen through the eyepiece. For the 60 mm adapter, the monitor's FOV was the same with the FOV of the 10× eyepiece and bigger than the 12.5× one (see Figure 7.3). The trade-off between image size and choice of video objective adapter is listed in Table 7.1.

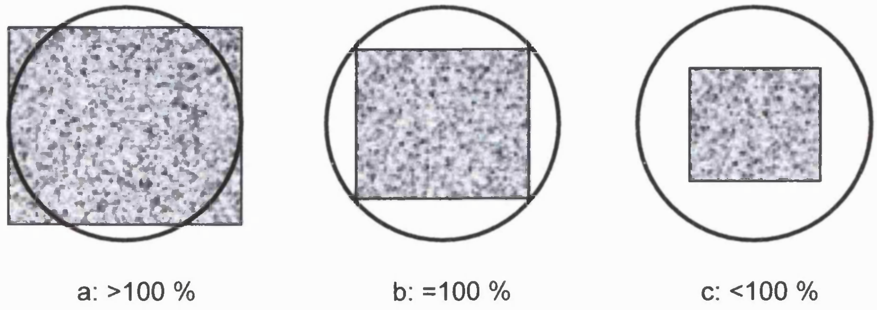


Figure 7.3: Monitor FOV (textured area) vs. Binoculars FOV (circular area). A percentage greater than 100%, illustrated in (a), means that a larger field is visible on the display than through the binoculars. The two FOVs are the same in (b), while in (c) the field visible on the monitor is smaller than the field visible through the binoculars.

Focal Length of Video Objective	Eyepiece Magnification (10×)	Eyepiece Magnification (12.5×)
60 mm	100 %	125 %
85 mm	70 %	88 %
105 mm	60 %	75 %

Table 7.1: Monitor vs. Binoculars FOV for three, different focal length, Zeiss video objective adapters.

7.2.1 Modified design characteristics

A 60 mm focal length video objective adapter provides an optical link between the microscope's beam splitter and the video camera by projecting a 1:1 magnified image to the CCD plane. When an adapter with greater focal length is chosen, a spacer has to be placed on the camera coupler in order to focus the image finely onto the CCD sensors (see Figure 7.4). At these working distances, most cameras are difficult to fix, plus they provide FOVs larger than required. Shorter working distances can be obtained by introducing spacers between the video objective lens and camera components.

In this way, camera focusing, which otherwise is performed as described in §6.1.1, is improved and the zoomed image is produced at the correct scale. For the case of the 85 and 105 mm Zeiss adapters tried by the author, images were produced at 1.4:1 and 1.7:1 scale ratios respectively. To achieve this correction, the original design of the camera coupler (see Figure 6.4) had to be modified.

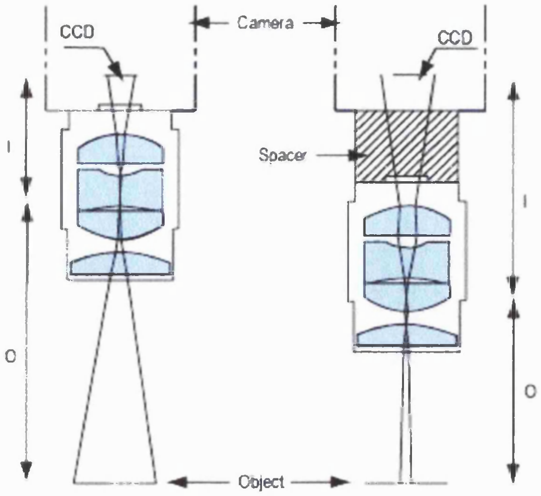


Figure 7.4: Ray tracing inside the video coupling system. When a spacer element is placed between the camera and video objective lens, the image distance increases and the object distance shortens. Therefore, the magnification of the system increases and the FOV decreases as more spacers are inserted (i.e. image distance increases).

A new design considered spacer inclusion and new camera position adjustments. Based on the design characteristics of Zeiss commercial camera couplers, it was determined that the relay lens cup (10 mm height) must be recessed inside the coupler, at an approximate distance of 4.4 mm from the bottom end. To manage this, and to ensure that the lens does not move during the procedure, the lens cup must be screwed into the coupler's housing. This can be done by threading the inner surface of the coupler housing, up to 14.4 mm from the bottom end. In this arrangement, the bottom plate of the lens (3 mm thickness) is placed 6.5 mm from the coupler's lower surface.

Other design modifications comprised making longer camera tubes which can accommodate video objective lenses with large focal length (i.e. > 60 mm). It is recommended that the length of the tube does not exceed the focal length of the adapter as this could affect the coupler's stability[30]. The modified design is illustrated in Figure 7.5. When compared to the original the video objective lens-camera coupler design (see Figure 6.4), the new modified design presents the surgeon with the option of choosing

[30] The outer surface of the camera tube has a trimmed vertical strip which allows safe locking at the focal position. This is done by placing a small Allen screw into the trimmed cavity. When the screw locks very low into the cavity a great part of the tube is outside the coupler, making the device susceptible to accidental movements.

different image magnification factors in the same manner s/he chooses different eyepiece magnification.

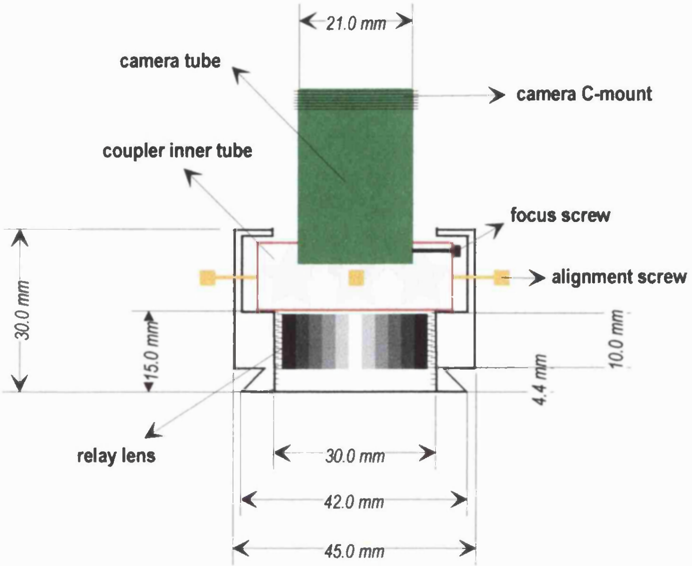


Figure 7.5: Side view of coupler's modified design.

7.2.2 Aperture-priority design

A good technique for increasing image depth in microscopic pictures is to reduce the NA of the double iris diaphragm positioned between the objective and the eyepieces (see Equation 3.6). The diaphragm is opened and closed using a wheel or lever in the microscope body housing. There are actually two diaphragms, one for each of the channels, in the CMO stereo microscope design. When selecting a low F-number (opening the aperture) the background fades and a crisp image of the ROI is created. A higher F-number value (closing the aperture) brings the entire range from foreground to background into focus. The larger the F-number, the greater the range of the image brought into clear focus. The relationship between the iris microscope's diaphragm and the depth of focus[31] in the eyepiece is illustrated in Figure 7.6.

[31] The term depth of focus, which refers to image space, is often used interchangeably with depth of field, which refers to object space. This interchange of nomenclature can lead to confusion, especially when the terms are both used specifically to denote depth of field in microscope objectives.

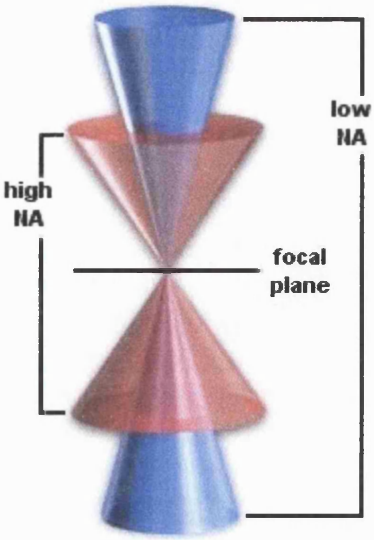


Figure 7.6: The depth of focus in the eyepiece is determined by the diaphragm's NA, that is the distance from the nearest object plane in focus to that of the farthest plane also simultaneously in focus.

Closing the iris diaphragms also produces a decrease in overall light intensity, hence degrading image brightness in both film and video camera systems. The optimum setting for microscope diaphragms is determined by experimentation. As the diaphragms are slowly closed, the image begins to display more contrast as illumination intensity slowly fades. At some point, depending upon the optical configuration of the microscope, the image begins to degrade and specimen details exhibit diffraction phenomena while minute structural details disappear. The best setting is a balance between maximum image depth and maximum contrast as seen in the eyepieces, on film, or in video images.

The technique explained above was used as part of the iterative design process in order to improve image depth, and consequently the interface's depth of field. This was done by inserting a push-in light stop into the beam splitter receptacle end of the 105 mm Zeiss TV objective (see Figure 7.7). The metallic component (with 5 mm aperture diameter) reduced the amount of light entering the video-lens, bringing at the same time a greater part of the image background and foreground effectively into focus. Reduced brightness effects were negligible, as the Panasonic GP-US522 allows automatic shutter speed and white colour balance to match the image luminance.

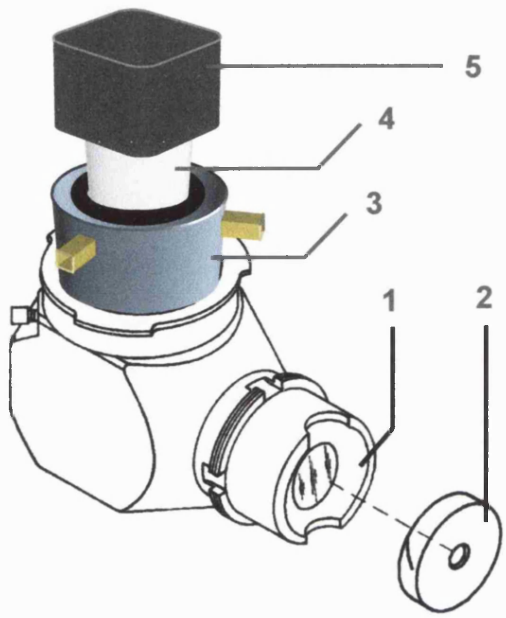


Figure 7.7: Aperture-priority modified design of the video coupling system. The diagram illustrates: 1) the beam splitter receptacle end of 105 mm Zeiss Video Objective adapter, 2) the push-in light stop, 3) the coupler housing, 4) the camera tube, and 5) the video camera.

7.3 Judgment study

When describing the physiological mechanisms of stereopsis in Section 3.2, it was noted that the human visual system operates in such manner that accommodation and convergence are linked. As a result, the disparate retinal images of contours far behind or in front of the fixation plane are blurred, and in turn, no strong stimuli arise to fuse them. Stereoscopic display techniques (described in §3.2.3) require the viewer's eyes to converge at a perceived depth a long way off the display plane but still focus on it. This implies that the amount of horizontal disparity in stereoscopic images (see Equation 3.3) should be limited in a small region (around 24 arcmin) so that the convergence/accommodation link is not stressed [112]. This condition does not allow a large amount of depth to be reproduced. For typical desktop displays with a viewing distance in the region of 700 mm this puts the comfortable perceived depth range as little as 50 mm in front and 60 mm behind the display plane. One immediate implication is that most 3D scenes will have depth compression when they are shown on such display set-ups.

A human factors study using a 13.8" LCD autostereoscopic display from Sharp Laboratories of Europe showed that the limits of depth that can be fused comfortably by viewers should be relaxed from around 60 mm behind and 50 mm in front, as far as 500 mm behind and 200 mm in front [113]. A follow-up study [29] suggested the need for a human factors-based technique that can control the amount of perceived depth range in the viewer-display space by manipulating the camera-scene space. Another subjective evaluation of viewing comfort in stereoscopic image presentation using a 21" NOKIA (445X) two-mirror stereoscope revealed that the creation of a "depth-of-interest" mini-volume could indeed improve viewing of stereoscopic images [114]. The technical approach is based on 'selective blurring' of stereoscopic images, depending on their position around the point of fixation. A simulation of the depth of focus can be achieved by changing the camera lens aperture diameter. Depth mapping over a range of apertures showed that imaging with an aperture diameter around 5 mm should be preferred to avoid distressing physiological phenomena (eye-pressure, strange pull on the eye-muscles, eye-ache). Similar psycho-optical experiments have shown that simulations of disparity magnitude and depth of focus can indeed help to extend the comfortable depth range for 3D displays [115].

In general, there is a need for creating 3D content in very much the same way as one creates photographic images by setting a stereo camera position, separation, direction and FOV diameter depending on the application and human factors considerations. This can only be achieved by simulating and controlling depth parameters, such as aperture and disparity magnitude.

7.3.1 Aperture study

Following prototype modifications, a laboratory experiment was conducted to subjectively evaluate aperture-priority design effects in image depth perception. The experimental sample comprised 10 volunteers (nine male and one female, all between 25 and 45 years of age) who had normal or corrected-to-normal vision, and were able to perceive depth from stereo vision. All subjects were familiar with stereoscopic image presentation (eight of them are computer graphics specialists, two are ENT surgeons), but were naïve as to the research question under investigation.

Five pairs of still, stereoscopic images (see Figure 7.8) were captured from the aperture-priority design modified prototype stereo camera system (equipped with the 105 mm Zeiss adapter). Large and small aperture images from each pair were displayed on the Sharp Micro-optic 13.8" autostereoscopic monitor. Viewers were positioned 540 mm (with a tolerance of ± 50 mm) away from the display Active Window.

A method of constant dichoptic stimuli presentation was used for each stereoscopic image pair. All pairs were tested in Two-Alternative Forced-Choice (2AFC) trials. Each trial consisted of one stimulus (large or small aperture) presented for five seconds, followed by the second stimulus presented for five seconds. The order of presentation was randomly determined for each trial[32]. At the end of the trial subjects were asked to judge which stimulus presented more depth. In some cases, trials had to be repeated due to subject's reluctance to make a decision. Trials were controlled by the author through the SGI Onyx2 Infinite Reality supergraphics computer.

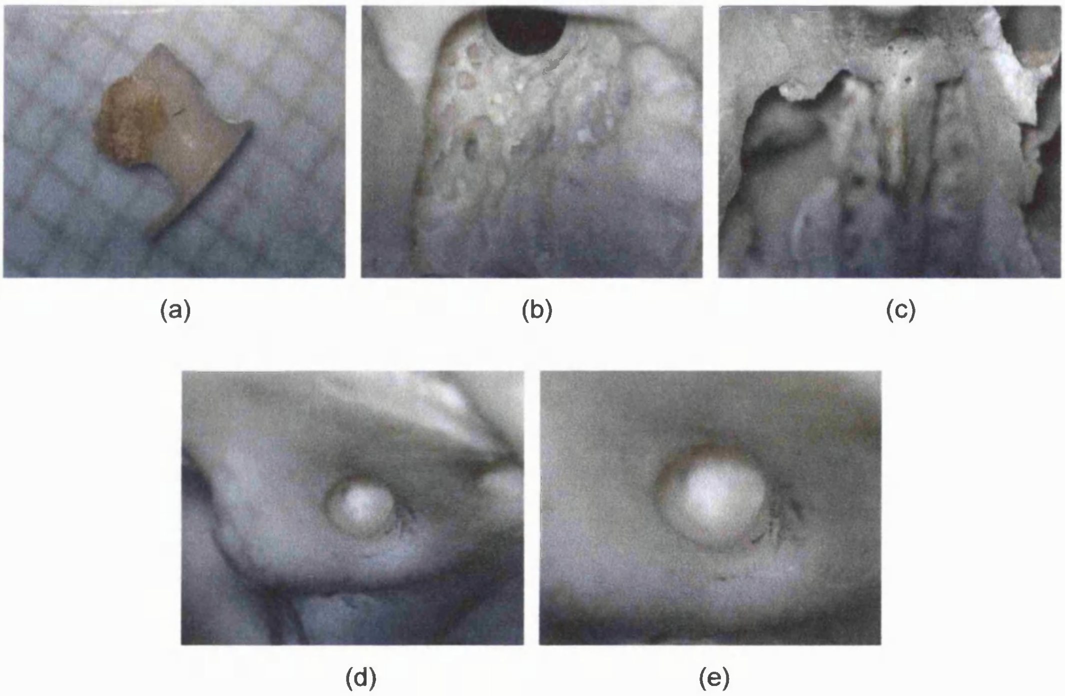
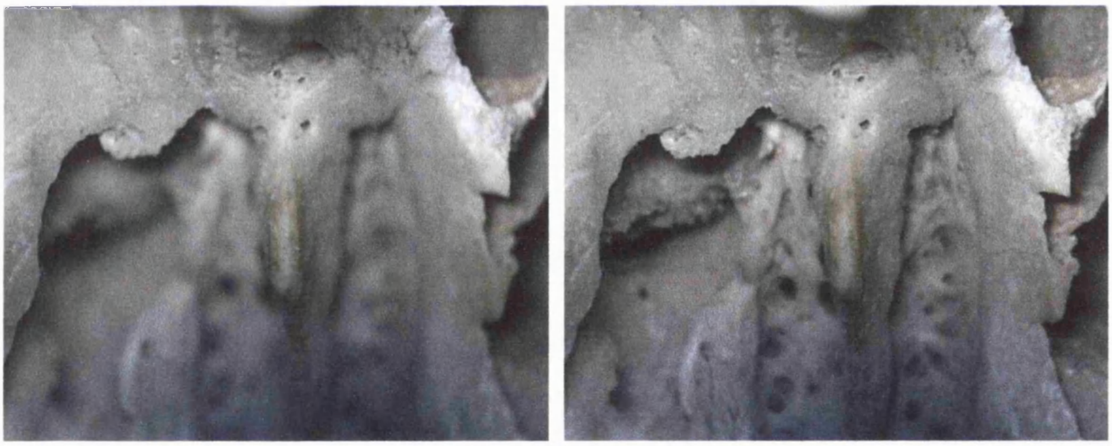


Figure 7.8:Test images of the aperture-priority design evaluation study (only one of the two images of the stereoscopic pair is shown here): (a) grommet, (b) middle ear canal, (c) ithmoid, (d) inner ear canal -10x-, (e) inner ear canal -16x-.

[32] 2AFC testing is very useful in comparing visual performance, and hence it is possible to provide feedback to the designer about the effectiveness of the viewing medium. Usually, the data obtained from testing consist of a list of the stimuli presented and the observer's response to each stimulus.

Experimental results of viewer judgment showed that all subjects agreed that the DOF of small aperture stereoscopic pairs was better than the one produced by large aperture pairs. When asked to judge on image brightness, all subjects stated that no significant difference between large and small aperture pairs was detected. This implies that although the increase of the lens F-number (from 6.5 without the push-in light stop installed, to 17.5 with the push-in light stop installed) improved the DOF, it did not affect image brightness negatively. This was achieved by performing automatic white and black balance control (camera gain was off during image capture) before images were captured.

Overall, the study indicated that the depth appreciation of stereoscopic images captured from a surgical microscope could be improved by increasing the scene's depth of focus. Stereoscopic images with a depth-selective spatial low-pass post-filtering mechanism can appear in focus over almost the whole depth range mapped. The effect is shown in Figure 7.9.



(a) (b)

Figure 7.9: Depth of focus effects in image capture from the aperture-priority stereo camera system. Image (a) illustrates a microscopic view of the ithmoid captured with a large aperture lens. Image (b) illustrates the same target captured from a smaller aperture lens. By looking at the two images, the reader can see that the blurred background of image (a) appears more focused in image (b).

7.3.2 Disparity magnitude study

When discussing the principles of binocular imaging in Chapter 3 it was stated that ideally, in a stereo camera system with parallel axes, the separation between the two cameras should be equal to the interpupillary distance (see Equation 3.2). Furthermore, to ensure that corresponding points in the left and right images are perceived in the display plane, images must be cropped during or after capture. With physical cameras, this effect is accomplished by setting the camera sensors to offset positions. For the prototype discussed in this thesis, camera cropping is established during the camera calibration process (see page 98).

In surgical microscopes, the default camera separation is set to 25 mm, the distance between the centres of the two zoom lenses inside the instrument, and cannot be increased further due to the limiting size of the instrument's housing. However, the prototype's coupler design allows the camera offset position to be as high as 5 mm away from the cyclopean axis. This enables the designer to introduce horizontal disparities to captured stereoscopic images. The magnitude of these disparities can be controlled by selecting various offset positions from the x-axis origin and symmetrically applying them to both cameras.

Disparity steps can be set by using the camera calibration grid. If the FOV and the overall magnification are known, the offset position can be calculated from the relative distance between the two cameras' origins on the grid. For example, if the final lens arrangement includes a total magnification power of 10.2× (microscope objective 0.6×, zoom 10×, and video objective 1.7×), the camera will capture only 60% of the total FOV. Since the CCD size is known, the offset position can be calculated. The example is illustrated in Figure 7.10.

The ability to set a symmetrical disparity step to both cameras allows the designer to control the location of the 3D image in the viewer-display space. In the default position, both cameras are aligned in the Zero-Disparity-Plane and the two monoscopic images have a *zero*-parallax setting. By shifting the right eye image to the –x direction and the left image to the +x direction, one can create *negative* (crossed) parallax points and consequently, put a great part of the camera-scene space in front of the display plane. Objects with negative parallax appear within the viewer space. Similarly, shifting the right eye image in the +x direction and the left eye image in the –x direction produces *positive* (uncrossed) parallax points and hence, allows the designer to place objects of the camera-scene space behind the display plane. The size of disparity magnitude is limited by breakdown of fusion and the convergence/accommodation link.

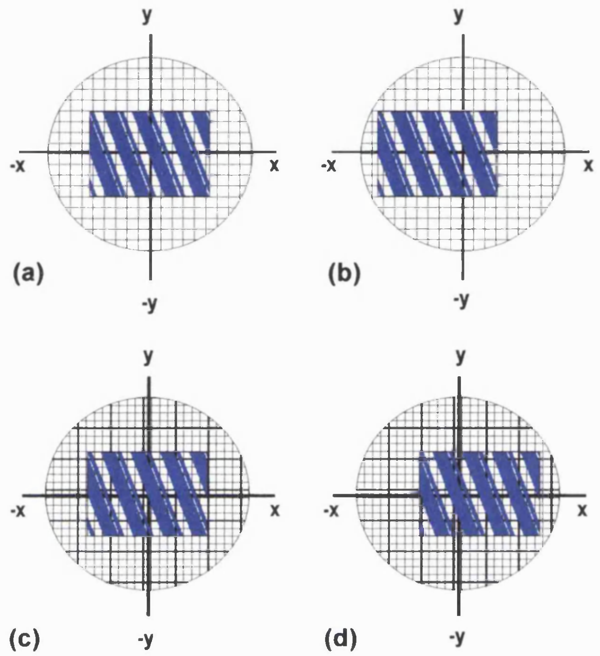


Figure 7.10: The calibration grid is used to calculate the amount of camera offset in the capturing scene. In (a) the camera is placed at the centre of the FOV while in (b) the camera has been shifted 0.3 mm to the +x direction. In (c) the camera is placed again at the centre of the FOV while in (d) it has moved to an offset position, -0.12 mm from the x-axis origin.

7.3.2.1 Experimental protocol

A second judgment study was conducted to evaluate viewing comfort and DOF perception in disparate stereoscopic images, captured from the modified prototype stereo camera system (105 mm Zeiss adapter, aperture-priority design) using the technique explained above. The experimental sample consisted of the same volunteers who participated in the aperture study.

Stereoscopic images of a target under the microscope were captured at five consecutive camera disparity steps to either side of the x-axis origin, ranging each eye's camera offset position from –1.5 mm to 1.5 mm. Each stereoscopic pair was created by combining symmetrically disparate left and right eye images from the camera origin (*e.g.* left camera step = -0.9 mm, right camera step = 0.9 mm). Overall, eleven pairs were created: five with uncrossed disparities, five with crossed, and one at the default camera position. The minimum disparity of stereo pairs was 0.6 mm while the largest was 3 mm. In a second series of the same experiment, but with different 3D content, each camera offset position ranged from –0.61 mm to +0.61 mm. The minimum disparity step between the stereo pair was 0.24 mm while the maximum 1.23 mm. Images from both series of the experiment were displayed on the Sharp Micro-optic 13.8" autostereoscopic monitor. As in previous laboratory experiments, viewers were positioned 540 mm (with a tolerance of ± 50 mm) away from the display Active Window.

The viewing comfort and DOF appreciation of uncrossed and crossed stereoscopic images from both series of the experiment were judged separately in **ranking** trials. In each trial, viewers were initially presented with the stereoscopic pair of the default camera position. After ranking the viewing comfort/DOF of the default image[33], viewers were shown pseudo-random presentations of uncrossed and then crossed disparity stereo pairs. During the trial, a crossed disparity 3D image was never followed by an uncrossed disparity one, and vice versa. A judgment on viewing comfort/DOF was made at the end of each pair presentation, before the next pair was exhibited on the display. All trials were controlled by the designer. The user's response was registered on a study questionnaire, shown in Table 7.2.

	Viewing Comfort/DOF				
Disparity offset	-2	-1	0	1	2

Table 7.2: Rank questionnaire of disparity offset vs. Viewing comfort and DOF (the total number of rows depends on the disparity offset positions tested in the experiment). Rank values vary from –2 to 2. For viewing comfort: -2=much less comfortable, -1=less comfortable, 0=comfortable, 1=more comfortable, 2=much more comfortable. For DOF: -2=very annoying, -1=annoying, 0= good, 1=better, 2=much better.

7.3.2.2 Results

Subjective judgment scores from the first series of the experiment were grouped into a contingency table of frequencies in order to test the effects of camera offset positions to viewing comfort and DOF appreciation. A Kruskall-Wallis one-way analysis of

[33] During the study, the viewing comfort and DOF appreciation of the default camera position were used as control parameters. As a result, the viewing comfort of the default 3D image was rated as 'comfortable' by all subjects. Similarly, the DOF appreciation of the 3D content for this position was always rated as 'good'.

variance (1-way ANOVA) was carried out to test for a difference between the medians of disparity groups. The non-parametric test compares multiple groups with ordinal data, using a frequency table as the input. The data is a matrix where each row represents the data in a group, and each column an ordinal value. Therefore, the number in the cell represents the number of cases in that group with that value. Raw data are shown in Appendix B.

Statistical results (see Table 7.3) showed that the creation of disparate stereoscopic images using the technique described in this paragraph affects both the comfort, and DOF appreciation of 3D viewing. Plot diagrams of Viewing Comfort/DOF mean rank vs. image disparity magnitude (see Figure 7.11) revealed that both visual parameters were affected by horizontal image translations, especially after disparities exceeded the 0.9 mm offset position for crossed and uncrossed disparities (mean rank<50%). At this stage, image fusion breaks down and the subjects are not quite able to see stereoscopically.

Kruskall-Wallis statistical parameters	**Viewing Comfort**	**Depth of Field**
Kruskall-Wallis statistic *H*	59.46	28.30
p-value	<0.0001	0.0015

Table 7.3: Kruskall-Wallis non-parametric test results of viewing comfort and DOF appreciation in Series I of the magnitude disparity study.

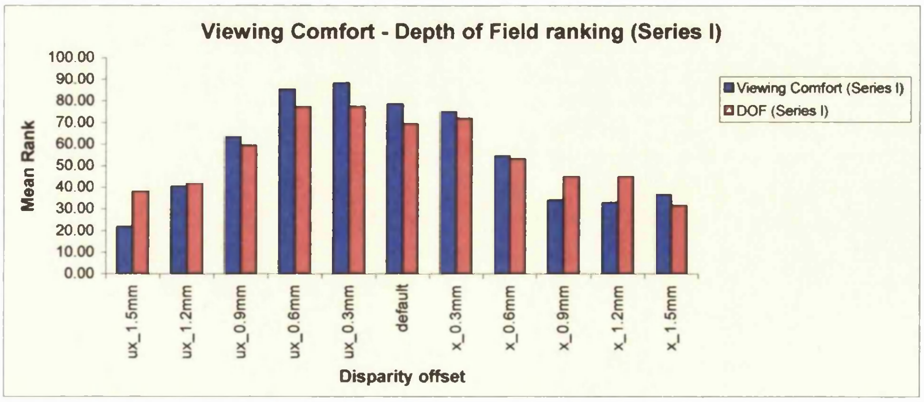


Figure 7.11: Plot diagram of Viewing comfort (blue) and Depth of Field (red) against uncrossed and crossed disparities.

From Figure 7.11 the reader can see that the mean ranks of viewing comfort and depth sensation of closely uncrossed (ux_0,3mm and ux_0.6mm), as well as crossed (x_0.3mm), stereoscopic images are higher than the mean ranks of the default one. These effects were examined further in the second series of the experiment, where the magnitude of the disparity step was smaller. A second Kruskall-Wallis 1-way ANOVA test showed that although a significant difference did not seem to exist between different levels of disparity magnitude and depth percept (*H=11.36, p-=0.33*), viewers tend to perceive uncrossed images of low disparity more comfortably than either the default or crossed ones (*H=29,32, p=00.11*). The effect can be seen clearly in Figure 7.12.

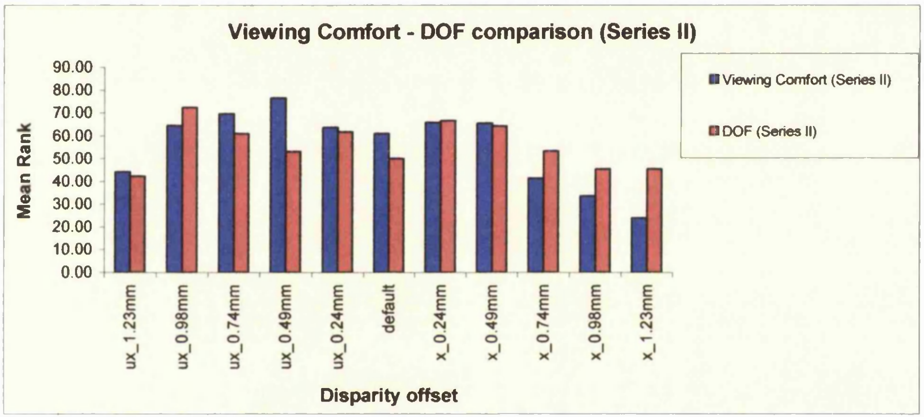


Figure 7.12: Plot diagram of Viewing comfort (blue) and Depth of Field (red) against uncrossed and crossed disparities. Greatest viewing comfort is found at 0.49mm uncrossed disparity offset.

7.4 Discussion

7.4.1 Field experiment

The purpose of the field experiment in this cycle of the design process was to examine the effectiveness of the human-machine interaction when a surgeon uses the 3D viewing interface as a stand-alone unit inside the operating theatre. The choice of the surgical task to be examined was based on its simplicity and representation, even though the interface is bound to be used in much more complex ENT tasks. Starting from a simple field experiment is more advantageous than trying to operate the interface directly in its intended environment, in this case AR microsurgery.

A heuristic evaluation of the experiment, revealed two major findings. The first finding regarded the size of the display, and the limitations that it imposed on moving the unit at various positions around the operating table. After observing the surgeon performing the grommet insertion, the advantages and disadvantages of potential display positions around the table were examined. Unfortunately, in many of the positions tried, the display's optimum viewing distance could not be achieved and the surgeon had to re-adjust to uncomfortable postures in order to experience the stereoscopic effect. The second important finding of the field experiment related to the psychological absence of imersiveness when the user was using the 3D interface. User comments emphasised that the loss of this element had severe impact on the overall depth perception of stereoscopic images. The latter were also affected by the lack of substantial 3D image magnification. Overall, the study received some very positive feedback, especially after taking into consideration the intended application of the interface. Particular praise was made regarding the viewing comfort of the interface and the easy adaptation of stereo vision.

Following the field experiment, an iterative design procedure was set up to examine potential trade-offs between FOV and magnification. Bearing in mind that higher magnification powers as well as adequate DOF are required from the interface, the designer proposed an aperture-priority modified design of the original camera coupling system (illustrated in Figure 7.7). This modification offers a pre-processing technique which improves the depth perception of 3D scenes.

7.4.2 Judgment study

The Sharp prototype autostereoscopic display requires the viewer's eyes to converge within a pre-defined depth range in the viewer-display space (set by the camera inter-axial separation in the camera-scene space) while fixating on the Active Window plane. In this way, the accommodation/convergence link is not stressed. Previous work on 3D image composition has suggested that 3D content designers are better off making scenic views appear to be seen through a window [116].

In a laboratory-based disparity simulation study, it was found that small horizontal translations of the stereo cameras to divergent offset positions away from the x-axis origin create uncrossed image disparities which increase the user's viewing comfort.

This finding resulted from a non-parametric statistical assessment of subjects, specialised in the fields of computer graphics and surgery. When viewed through the Sharp prototype, uncrossed images recede inside the display, with a small portion of the foreground still placed in front of the Active Window plane.

In ENT operations, the surgeon is more likely to work inside a small cavity than an open surface. This implies that the central part of the image will be focused in the background while the peripheral parts will appear blurred in the foreground. The cognitive effect of this condition is that the surgeon presumes that closer objects (peripheral vision) should appear with more disparities than distant ones (central vision). This explanation comes in line with experimental results and suggestions that when compositions are scenes in the real world, most of the stereoscopic image should lie inside the display space.

The necessity to map the depth range of an autostereoscopic interface was outlined at the beginning of Section 7.3. In microscopic 3D viewing the camera-scene space is dominated by the instrument's optics and the stereo camera coupling system. This implies the need for exploiting the trade-offs between FOV, image magnification, resolution, brightness and other primary and secondary visual human factors. Such exploitation can be achieved through iterative design techniques, followed by subjective assessment of potential users and scientists who are specialised in the field of computer vision.

CHAPTER 8

DESIGN CYCLE IV: INTEGRATED DESIGN

In general, the final cycle of the design process of a human-machine interface involves system implementation and installation in the intended working environment. Since most of these systems present some level of HCI, the phase usually requires testing all aspects of usability (*e.g.*, hardware and software acceptance tests, quantitative and qualitative analysis of control experiments, training plan, user and hardware manual, help system, etc.). If the new product is a CAS intervention system using AR techniques, the design phase comprises implementation and installation of hardware components in the operating theatre (viewing interface, tracker, and computer workstation). Furthermore, it includes evaluation tests of computer graphics techniques (optical system calibration, pre-operative image registration and microscope-tool-patient tracking) and qualitative assessments of human factors issues in presenting real and virtual objects in the augmented scene.

The last cycle of the design process explained in this thesis presents an extensive analysis on the communicative and broadcasting HCI parameters of the CAS system developed at the ILO, starting from the characteristics of augmented stereo video images and their real time transmission. Scientific work described in the following pages refers to the CAS system demands, I/O control management, physical limitations and network solutions for an AR microsurgical system. Special attention is paid in the cabling arrangements of hardware modules (*e.g.* video cameras, sync generators, light-to-copper modules, etc) carried into the operating theatre, as hospital regulations are very strict in respect of loose cables on the theatre floor. The microscope, together with the remaining instruments used in the operating theatre, must be in a safe and ergonomically correct position around the operating table. The layout of equipment must be arranged in such way that the surgical procedure is comfortable for the surgeon and non-hazardous for the patient.

8.1 Digital video communications

8.1.1 The digital image

The format of analogue video images was discussed in §3.3.2. The broadcasting of digital video information requires first transformation of the analogue video frame into a digital one. In this case, sampling rate and resolution become affecting factors to image quality. The former International Radio Consultative Committee (CCIR), now known as the International Telecommunications Union-Radio (ITU-R), has introduced a standard method (CCIR601) for converting an analogue image format into a digital one for which the YUV information is sampled at eight bits.

In the CCIR601 standard for digital television, the PAL analogue Y component of each one of the 625 (or 525) active lines is digitally quantised to 720 (out of a possible total 864) samples. The active area of the final digitised image occupies a matrix of 720×576 pixels. The U and V chrominance components still produce an image of 625 lines but with halved sampling rate, containing 360 samples for each one of them. The active area of the digitised image remains within the same 576 pixels, ensuring that the vertical resolution remains the same, while the horizontal resolution is halved. As a result, it is only the odd-numbered Y pixels in each of the 635 lines that have a UV associated pixel. In the NTSC system, each luminance component produces an image of 525 lines, each containing 720 samples. The area of the digitised image is 720×486 pixels. Again, the horizontal resolution is halved while the vertical resolution remains the same during sampling of the chrominance components [117]. The above sampling format is described as the 4:2:2 (2:1 horizontal downsampling, no vertical downsampling) component system. The bit rate of the two systems is the same and for the PAL SECAM system is calculated as follows:

25 frames per second

8 bits per sample

720×625 Y samples $25 \times 8 \times [(720\times625) + (360\times625) + (360\times625)] = 180$ Mbps

360×625 U samples

360×625 V samples

8.1.2 Video signal transmission

In order for a consultant surgeon to display his/her surgical approach to 50 junior surgeons, one would think that trying to fit all 50 of them in the operating theatre might not be the best of ideas. What about when the operation involves the removal of a brain tumour and requires the use of the surgical microscope? The surgeon would then not only need to show his methodology to all 50 students, s/he would also need to repeat it 50 times so that all students can appreciate the viewing conditions while operating. A more appropriate idea would be to somehow manage to display the viewing conditions to a remote location away from the operating theatre, preferably to a lecture theatre. To do that, one would need to extract the viewing images, as these are seen through the microscope, and display them to a stereoscopic monitor.

Recent advances in surgical optics have allowed the use of video cameras for displaying and documenting surgical procedures (as discussed in Chapter 3). The video cameras use analogue electronics, capturing moving RGB pictures. These pictures are normally displayed to an RGB monitor, placed next to the remaining monitoring equipment in the operating theatre. By doing that, it is possible for the surgeon to check on his/her actions and other consultants to monitor and comment on the approach.

Extracting optical images from the surgical microscope offers the opportunity of displaying the visual information at some distance away from it. With the use of copper cabling, the analogue RGB signal can travel a maximum distance of a few metres before it begins to degrade in intensity and phase. Converting the signal into a digital format and then transmitting allows integrated signalling for tens of metres before amplification and synchronisation is needed. This is because after a maximum distance, the moving electrons that transfer the signal begin to scatter randomly, thus weakening the intensity of the signal. Moreover, electron scattering increases the cable's thermal conductivity, thus increasing the resistance of the path the electrons travel through. The latter alters the frequency of transmission. When the images are transmitted by means of light the signal is transferred through photons, thus allowing integrated signal transmission for several kilometres without any pre-conditions [118].

Both cases of signal transport consider issues such as attenuation, usable bandwidth, electromagnetic field interference and network structure. Attenuation results from cable jumpering and in-line or branch nodes. When using optical fibre, the attenuation loss of the signal is around 0.1 dB per kilometre. For copper cables, the attenuation loss is about 1 dB every 30 metres. This poor performance is due to the cable's physical properties. In terms of bandwidth, maximum transmission rate of about 1 Gbps is possible using coaxial cable whereas transmission rates in excess of 10 Gbps are possible using optical fibres. Real-time moving images of a video picture contain enough information to require the additional bandwidth capabilities of fibre [119]. Moreover, when an electronic signal travels through standard coaxial cable it becomes sensitive to nearby electromagnetic signals, generated from transformers and motors. On the other hand, light waves are not affected by external crosstalk [120].

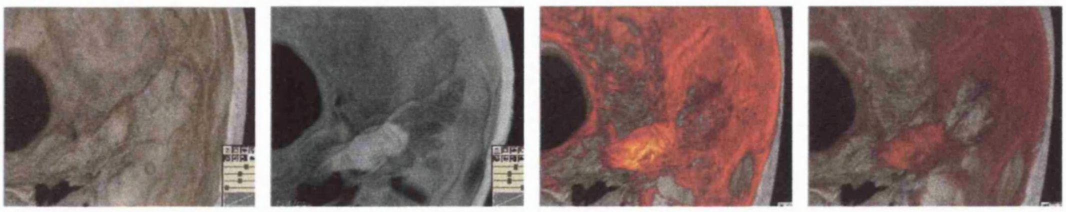
So far, only the unidirectional one-to-one link of a RGB video signal has been considered. The circumstances become more complicated when the RGB channel needs to be viewed at several locations. The link then changes to a multipoint system where the original signal is branched and transmitted to all various endpoints. This type of transmission is called broadcasting. Television networks use broadcasting to transmit live colour pictures to millions of end users. Broadcast channelling can be single-mode or multi-mode, permitting more than one image to be transmitted simultaneously. More demanding systems, require data exchange and image display capabilities between different endpoints. When every endpoint can reach each of the others the system is called a network. To be maximally useful, each end must be able to provide data to any other end. A more detailed analysis of broadcasting computer networks can be found in Appendix C.

8.2 System demands

As it has been demonstrated, the CAS system has communicative and broadcasting applications. Initially, it enables the microscope's live optical images to be transmitted to the computer mainframe. While residing in the server's memory, the images are augmented with the 3D reconstructed, computer generated MRI/CT data. The stereoscopic images must be transmitted in concurrent pairs of left and right eye monoscopic images. After image-processing techniques have been applied, the temporal augmented stereoscopic pair must be transmitted to the 3D display inside the operating theatre. The same images may also be displayed in the lecture theatre and in the computer laboratory or another remote location, such as the seminar room.

Each optical video image undergoes a series of image processing techniques while residing in the server's graphics memory. The most crucial, image augmentation, includes the following steps:

- Creation of a base picture layer containing the real world digital image frame with an active area of 720×576 pixels (see Figure 8.1a),
- Overlay of a second layer that includes the virtual digital image (see Figure 8.1b). This second layer may be given an amount of transparency, identified by the image's alpha (α) value. The latter is part of a 10-bit sampling rate where the 4:2:2 bits are devoted to luminance and chrominance sampling and the remaining two bits define the image's degree of transparency. Alpha sampling is an internal sampling procedure and does not affect the overall sampling rate from analogue to digital transmission,
- Overlay of a third layer that contains the digitised real world image with some degree of opacity. The later is chosen by the user and can vary throughout the process. It serves the purpose of increasing or reducing the opacity of the real world visual information before the augmented image (see Figure 8.1d) is finally presented to the viewer,



(a) (b) (c) (d)

Figure 8.1: Image augmentation of a cadaver skull using "sandwiched" digital video frames. Frame (a) is captured from the stereo camera system while frame (b) is generated from the SGI video card. Frame (c) presents the "augmented virtuality" operating scene, while frame (d) displays the "augmented reality" scene.

The physical transmission rate in bytes of the transmitted augmented video image can be calculated as follows:

pixel area of frame × no. of colour layers × no. of frames per second = image size

For a real time, live, PAL system, video augmented image the final size is:

$$(720\times576) \times (3) \times (25) \approx 31 \text{ Mbytes per second, or}$$

$$31 \times 8 = \mathbf{248\ Mbps},$$

It is estimated that the synchronous transmission of the stereoscopic pair will require a bandwidth of about 500 Mbps. Synchronism of the video devices involved in the application is vital for successful broadcasting of the signal. This implies that the video cameras attached on the microscope, the server's frame grabber, the 3D TV and other TV monitors must be functioning under the same synchronisation pulse. This can be achieved by connecting all the video devices to a single sync generator. In serial transmission, the sync generator identifies the correct transmitted sequences of the bit information and as the bit pattern is decoded, all devices are synchronised to the same logic transmission. The mechanism is called genlock and ensures that all devices operate under the same clock conditions. The vertical and horizontal resolutions of the sync pulse are set at a specific frequency by the generator. The clock works in binary mode and during operation, it produces a potential difference of 7 Volts[34].

Another broadcasting parameter in the system is the display on the Silicon Graphics screen. As the user must be able to view and perform computer applications in the operating theatre or in the lecture theatre, it is important to have a remote computer terminal at these locations. This implies that the VGA signal of the computer monitor must be transmitted from the server's location to the computer terminal, wherever the latter is placed. The monitor signal is a standard RGB signal for which the sync pulse is 'hidden' inside the green colour signal ('sync on green'). The image size is normally 1280×1024 pixels. The overall transmission of the signal occupies roughly 335 Mbps of the total bandwidth.

[34] During operation, the clock switches between two discrete values, 0 & -7 Volts.

The system must also be communicative. The users must be able to interact with the computer at any time during the surgery. They might need to manipulate the appearance of the computer-generated data or to perform additional software applications. These applications are executed through buttons pressed on the user's mouse or by specific keyboard commands or through specialised input devices. For example, with any mouse motion the user end sends an appropriate signal to the server which denotes an action. Similarly, a keyboard command abets the server to execute an application. The above sequence of steps for both devices states a bi-directional link between the communicative user terminal and the computer's mainframe. In terms of bandwidth, the above connections occupy about 512 kbps of active bandwidth. This is similar to the data transfer mode through a personal computer's RS232 serial port; only it is twice as fast.

Another communicative link between the server and the user in the augmented reality system involves the tracking of medical instruments, such as the surgical microscope and the surgical burr, as well as the patient's head position. As mentioned earlier, all instruments, together with video and virtual data must be unified under a single co-ordinate world so that all visual information is correctly correlated. Tracking of the microscope, burr and patient's head is carried out using a tracking device called POLHEMUS. The link between the POLHEMUS device and the computer's server is established using one of the computer's RS232 serial ports. The transmission is uni-directional and takes place at 64 kbps.

The CAESAR system can be a real-time surgical tool for microsurgery or for surgical training of trainee surgeons. In a different case, the trainee surgeon may be asked to perform particular surgical tasks on a cadaver's head. This can be done in the computer laboratory. As a result, all the dedicated operating theatre links must appear in the computer laboratory too.

8.3 Broadcasting I/O management

The original theoretical idea of introducing live optical images obtained from the microscope to the Onyx2 Supergraphics computer was based on the concept of mixing live optical frames with 3D rendered computer images. The insertion and extraction of data had to be done by using an image frame grabber, called DIVO (DIgital Video

Option). Upon completion of problems such as camera capabilities, conversion of signals, type of display etc., the author managed to build a complete system whose architecture is shown in Figure 8.2 below.

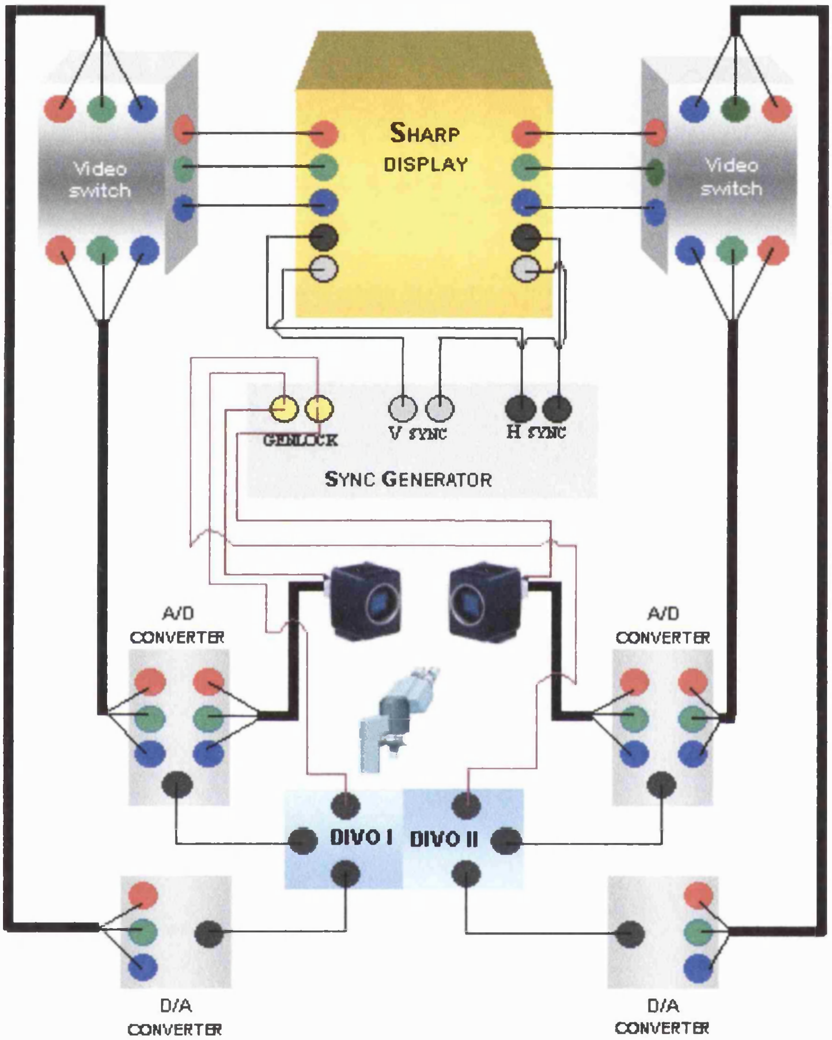


Figure 8.2: System Architecture diagram.

In the figure above, one can see that the monoscopic analogue RGB video images are originally converted to 8-bit CCIR601 fast serial digital video before they are input to the DIVO cards. The processed data sit in temporal memory stacks before they are extracted through the DIVO output channels. Conversion of the serial digital augmented monoscopic video images to RGB format occurs at the 3D display end, just before displaying. The latter ensures lossless transmission and full recovery of visual information. The successful transmission of such images is accomplished using a sync generator that genlocks the pair of images coming out of the two DIVO channels. A similar synchronisation technique is used for the correct setting of horizontal and vertical sync pulses on the 3D display. The video switches appearing in the figure are used for the optional switch from live augmented pictures to pure optical ones. During surgery, the surgeon does not need to use augmented reality throughout the entire procedure but only at the time of burring or approach to the tumour region. The switches offer the option of performing operations in the same way that they are performed with the stand-alone surgical microscope.

A number of different variations of the original architectural designs were tried, some of which proved to be unusable while others became subject of further investigations. In one of these variations, we managed to develop a method in which the Onyx2 generated separate component analogue PAL RGB images of the augmented data. This was achieved by upgrading the internal architecture of the computer. More specifically, the author managed to extract the two live augmented images as these appear on the computer monitor by using the computer's graphics pipelines. This was accomplished by setting the computer's multi-channel generator in such a way that each channel carries and displays graphical information equal to the screen size of the mono image. The video signal of the images in these channels were set to be in analogue PAL RGB format, and were directed to the 3D display to produce live 3D augmented pictures.

8.4 Network Communications

8.4.1 Health care and Information Technology

In a hospital environment, a network would be used to transmit diagnostic images from one department to another, live video images from the operating theatre to the lecture theatre, DICOM 3 information, other computer data, as well as audio data transfer between medical specialists. To speed up data communication, people tend to use fibre optic cables for data transfer. Some groups [118] have tried to connect operating theatres, lecture halls, Intensive Care Units, radiology departments and video conferencing studios through the construction of a Local Area Network (LAN). The connection set-up is illustrated in Figure 8.3 below.

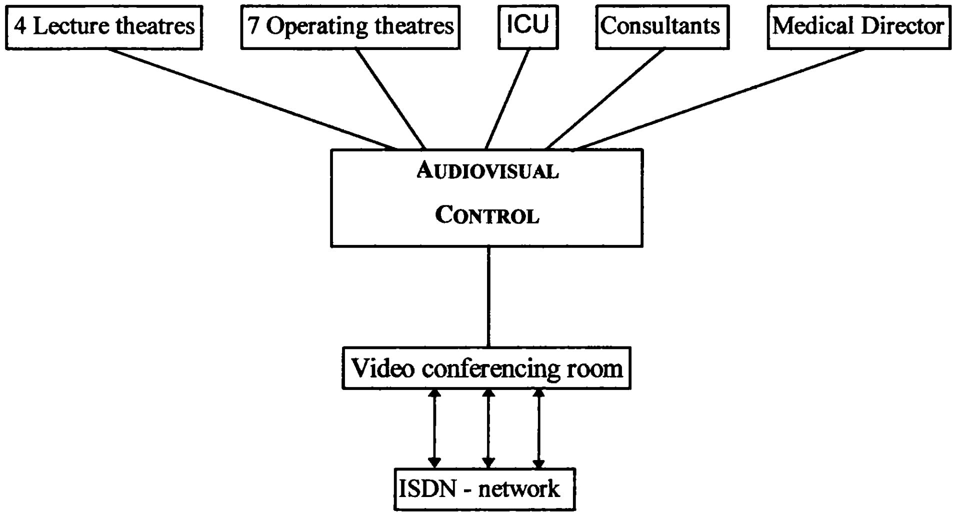


Figure 8.3: "In-house" audio/video communication at the Department of Surgery, Technical University Munich.

The Institute of Laryngology & Otology is located at an approximate distance of 150 metres from the hospital's operating theatres. The computer mainframe room is situated at ground level, with the audio-visual control unit and the lecture theatre stage located right above it, on the first floor and at near distances of 15 and 30 metres respectively. The Institute's computer laboratories are placed on the second and third floor of the building, at distances of 35 and 40 metres respectively. Linking them together is difficult because there is no existing cable duct to house the fibres. Hospital regulations are strict for cabling layout. All cables must be hidden and out of reach for both staff and visitors. Safety standards suggest that new cabling should preferably pass through existing holes. Layout of new cables must satisfy all criteria set out by the hospital's management, electrical and mechanical groups.

Fibre optics cabling allows the transmission of signals without any significant loss of information. The installation of fibre optics offers the advantage of transmitting signals over long distances.

The author has designed a system that meets the demands of signal transfer and data presentation, according to the various locations of display [121]. The transmission originates in the operating theatre with the acquisition of the real world video signal and the setting of the virtual environment. The signals arrive at the Onyx2 mainframe site for augmentation. They are then transmitted back to the theatre, and mirrored to the central computer laboratory and the MSc students' lecture theatre. This set up is shown diagrammatically in Figure 8.4 below.

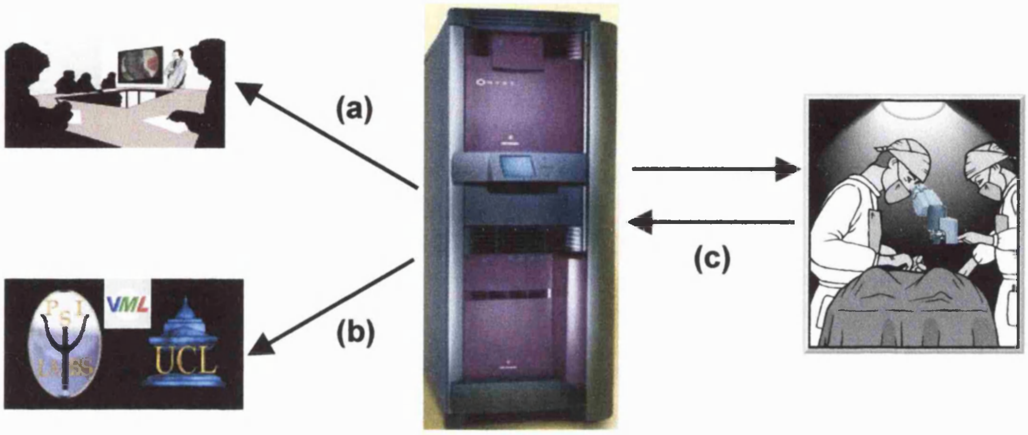


Figure 8.4: "In-house" audio/video communication at the Royal National ENT Hospital at UCL, London. Arrows illustrate direction of information transfer while letters in brackets symbolise distance between communication ends (a=150 metres, b=15 metres, c=30 metres).

8.4.2 Implementation of a Dedicated Optical Fibre Network for Real time Stereo Imaging

Following the design of a network that makes the use of the AR microsurgical system possible, the author tried to implement such a system in a hospital environment at the Royal National ENT hospital, London.

Consideration was given to the ease of implementation, costing and the distribution of connection points. In view of the distance problem and complexity of signalling format and its bandwidth constraints, fibre optic links work out to be the optimal choice against copper cabling. Optical fibre cabling has excellent noise immunity and minimal signal attenuation. The cable size is also very small (~8 mm diameter for 24-core) and therefore can be used in narrow ducts. The choice of using single mode or multi-mode fibres depends on the application. With single mode fibre the signal travels in one direction alone (from point A to point B) and can use the full bandwidth of the fibre core. With multi-mode fibre, the signal travels bi-directionally. The core's bandwidth for each direction is halved and cannot exceed the transfer rate of 140 Mbits per seconds. It also requires complicated and expensive signalling components, (multiplexers) which means further cost.

For cable planning, the author allocated more fibre channels than required to allow for redundancies so that one can use the additional cable in case of signal degradation/disconnection problems and for future expansion. Multiplexing technology was considered for low bit rate transmission like RS232 serial links, mouse, keyboard and voice transmission, but the cost of multiplexing and de-multiplexing exceeds the cost of individual signalling equipment [122]. Since one can pack more fibres in a single cable, it was decided to use "dedicated" point-to-point links. Patch panels are used at strategic locations to distribute the signals. One of the advantages of working with light waves is that they can be split using one-to-two channels optical fibre couplers to multicast the signals to two locations simultaneously, without any signal degradation. This proves to be a very useful and cost saving characteristic of a fibre link.

The signalling components for transmission and reception also form part of the overall costing formula. Each of the devices has its unique signalling properties. Figure 8.5 shows the use of fibre for the case of augmented reality application.

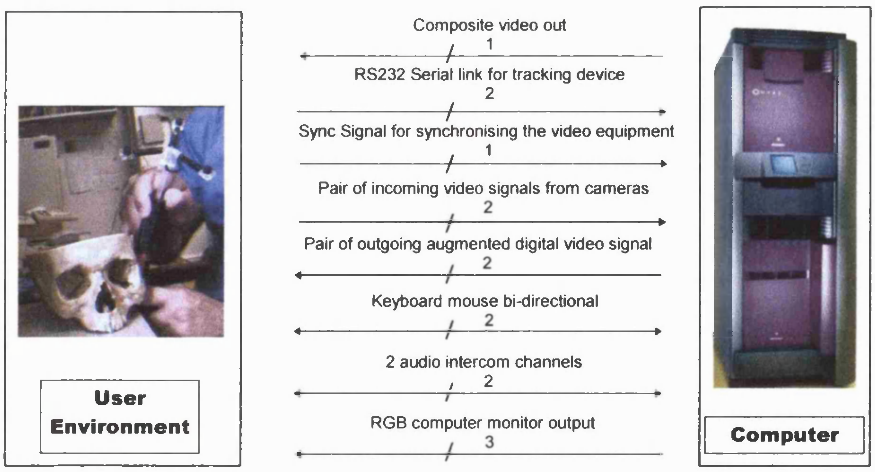


Figure 8.5: Fibre Optics dedicated links for CAS Microsurgery.

The budget allocated for the work was around £20,000. Several industrial suppliers were contacted to bid for the job. Upon consultation with their sales and support engineers, it was realised that the signal transmission of our Super Graphics computer was the principal concern to most suppliers who have basic capability in installing standard fibre computer networks. As a result, it was decided to design a prototype architectural system and select a specialist optical fibre company that has in-house design skill. The author came into close collaboration with two industrial contractors, Fibre Optic Communications Ltd for the cable installation and Cove Communications Ltd, for the supply of the fibre optics equipment.

The optical fibre network system relies on a central patch panel adjoining to the Super Graphics computer to distribute its signals to the three different locations (see Appendix C). At each location, depending on the usage, a sub hub is used to redistribute the signals to the respective rooms. At the theatre end, the signals are carried using a 24-core fibre optic cable. A sub-hub at the theatre office, patches the signals to each of the three operating theatres, which are about 15 metres from the office. Another 24-core fibre connects the central patch panel to the computer laboratories. A sub-hub is used in the first computer laboratory to distribute the signal to the floor above it, where the second computer laboratory is sited. The author also used a 16-core fibre optic for the connection between the lecture theatre and the central patching hub. A sub hub is located here to allow local distribution of audiovisual signals over a dedicated 12-core fibre to the audio-visual control room and a 16-core to the presentation stage. A graphical presentation of this architectural design can be seen in the following diagram (see Figure 8.6).

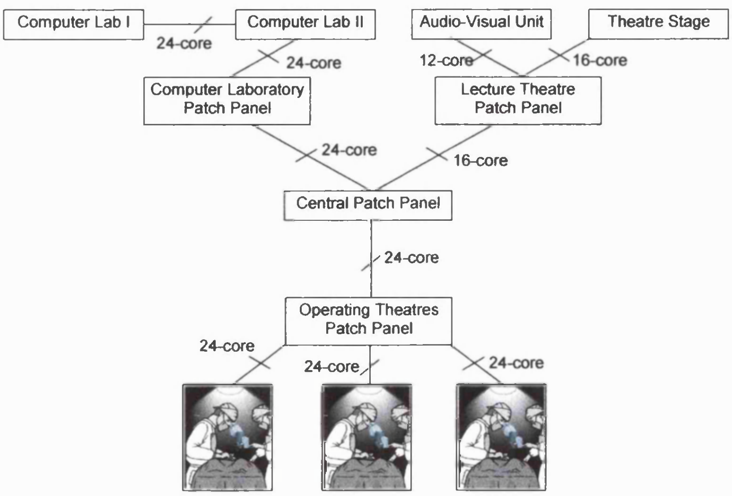


Figure 8.6:Fibre Optics Architectural Design at the Royal National ENT Hospital, London.

Network designs are usually justified by their *openness*. An open system is advantageous because of its ease to use, ease to change and ability to serve a wide variety of resources that are in use concurrently. The 'dedicated-links' Optical Fibre Network at the Royal National ENT hospital is designed in such manner so that it can be used for various applications. In one particular case, the system is used to link a permanent computer terminal in the lecture theatre to the computer's mainframe for teaching purposes. With this configuration, the user is offered direct access to software applications and subsequently, the viewing conditions of the augmented environment (as was mentioned earlier, the software allows the user to increase or reduce the opacity of the real world image). It is believed that by doing this, the whole procedure becomes more interactive, with the audience actively participating in the surgical procedure. Fibre redundancies in the computer laboratory enable the location to become a mirror site of the operating theatre. Hence, users are allowed to perform cadaver dissections and phantom studies outside the operating theatre. This is very important for junior surgeons training as it offers them an alternative place of practice while reducing the number of people in the operating theatre.

CHAPTER 9

Conclusions

So far, surgery has been relying on the surgeon's ability to visualise in 3D the complex anatomy seen in the diagnostic scans, and relate it to the surgical scene. Surgical planning is based on previous experience acquired by the surgeon and bibliographic references. Surgical simulations are dependent on the virtual appreciation of the real world operating scene. Unfortunately, virtual interpretations of the real world are not always accurate and can mislead the course of surgery. Still, they could form the basic practice of training surgeons during the early stages of learning on how to perform surgery. The customary approach to delicate microsurgical operations will always be biased by the visual feedback the surgeon has after skin incision or bone burring has started.

A case for which the above argument is best explained is mastoid surgery, as it is associated with major complications. Even in the best hands, damage to the inner ear occurs in over 2% of operations, and causes a profound untreatable hearing loss [123]. Facial nerve (see Figure 9.1) damage is another potential problem. The incidence of this complication ranges from 1% to 5%, depending on the size of the tumour. These figures further increase when operating on a previously operated on, or irradiated field. The incidence of facial paralysis increases in younger patients because the facial nerve follows a different path in children. Uncertainty of facial nerve anatomy in the young as well as individual anatomic variation between patients can lead to inadequately performed mastoidectomies. These patients continue to have discharging ears and present to the outpatient clinics frequently (20% of primary operations) and there is a 1 in 200 lifetime risk of a person developing an intra-cranial complication from an operated ear [124]. This inflicts an enormous and avoidable cost on the health service, which could be reduced by providing surgeons with more detailed information on the individual patient anatomy.

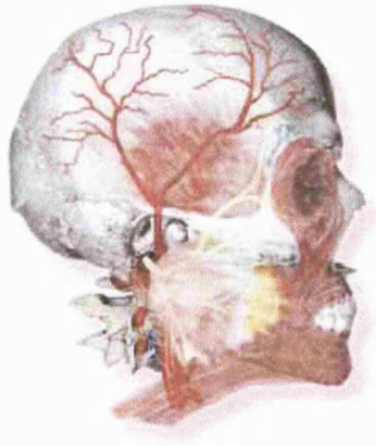


Figure 9.1: Illustration of the facial nerve and its neuronal branches as they spread around the face (nerve coloured in yellow).

The purpose of CAS is to enhance surgeons' knowledge of patient anatomy during surgical interventions. At first, this was very difficult, but as the processing power of computers increased, interactive 3D visualisation of pre- and intra-operative diagnostic scans became more feasible. One of the remaining obstacles in the use of CAS systems has been the visual presentation of such information, as most commercial systems display virtual scenes on 2D computer monitors, thus requiring the surgeon to look away from the operating scene. The problem becomes more evident in microsurgery, where viewing conditions are already 'unrealistic', since the surgeon uses a stereo microscope to view the surgical lesion.

Recently, scientific research in the field of computer medical graphics has shown how AR can be used to overcome the aforementioned problem. By overlaying registered VR models of patient anatomy on real world operating scenes, CAS AR systems equip surgeons with 'X-ray' vision. Some of these systems were described in Chapter 4, along with their advantages and disadvantages.

A common limitation of current CAS AR systems is that they do not decrease the stress factor of microsurgery. One of the reasons that they have failed to meet expectations is due to the lack of ergonomic viewing interface design. The problem originates from the fact that surgeons are required to either wear special goggles, or use the microscope, to view the augmented operating scene. In the extreme environment of the operating theatre, this reduces the level of HCI, thus making surgeons revert to the conventional method.

Matching the display to the specific tasks and needs of the user can be complicated, however, requiring expertise in vision sciences as well as display technology. In this thesis, the author has described and evaluated the design of a viewing interface for computer-assisted microsurgery that overcomes some of the above HCI limitations, by replacing the viewing end of a surgical microscope with an autostereoscopic 3D display, thus equipping surgeons with 'electronic' eyes.

9.1 Major scientific contributions

This work contributes a system which has undergone informal usability testing in the context of real domain experts doing real work, but it also presents the results of laboratory-based experimental evaluations which illustrate general behavioural principles. Some of the informal design issues, which user observations have suggested, are now supported by formal experimental evaluations. Moreover, the selection of experimental evaluations has been tempered by design experience, so there is some assurance that subjective and objective assessments apply to design in a meaningful way. Together, these approaches make a decisive statement that using the prototype discussed in this thesis can result in improved computer-assisted microsurgery.

The major contributions of the author's work are:

- *Human-centred design*: The author contributes an ergonomic interface design which, by creating physical and behavioural aspect models of conventional ENT surgery, is based on the surgeon's mental model during the microsurgical procedure,
- *Application to ENT surgery*: The author contributes the design, implementation and usability testing of a stereo video camera system which captures stereo pairs of images from a surgical microscope and displays them in an autostereoscopic 3D display,
- *Mechanical coupler design*: The author contributes a coupling component design which can align two disparate camera views onto the display plane of

a 3D display (stereoscopic or autostereoscopic), and can be optimised for various focal settings,

- *Training and teaching at remote locations*: The author contributes the design and implementation of a dedicated optical fibre network for CAESAR surgery in three operating theatres and a computer laboratory, as well as real time stereo imaging at a remote seminar room,
- *Research methodology*: The author contributes a thesis which provides a case study for an interdisciplinary approach to the design and evaluation of new human-computer interfaces,

From a task analysis of conventional microsurgical operations, it was realised that there are strict limitations as to which direction the microscope can be physically expanded. Even though microelectronics technology can offer these days inexpensive, state-of-art optical and video components, there are always additional human factors that need to be considered when altering the original design of the instrument. The physical aspect model of microscope-affected tasks derived from the first design cycle suggested that a "sideways" or "away from the surgeon" (z-axis, see Figure 5.10) addition of new modules could damage the instrument's counterbalance and overall movement, particularly tilt. As a result, an "upwards" (parallel to the eyepiece) module design was chosen.

The broader vision of this research is eventually to replace the microscope binoculars with a more natural viewing medium and limit the instrument's use to a video magnifying "torch". In this way, surgeons would be completely relieved from the burden of immersive interfaces. This has been delivered through the development of the stereo video camera system shown in Figure 6.17. The main characteristics of the design are the parallel-camera viewing geometry and open-architecture which ensures that the system can be used for different microscope brands. This openness is extended further, as modifications in the design of the mechanical coupler enable users to adjust the system to various FOV and DOF settings, therefore optimising the appearance of 3D content on the autostereoscopic 3D display.

The success of a CAS system depends on its ability to process and display accurately high levels of interaction in real time. This becomes more difficult when stereoscopic viewing is required, as in the case of microsurgery. Up to date supergraphics computers, with digital video frame grabbers and graphics pipelines, can provide sufficient computing power for real time implementation of CAS applications. However, such computers are bulky and need to be stored in air-conditioned rooms. The optical fibre network, described in Chapter 8 of this thesis, offers an end-to-end digital audiovisual communications platform for CAS surgery in any of three operating theatres and the hospital's imaging laboratory, all located far away from the computer mainframe room. Furthermore, it allows real time stereo multicasting of the surgical procedure in a seminar room, also located far away from the supergraphics computer. The network is implemented with fibre redundancies, thus providing the infrastructure for future tele-robotic surgical applications.

The research methodology of this case study is statistically correct. Statistical analysis methods need to be carefully chosen during product evaluation. As part of the research work included some quantitative and qualitative analysis of usability and effectiveness testing, the appropriate statistical analysis techniques were selected. A paired t-test analysis was chosen for the quantitative analysis of the interface's accuracy and precision. The need for a parametric test arose from two main reasons. First, the nature of the experiment (task execution by an operator who uses two different viewing media in succession) was such that suggested a matched comparison[35]. Secondly, sample homogeneity (86% of participants being ENT surgeons) and a quick inspection of results from a single operator suggested that experimental data form a Gaussian distribution. For the qualitative analysis of the interface's effectiveness, a Kruskall-Wallis non-parametric test was selected. The choice of this analysis technique was because the author needed to test viewing comfort and depth appreciation over a large set of stereo camera disparities.

[35] If the author had chosen free-viewing measurements as a third matched group data, an alternative method of estimating statistical error would be repeated-measures ANOVA. However, the "golden standard" value was the actual position of the selected point as it was measured with a great deal of haptic feedback.

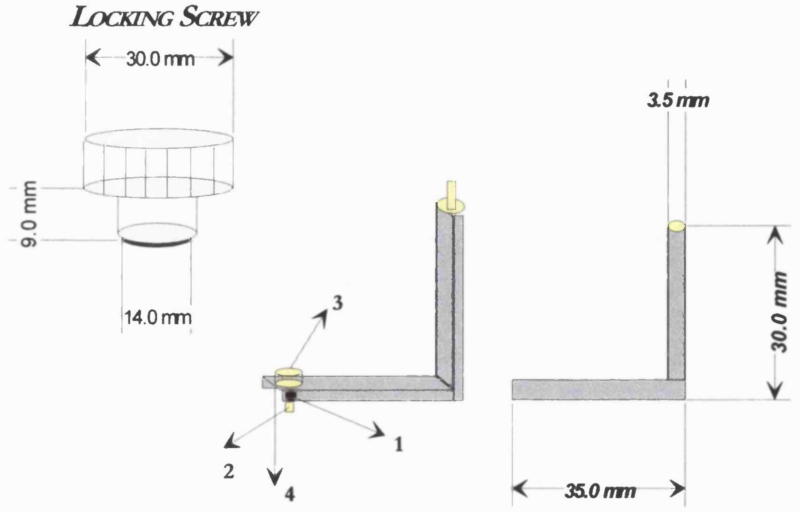
9.2 General remarks

Overall, this research has tackled some difficult problems of perceptual issues in stereo video such as camera separation, expanded DOF and the accommodation-convergence conflict. For example, it is known that 3D displays require the viewer's eyes to focus on the display plane in order to see the objects clearly, regardless of the position of the objects. In fact, many researchers stress that disturbing the accommodation-vergence link is the major cause of eyestrain in stereoscopic displays. However, none cite any work that actually demonstrates that this is the case. However, all optical systems, including binoculars, goggles, microscopes, etc, disturb this link to some extent. Nevertheless, the human visual system is known to be flexible and robust, and can adapt to new situations quickly. Furthermore, when accommodative responses to changes in convergence distance caused by ophthalmic prisms were measured, it was found that although subjects initially under-accommodated the target, they quickly adapted for near target distances, with no significant after-effects [125]. Findings from psychophysical experiments, described in Chapter 7 of this thesis, have revealed similar results. In fact, results from depth mapping experiments suggested that viewers experience greater viewing comfort and depth appreciation when presented with slightly disparate stereoscopic images captured from the prototype interface (see Figure 7.11). These findings come into agreement with suggestions that stereoscopic cameras used for endoscopic surgery, where DOF is restricted, can reduce the distracting effect of disparities in the image [126].

One of the main drawbacks of the new viewing interface has been the bulky structure of the prototype Sharp autostereoscopic 3D display. This problem became more noticeable during the field experiment described in Chapter 7, as the number of possible viewing positions around the operating table was limited to three. From these only Position 1, where the 3D display is situated 400 mm left to the surgeon, promises some degree of ergonomic viewing which in turn is limited by the fact that the surgeon still needs to rotate his head. In early practice this might not be considered as a limitation, but in the end its application might inflict problems not very much different from the ones occurring with microscopic viewing, which this research work has been trying to tackle. Furthermore, in such a set-up the shortest viewer-display distance is limited to 40 cm (unless the arrangement of other staff and equipment inside the operating theatre is re-organised). This means that ideally, a larger 3D screen (≥18") is needed but that would infringe the cost-effectiveness of the interface. Nevertheless, some scientific groups have tried to use autostereoscopic 3D video technology in ophthalmic surgery [127], as shown in Figure 9.2.

Figure 9.2: Operating theatre surgeon-display space arrangement using the Dresden D4D® 3D display.

A more natural approach should be preferred for the viewer-display space set-up, one that would not change the traditional general posture of surgeons during operations, neither challenge product cost-effectiveness. This could be achieved if a small size LCD screen could be mounted on the microscope. In an attempt to establish such an arrangement, the author designed and manufactured a metallic sidearm of inverted L-shape in a way that the microscope optical body is attached at the lower end of the arm. The arm itself hangs from a Zeiss microscope suspension arm, thus lowering the microscope's position by 50 cm and translating to the left by 30 cm (see Figure 9.3). By employing the sidearm in such way, the microscope disappears completely from the surgeon's line of sight while the display can be brought as close as 20 cm from the surgeon, a problem that the author encountered when placing the display in Position 3 during the field experiment.



1. Star knob (with round knob): 48 mm inner diameter
2. Pivot point (with top rotation shaft) insertion into coupling (locking into holding pin)
3. Locking screw
4. Maximum load bearing: 18 kg

Figure 9.3: Schematic diagram of the microscope sidearm extension.

The arrangement could be optimised if black flaps, extended from the display boundaries, were used to limit the surgeon's peripheral vision, thus improving his/her appreciation of relative depth (see Figure 9.4). The latter could be further optimised if the stereo camera system is calibrated so that the 3D image appears to reside slightly inside the monitor, as it was found from depth mapping experiments during the Interface-Optimisation cycle of the design process.

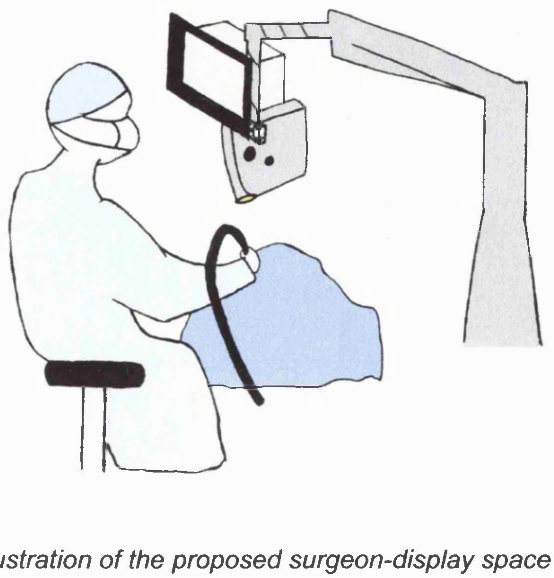


Figure 9.4: Illustration of the proposed surgeon-display space arrangement.

Unfortunately, the Sharp display used in this study could not be integrated with the sidearm due to monitor's heavy weight. Until a lighter 3D display is selected, the current prototype could be used in combination with a reflecting mirror. With this arrangement, shown in Figure 9.5, the display could be placed next to surgeon while a rotating mirror, mounted on the microscope sidearm, reflects light from the display screen towards the surgeon's eyes.

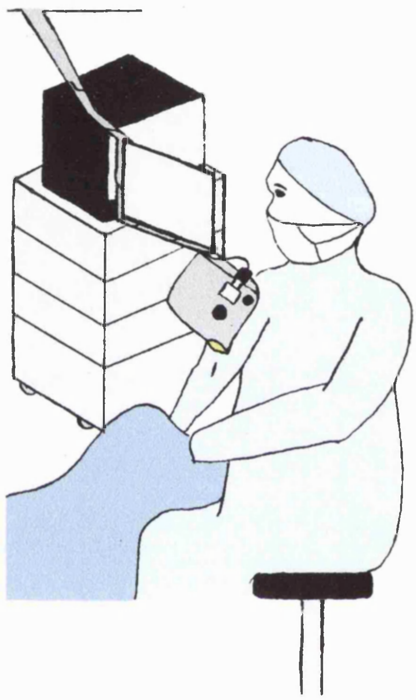


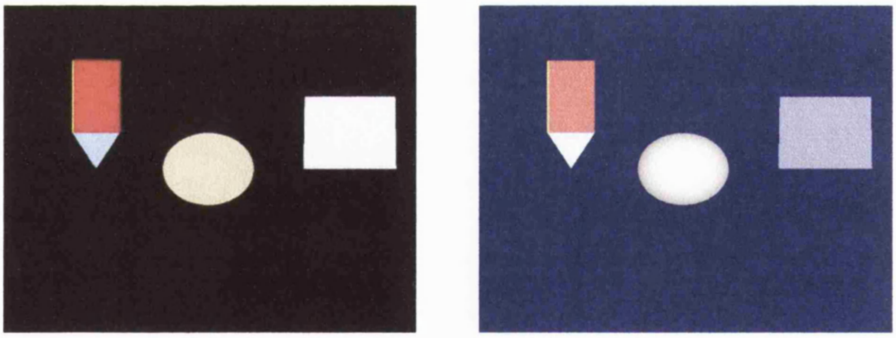
Figure 9.5: Illustration of the proposed surgeon-display space arrangement using a silver-coated reflecting mirror.

9.4 Future work

The autostereoscopic viewing interface described in this thesis was designed as part of a new computer-assisted surgical tool that aims to improve the way surgeons carry out microsurgical operations. This is done by augmenting stereo video pictures of a target under the microscope with 3D virtual reconstructions acquired from diagnostic data, and displaying the live augmented views to an autostereoscopic 3D display. In its current state, the tool cannot be fully implemented inside the operating theatre due to the lack of an accurate tracking device. As a result, the full potential of the viewing interface could not be tested during this research study. Nevertheless, the display of augmented stereoscopic images may damage the interface's effectiveness due to accommodation mismatch (not credited to image registration or camera calibration) between real and virtual objects and absence of shadow cues [128]. The magnitude of these effects should be the subject of research in follow-up studies for further interface optimisation.

Accommodation mismatch could be studied in laboratory-based phantom experiments where users are asked to make 3D measurements of the dimensions, locations, or distances between anatomical structures within the augmented microscopic scene [129].

The absence of shadow cues is inevitable in augmented scenes as it is impossible to create virtual models with realistic shadows that appear to fall on real objects, especially under complex, real world conditions. Still, shadows are very important cues in stereo video and their absence from mixed reality systems can greatly impair the performance of the surgical task. The author has identified this important factor in preliminary depth mapping studies of virtual objects (see Figure 9.6) with dimensions similar to anatomical structures found in the temporal bone region.



(a) (b)

Figure 9.6: OpenGL virtual models: a) no shadows, b) shadows and microscope lighting.

The design prototype interface discussed in this dissertation has not been fully integrated in the natural context of microsurgery. A major step towards integration would be to implement the display arrangements illustrated in Figure 9.4 and Figure 9.5 in the operating theatre, depending on the choice of 3D display. This implies conducting work in mechanical component design and evaluation of other autostereoscopic 3D displays.

A long-term research aim would be to test the interface's usability and effectiveness in other microscopy applications where depth perception is critical, such as machine vision and material mechanics [130].

9.5 Epilogue

Up to date digital product development and VR applications have been implemented focusing on system and data technology, thus leading to a high demand for user-centred interfaces. During the last decade, many world leading consumer electronics companies have invested heavily in 3D display technologies, primarily driven by considerations of product quality and cost-to-market.

Nowadays, newly created autostereoscopic 3D displays are maturing into valuable tools in human-machine interfaces and advanced HCI systems. However, opto-electronics companies are mainly interested in creating 3D displays and not 3D content [131]. As a result, the success of a viewing interface in imaging applications depends on the design and the associated process followed for a particular application. In this thesis, the author explained the scientific methodology followed for the case of ENT surgery.

Microsurgical operations incorporate a system element that is absent in the most common variety of CAS systems: the microscope. Not only must the workings of the 3D display be appropriately designed for the maximum comfort and efficiency of operators and the maximum utility of the tool, but the needs of designing an interface which does not obstruct the microscope's functionality must also be satisfied, and with the highest priority. In fact, the presence of the microscope is what necessitates the existence of the interface in the first place and as a result, the demands placed on the interface by the presence of the microscope must be considered paramount.

BIBLIOGRAPHY

- Park, David. (1997). *The fire within the eye; a historical essay on the nature and meaning of light*, Princeton University Press.
- Ruestow, Edward (1996). *The microscope in the Dutch Republic: the shaping of discovery*, Cambridge University Press.
- Banerjee, Prashant and Zetu, Dan (2001). *Virtual Manufacturing*, John Wiley & Sons Inc.
- Comfort, Alex (1984). *Reality and Empathy: Physics, Mind, and Science in the 21st Century*, State University of New York Press.
- Akay, Metin and Marsh, Andy (2001). *Information Technologies in Medicine*, John Wiley & Sons, Inc.
- Pimentel, Ken and Teixeira, Kevin (1993). *Virtual Reality: through the new looking glass*, Intel/Winderest/McGraw-Hill, New York.
- Gregory, Richard (1998). *Eye and Brain: The Psychology of seeing*, Oxford University Press (5th edition).
- Gupta, Shanti (1981). Multiple statistical decision theory: recent developments. In *Lecture notes in statistics*, Vol. 6., Springer-Verlag, New York.
- Bofinger, Eve (1987). The frontiers of modern statistical inference procedures. International conference on inference. In *Proceedings and discussions of the second international conference on inference procedures associated with statistical ranking and selection*, American Sciences Press.
- Werner, John (1990). *Visual perception: the neurophysiological foundations*, London Academic Press.
- Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*, Pinter Publishers, New York, St. Martin's Press.
- Harris, Don (1997). International conference on engineering psychology and cognitive ergonomics. In *Engineering Psychology and cognitive ergonomics*, Aldershot, Ashgate.
- Schaphorst, Richard (1996). *Videoconferencing and videotelephony: technology and standards*, Boston, Artech House.
- Inglis, Andrew (1996). *Video engineering*, New York, McGraw-Hill (2nd edition).
- Preece, Jenny (1994). *Human-Computer Interaction*, Harlow, Addison-Wesley.
- Riley, Martyn (1997). *Digital video communications*, Boston, Artech House.
- Reed, David (1992). *Residential fiber optic networks: an engineering and economic analysis*, Boston, Artech House.
- Wood, Alexander (1954). *Thomas Young, natural philosopher, 1773-1829*, Cambridge University Press.
- Duke P. J. (1990). *Modern microscopies: techniques and applications*, New York, Plenum Press.
- Baecker, Ronald (2000). *Readings in Human-Computer Interaction: toward the year 2000*, San Francisco, Morgan Kaufmann Publishers (2nd edition).
- Yanagisawa, Eiji (1997). *Colour atlas of diagnostic endoscopy in otorhinolaryngology*, New York, Igaku-Shoin.
- Shed, Donald (1999). *Historical landmarks in head and neck cancer surgery*, Pittsburgh, American Head and Neck Society.
- Woodbridge, Frederick (1983). *Aristotle's vision of nature*, Westport, Greenwood Press.

- Arbib, Michael (1986). *The construction of reality*, Cambridge University Press.
- Dowdy, Shirley (1983). Statistics for research. In Wiley (Ed) series *Probability and mathematical statistics*, New York.
- Devore, Jay (1997). *Statistics: the exploration and analysis of data*, Belmont, Duxbury Press.
- Howard, Ian (1995). *Binocular vision and stereopsis*, Oxford University Press.

REFERENCES

1. Lavallée, S., Cinquin, P., Troccaz, J. (1997). Chapter 10, Computer integrated surgery and therapy: state of the art. In *Technology and Informatics*, Vol. 30, IOS Press.
2. Lapeer, R. J., Chios, P., Alusi, G. H., Linney, A. D., Davey, M. K., Tan, A C. (2000a). Computer Assisted ENT Surgery using Augmented Reality: Preliminary results on the CAESAR project. In *MICCAI'2000 Conference Proceedings*, October 2000.
3. Lapeer, R. J., Chios, P., Alusi, G. H., Linney, A. D., Davey, M. K., Tan, A C. (2000b). Augmented reality for ENT surgery: an overview of the CAESAR project. In *ISA/ICC-2000 Conference Proceedings*, December 2000.
4. Goldstein, E. (1989). *Sensation and Perception* (3rd Edition). Belmont, California, Wadsworth Publishing.
5. Pastoor, S., Wöpking, M. (1997). 3-D Displays: A review of current technologies. In *Displays*, Vol. 17, pp. 100-110.
6. Durlach, N., Mavor, A., (Ed) (1995). *Virtual Reality: Scientific and Technical Challenges.* Washington, D.C.: National Academy Press.
7. ACM SIGCHI (1992). *Curricula for Human-Computer Interaction.* ACM Special Interest Group on Human-Computer Interaction.
8. Wilson, J. (1998). A Framework and a Context for Ergonomics Methodology. In Wilson & Corlett (Eds.) *Evaluation of Human Work: A Practical Ergonomics Methodology* (2nd Edition), London, Taylor & Francis, pp.1-40.
9. Rogowitz, B. (1985). A Psychophysical Approach to Image Quality. In *Proceedings of SPIE – Image Quality: An Overview*, Vol. 549, pp. 9-13.
10. Miller, R. (1976). The Human Task as Reference for System Interface Design. In *Proceedings of User-Oriented Design of Interactive Graphics*, pp. 97-99.
11. Gould, J., Boies, S., Lewis, C., (1991). Making Usable, Useful, Produtivity-Enhancing Computer Applications. In *Communications of the ACM*, Vol. 32(10), pp. 74-85.
12. Gibson, W. (1984). *Neuromancer*, Victor Gollanez Ltd.
13. Aukstakalnis, S. and Blatner, D. (1992). *Silicon Mirage - The Art and Science of Virtual Reality.* Berkeley, CA, Peachpit Press.
14. Carr, K. (1995). Introduction – 1.3 Themes from Simulated and Virtual realities. In Taylor & Francis (Ed) *Elements of perception*, pp. 5.
15. Heim, M. (1993). *The Metaphysics of Virtual Reality*, Oxford University Press.
16. Milgram, P., Drascic, D. (1994). Augmented reality: A class of displays on the reality-virtuality continuum. In H. Das (Ed) *SPIE Proceedings: Telemanipulator and Telepresence Technologies*, Vol. 2351, pp. 282-292.
17. Kalawsky, R. S. (1993). *The Science of Virtual Reality and Virtual environments*, Addison-Wesley (Ed).
18. Young, T. (1802). On the Theory of light and colour. In *Philosophy Transactions*, Vol. 92, pp. 12-48.
19. Fischer, F. P., Wagenaar, J. W. (1954). Binocular vision and fusion movements. In *Documenta Opthalmologica*, Vol. 7, pp. 359-391.
20. Richards, W. (1971a). Independence of Panum's near and far limits. In *American Journal of Optometry and Archives of American Academy of Optometry*, Vol. 48, pp. 103-109.
21. Julesz, B. (1971). 2. Foundations of Cyclopean Perception. In *Foundations of Cyclopean Perception*, University of Chicago Press, Chicago, pp. 14-54.

22. Shebilske, W. L., Nice, D. S. (1976). Optical insignificance of the nose and the Pinocchio effect in free-scan visual straight-ahead movements. In *Perception and Psychophysics*, Vol. 20, pp. 17-20.

23. Edwards, E. K., Rolland, J. P., Keller, K. P. (1993). Video see-through design for merging of real and virtual environments. In *Proceedings of IEEE Virtual Reality International Symposium (VRAIS '93)*, Seattle, pp. 223-233.

24. Longuet-Higgins, H. C. (1982). The role of the vertical dimension in stereoscopic vision. In *Perception*, Vol. 11, pp. 371-376.

25. Spottiswoode, R., Spottiswoode, N. L., Smith, C. (1952). Basic principles of the three-dimensional film. In *Journal of the Society of Motion Picture & Television Engineers*, Vol. 59, pp. 249-285, October 1952.

26. Lipton, L. (1982). *Foundations of the Stereoscopic Cinema: a Study in Depth*, Van Nostrand Reinhold Company.

27. Woods, A., Docherty, T., Koch, R. (1993). Image distortions in Stereoscopic Video systems. In John O. Merrit, Scott S. Fisher (Ed) *Stereoscopic Displays and Applications IV, Proceedings of the SPIE*, Vol. 1915, pp. 36-48.

28. Grinberg, V. S., Podnar, G., Seigel, M. W. (1994). Geometry of Binocular Imaging. In *Stereoscopic Displays and Applications IV, Proceedings of the SPIE*, Vol. 2177, pp. 56-65, February 1994.

29. Jones, G., Lee, D., Holliman, N., Ezra, D. (1999). Controlling Perceived Depth in Stereoscopic Images. In *Stereoscopic Displays & Virtual Reality Systems VIII, Proceedings of SPIE*, Vol. 4297 (12), pp. 42-53.

30. Nothnagle, P. E., Chambers, W., Davidson, M. W. (2001). Introduction to Stereo Microscopy. In *NICON MicroscopyU, NIKON Corp*, www.microscopyu.com.

31. Innoué, S. (1997). Solid-State detectors and cameras. In Spring & Kenneth (Ed) *Video Microscopy: the fundamentals*, (2nd Edition), pp. 278-296.

32. Masland, R. H. (1996). Unscrambling Colour Vision. In *Science*, Vol. 271, pp. 616-617.

33. Janesick, J. (1987). Sky on a Chip: The Fabulous CCD. In *Sky & Telescope*, pp. 238-242, September 1987.

34. Zworykin, V. K., Morton, A. G. (1954). *Television: The Electronics of Image Transmission in Colour and Monochrome*, (2nd Edition), John Whiley & Sons (Ed), pp. 10-26.

35. Inglis, A. *et al.* (1996). Colour Video Fundamentals. *Video engineering*, (2nd Edition), McGraw-Hill (Ed), pp. 31-46.

36. Wheatstone, C. (1838). Contributions to the physiology of vision - On some remarkable and hitherto unobserved phenomena of binocular vision. In *Philosophical Transactions of the Royal Society*, Vol. 128, pp. 371-394.

37. Kaplan, H. (1990). New jobs for liquid crystals. In *Photonics Spectra*, pp. 82-84, April 1990.

38. Walba, D. M. (1995). Fast ferroelectric liquid-crystal electro-optics. In *Science*, Vol. 270, pp. 250-251.

39. Higgins, T. V. (1995). Electro-optic displays deliver a feast for the eyes. In *Laser Focus world*, pp. 121-132, October 1995.

40. Marko, M., Leith, A. (1992). Contour-based 3D surface reconstruction using stereoscopic contouring and digitised images. In Kriete (Ed) *Visualisation in Biomedical microscopies*, pp. 45-73, VCH, New York.

41. Background on creating Images for CrystalEyes® and SimulEyes®. *Stereographics Developers' Handbook,* Stereographics Corporation, 1997.

42. Cruz-Neira, C., Sandin, D. J., DeFanti, T. A. (1993). Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In *Computer Graphics (SIGGRAPH '93 Proceedings)*, Vol. 27, pp. 135-142, August 1993.

43. Czernuszenko, M., Pape, D., Sandin, D., DeFanti, T., Dawe, G. L., Brown, M. D. (1997). The ImmersaDesk and Infinity Wall projection-based virtual reality displays. In *Computer Graphics*, Vol. 31, Number 2, pp. 46-49, May 1997.

44. Fisher *et al*, (1986). Virtual Environment Display System. In *ACM Workshop on Interactive 3D Graphics*, Chapel Hill, NC, pp. 23-24, October 1986.

45. Bos, P. J. (1993). Liquid-crystal shutter systems for real time, multi-plexed stereoscopic displays. In D. F. McAllister (Ed) *Stereo Computer Graphics and Other True 3D Technologies*, Princetown University Press, pp. 90-118.

46. Mon-Williams, M., Wann, J. P., Rushton, S. (1993). Binocular Vision in a Virtual World: Visual Deficits Following the Wearing of a Head-Mounted Display. In *Ophthal. Physiol. Opt.,* pp. 387-391.

47. Harris, M. G. (1994). Optic and Retinal Flow. In Smith and Snowden (Ed) *Visual Detection of Motion.* Academic Press, London.

48. Woodgate, G. J., Ezra, D., Harrold, J., Holliman, N. S., Jones, G. R., Moseley, R. R. (1997). Observer Tracking Autostereoscopic 3D Display Systems. In *Proceedings of SPIE*, Vol. 3012, pp. 187-198, February 1997.

49. Schwerdtner, A., Heidrich, H. (1998). The Dresden 3D Display (D4D). In *Proceedings of SPIE*, Vol. 3295, pp. 203-210, January 1998.

50. Van Berkel, C., Parker, D. W., Franklin, A. R. (1997) Characterisation and Optimisation of 3D-LCD Module Design. In *Proceedings of SPIE*, Vol. 3012, pp. 179-186, February 1997.

51. Van Berkel, C., Parker, D. W., Franklin, A. R. (1996). Multiview 3D-LCD. In *Proceedings of SPIE*, Vol. 2653, pp. 32-39.

52. Ezra, D., Woodgate, G. J., Omar, B. A., Holliman, N. S., Shapiro, L. S. (1995). New autostereoscopic display system. In *Proceedings of SPIE*, Vol. 2409, pp. 31-40, February 1995.

53. Portoni, L., Patak, A., Noirard, P., Grossetie, J., van Berkel, C. (2000). Real-Time Autostereoscopic Visualisation of 3D Medical Images. In *Medical Imaging 2000: Image Display and Visualisation, Proceedings of SPIE*, Vol. 3976, pp. 37-44.

54. Nakajima, S., Orita, S., Masamune, K., Sakuma, I., Dohi, T., Nakamura, K. (2000). Real 3D display system for use in medical applications. In Lemke *et* al. (Ed), *CARS 2000,* pp. 613-618.

55. Brooks *et al*, (1988). Using a manipulator for Force Display in Molecular Docking. In *IEEE publications*, CH2555-1, pp. 1824-1829.

56. Brooks, Frederick, P. (1986). Walkthrough-A Dynamic Graphics System for Simulating Virtual Buildings. In *ACM Workshop on Interactive 3D Graphics*, pp. 23-24, October 1986.

57. Foley, J. (1987). Interfaces for Advanced Computing. In *Scientific American*, Vol. 257(4), pp. 126-135, October 1987.

58. Satava, R. M. (1993). Virtual Reality Surgical Simulator: The First Steps. In *Surgical Endoscopy*, Vol. 7, pp. 203-205.

59. Lorensen *et al*, (1995). The Exploration of Cross-Sectional Data with a Virtual Endoscope. In *Interactive Technology and the New Paradigm for Healthcare*, pp. 221-230, IOS Press, Amsterdam.

60. Satava, R. M. (1995). Medicine 2001: The King Is Dead. In *Interactive Technology and the New Paradigm for Healthcare*, pp. 334-339, IOS Press, Washington.

61. Barfield, W., Hendrix, C. (1995). Factors Affecting Presence and Performance in Virtual Environments. In *Interactive Technology and the New Paradigm for Healthcare*, pp. 21-28, IOS Press, Washington.

62. Spiegel, E.A., Wycis, H.T., Marks, M., Lee A. S. (1947). Stereotaxic Apparatus for Operation of the Human Brain. In *Science*, Vol. 106, pp. 349-350.

63. Zinreich, S. J., Tebo, S. A., Long, D. M., Brem, H., Mattox, D. E., Loury, M. E., Van der Kolk, C. A., Koch, W. M., Kennedy, D. W., Bryan, R. N. (1993). Frameless Stereotactic Integration of CT Imaging Data: Accuracy and Initial Applications. In *Radiology*, Vol. 188, pp. 735-742.

64. Alusi, G. H., Tan, A C, Linney, A. D., Raoof, K., Wright, A. (1997). Three-Dimensional Tracking with Ultrasound for Augmented Reality Applications in Skull Base Surgery. In Grimson *et al* (Ed) *proceedings of CVRMed-MRCAS '97*, pp. 511-517.

65. Birkfellner, W., Watzinger, F., Wanschitz, F., Ewers, R., Bergmann, H. (1998). Calibration of Tracking Systems in a Surgical Environment. In *IEEE Transactions of Medical Imaging*, pp. 737-822.

66. Rohling, R., Munger, P., Hollerbach, J. M., Peters, T. M. (1995). Comparison of Relative Accuracy Between a Mechanical and an Optical Position Tracker for Image-Guided Neurosurgery. In *Journal of Image Guided Surgery*, Vol. 1, pp. 30-34.

67. Pearl, R. (1984). *Gastrointestinal Surgery for Surgeons*, Little Brown (Ed), Boston.

68. Parisien, J. (1988). *Arthroscopic Surgery*. McGraw-Hill (Ed), New York.

69. Saleh, J. (1988). *Laparoscopy*, Saunders (Ed), Philadelphia.

70. White, G., White, R. (1989). *Angioscopy: vascular and coronary applications*, Year Book (Ed), Chicago.

71. Boutin, C., Viallat, J., Aelony, Y. (1991). *Practical thoracoscopy*, Springer-Verlag (Ed), New York.

72. Knyrim, K., Seidlitz, H., Vakil, N., Classen, M. (1990). Perspectives in electronic endoscopy: past, present and future of fibres and CCDs in medical endoscopes. In *Endoscopy* Vol. 22 (Suppl 1), pp. 2-8.

73. Gilani, S., Norbash, A. M., Ringl, H., Rubin, G. D., Napel, S., Terris, D. J. (1997). Virtual Endoscopy of the Paranasal Sinuses Using Perspective Volume Rendered Helical Sinus Computed Tomography. In *Laryngoscope*, Vol. 107, pp. 25-29.

74. Vannier, M. W., Marsh, J. L., Warren, J. O. (1983). Three-dimensional Computer Graphics for Craniofacial Surgical Planning and Evaluation. In *Computer Graphics*, 17-3, pp. 263-273.

75. Sutherland, C. J., Bresina, S. J., Gayou, D. E. (1994). Use of general purpose mechanical computer-assisted engineering software in orthopaedic surgical planning: advantages and limitations. In *Comput. Med. Imaging Graph*, 18-6, pp. 435-442.

76. Linney, A. D., Campos, J., Richards, R. (1997). Non-contact Anthropometry using Projected Laser Line Distortion: Three-dimensional Graphic Visualisation and Applications. In *Optics and Lasers in Engineering*, 28, pp. 137-155.

77. Sellberg, M, Murray, D., Knapp, D., Teske, T., Lattie, K., Vanderploeg, M. (1995). Virtual Human: An Automated Virtual Environment for Computer-Aided Instruction and Biomechanical Analysis. In Morgan *et al.* (Ed) *Interactive Technology and the New Paradigm for Healthcare*, pp. 340-348, IOS Press and Ohmsha.

78. Lapeer, R. J., Chios, P., Linney, A. D., Alusi, G. H., Wright, A. (2001). HCI: Chapter XV, The Next Step Towards Optimisation of Computer-Assisted Surgical Planning, Intervention and Training (CASPIT). In Chen (Ed) *Human Computer Interaction: Issues and Challenges*, pp. 232-246, Idea Group Publishing.

79. Wright, A. (1986). Structural changes in Human Cochlea during Drug Treatment. MD Thesis, University of London.

80. Mason, T. P., Applebaum, E. L., Rasmussen, M., Millman, A., Evenhouse, R., Panko, W. (2000). Virtual temporal bone: creation and application of a new computer-based teaching tool. In *Otolaryngology Head Neck Surgery*, Vol. 122(2), pp. 168-73, February 2000.

81. Moline, J. (1998). Virtual Reality for Healthcare: a survey. In Riva (Ed) *Virtual Reality in Neuro-Psycho-Physiology*, IOS Press, Amsterdam.

82. Grimson, W. E. L., Lozano-Perez, T., Wells, W. M., Ettinger, G. L., White, S. J., Kikinis R. (1996). An automated registration method for Frameless Stereotaxy, Image-Guided Surgery, and Enhanced Reality Visualisation. In *IFFE Transactions on Medical Imaging*, Vol. 15 (2), pp. 129-140.

83. Hill, D. L. G., Maurer, C. R., Maciunas, R. J., Barwise, J. A., Fitzpatrick, J. M. (1998). Measurement of Intra-Operative Brain Surface Deformation Under a Craniotomy. In *Neurosurgery*, Vol. 43, pp. 514-528.

84. Wilson, A. (Ed) (2000). Surgeons operate with real-time imaging. In *Image Processing Europe*, pp. 17-21, November/December 2000.

85. Blackwell, M., Nikou, C., DiGioia, A. M., Kanade, T. (2000). An Image Overlay system for medical data visualisation. In *Medical Image Analysis*, Vol. 4, pp. 67-72.

86. Levy, M. L., Chen, J. C. T., Moffart, K., Corber, Z., McComb, J. L. (1998). Stereoscopic Head-Mounted Display Incorporated into Microsurgical procedures. In *Neurosurgery*, Vol. 43, pp. 392-396.

87. Edwards, P. J., Hawkes, D. J., Hill, D. L. G., Jewell, D., Spink, R., Strong A. G., Gleeson, M. J. (1995). Augmentation of Reality in the Stereo Operating Microscope for Otolaryngology and Neurosurgical Guidance. In Computer Assisted Surgery (formerly Journal of Image Guided Surgery) Vol. 1(3), pp. 172-178.

88. Edwards, P., King, A. P., Maurer, C. R., deCunha, D. A., Hawkes, D. J., Hill, D. L. G., Gaston, R. P., Fenlon, M. R., Chandra, S., Strong, A. J., Chandler, C. L., Richards, A., Gleeson M. J. (1999). Design and Evaluation of a System for Microscope-Assisted Guided Interventions (MAGI). In Taylor (Ed) *Medical Image Computing and Computer-Assisted Intervention-MICCAI '99*, Vol. 1679 of *Lecture Notes in Computer Science*, pp. 842-851.

89. Preece, J. *et al.* (1994). *Human-Computer Interaction.* Addison-Wesley (Ed).

90. Held, J. (1999). Fitting the mental model: A new technique for computerised event recording. In D. Harris (Ed), *Engineering psychology and cognitive ergonomics*, London, Ashgate.

91. Johnson-Laird, P. N. (1988). *The Computer and the Mind.* Harvard University Press, Cambridge MA.

92. Norman D. A. (1988). *The Psychology of Everyday Things.* Basic Books, NY.

93. Veer, G. *et al.* (1995). Designing Complex Systems – a Structured Activity. In Ann Arbor (Ed), Designing interactive systems (DIS '95), MA, pp. 207-217.

94. Lim, Y., Sato, K. (2001). Development of Design Information Framework for Interactive Systems Design. In *Proceedings of fifth Asian International Symposium on Design Research*, Seoul.

95. Nardi, B. A. (1996). *Context and Consciousness: Activity Theory and Human-Computer Interaction*, Cambridge, The MIT Press.

96. Carroll, J. (1998). *Making use: Scenario-based Design of Human-Computer Interactions*, Cambridge, The MIT Press.

97. Yoshikawa, H. (1987). General Design Theory and Artificial Intelligence. In Bernold (Ed) *Artificial Intelligence in Manufacturing*, Elsevier Science, pp. 35-61.

98. Sato, K., Lim, Y. (2000). Physical Interaction and Multi-Aspect Representation for Information Intensive Environments. In *Proceedings of the 2000 IEEE Workshop on Robot and Human and Human Interactive Communication*, pp. 436-443, Osaka, Japan, September 2000.

99. Helmreich, R. L. (2000). On error management: lessons from aviation. In *British Medical Journal*, Vol. 320, pp. 781-785, March 2000.

100. Bennett J. (1984). Managing to meet usability requirements. In J. Bennett, D. Case, J. Sandelin, and M. Smith (Ed) *Visual Display Terminals: Usability issues and Health concerns.* Prentice-Hall Englewood Cliffs, NJ.

101. Chios, P., Tan, AC, Linney, A. D., Alusi, G. H., Wright, A., Woodgate, G., Ezra, D. (1999). The potential use of an autostereoscopic 3D display in microsurgery. In C. Taylor & A. Colchester (Ed) *Medical Image Computing and Computer-Assisted Intervention MICCAI '99*, vol. 1679 of *Lecture Notes in Computer Science*, pp. 998-1009, Springer, September 1999.

102. Faro Arm Silver 8, *Faro Technologies Inc*, Florida, USA, 1997

103. Caliper 3D for Windows manual, (1997). Faro Technologies Inc., Florida, USA.

104. Dodgson, N. A. *et al.* (1995). Autostereoscopic 3D Display in Laparoscopic Surgery. In H. U. Lemke *et al* (Ed) *Proceedings of Computer-Assisted Radiology (CAR '95)*, Springer-Verlag, pp. 1139-1144, Berlin.

105. Woodham, J., Javer A. (1998). "Heads Up FESS" – working from the TV monitor. In *FESS Special Feature, ENT News*, Vol. 7, no. 5, pp. 19-20, November/December 1998.

106. McGrath, J. (1994). Methodology matters: Doing Research in the Behavioural and Social Sciences. In Ronald Baecker *et al* (Ed) *Readings in human-computer interaction:toward the year 2000, Chapter 2, Design and Evaluation*, (2nd Edition), pp. 152-170.

107. Van Heerbeek, N., Ingels, K. J., Rijkers, G. T., Zielhuis, G. A. (2002). Therapeutic improvement of Eustachian tube function: a review. In *Clinical Otolaryngology* Vol. 27(1), pp.50-56, February 2002.

108. Appendix F: Examples of Desirable Specialty Clinical and Practical Skills. In *The Manual of Basic Surgical Training*. Published by The Royal College of Surgeons of England, http://www.rcseng.ac.uk/surgical/trainees/sho_bst/bluebook.pdf, September 1998.

109. Sand, R., Chiari, A. (Ed), (1998). *Stereoscopic Television: Standards, Technology and Signal Processing*, European Commission, Directorate General XIII-B, Brussels.

110. Lipton, L. (1997). Developer's Handbook, In *Stereographics,* Stereographics Corporation.

111. Nielsen, J. (1992a). Finding Usability Problems through Heuristic Evaluation. In *Proceedings of CHI '92*, ACM, pp. 373-380.

112. Yeh, Y-Y, Silverstien, L. D. (1990). Limits of Fusion and Depth Judgement in stereoscopic in Stereoscopic Color Displays. In *Human Factors*, Vol. 32 (1), pp. 45-60.

113. Action Research Laboratory, (2000). *Sharp 3D Screen, Evaluation Report: Stages 1-3*, In Private Communications to Sharp Labs.

114. Blohm, W., Beldie, P., Schenke, K., Fazel, K., Pastoor S. (1997) Stereoscopic Image representation with Synthetic Depth of Field. In *Journal of the SID.*

115. Wöpking, M. (1995). Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus. In *Journal of the SID*, Vol. 3(3), pp 101-103.

116. Riley, M. *et al*, (1996). Video Compression Techniques and Standards. *In Digital Video Communications*, pp. 16-19, Artech House.

117. CCIR Rec. 601-2 (1990), Encoding Parameters of Digital Television for Studios. In *Recommendations of the CCIR*, Vol. XI - Part 1, Broadcasting Service (Television), pp 95-104.

118. Tricker, R. (1990). Optoelectronic Line Transmission. In Newnes H. (Ed.) *Introduction to Fibre Optics*, pp. 23-29.

119. Reed, D. (1991). Introducing Fibre Optics into the Subscriber Loop. In *Residential Fibre Optic Networks*, pp. 34-45, Artech House.

120. Etter, M. *et al*, (1997). Practical application of telemedicine in the clinical field. In H. M. Lemke *et al* (Ed.) *Computer Assisted Radiology and Surgery*, Elsevier Science B.V., pp. 589-593.

121. Chios, P., Tan A C (2000). Implementation of a Dedicated Optical Fibre Network for Real Time Stereo Imaging. In A. Hasman *et al* (Ed.) *Medical Informatics in Europe –MIE 2000,*

volume 77 of Studies in Health Technology and Informatics, pp. 1151-1155, August 2000.

122. Harkness, *et al*, (1995). Mastoidectomy audit: Results of the Royal College of Surgeons of England comparative audit of ENT surgery. In *Clinical Otolaryngology*, vol. 20, pp. 89.

123. Reed, D. (1991). *Developing the Engineering Cost Models*. In Residential Fibre Optic Networks, pp. 79-96, Artech House.

124. Nunez, D., Browning, G. (1990). Risks of developing an otogenic intracranial abscess. In *Journal of Laryngology and Otology*, Vol. 104, pp. 468-472.

125. Cegalis, J. A. (1973). Prism distortion and accommodative change. In *Perception and Psychophysics*, Vol. 13(3), pp. 494-498.

126. Noble, L. (1996). A Single-Chip 3D Endoscope. In SPIE: Stereoscopic Displays and Applications VII, Vol. 2653, San Jose.

127. Gerding H. (2002). Ophthalmic Clinic University of Münster, *www.dresden3d.com*, Germany.

128. Milgram, P., Drascic, D. (1997). Perceptual effects in aligning virtual and real objects in augmented reality displays. In *Proceedings of the 41st Annual Meeting Human Factors and Ergonomics Society*, Vol. 41, pp. 1239-1944, Albuquerque, NM, September 1997.

129. Kim, M., Milgram, P., Drake, J. M. (1997). Computer assisted 3D measurements for microsurgery. In *Proceedings of the 41st Annual Meeting Human Factors and Ergonomics Society*, Vol. 41, pp. 787-793, Albuquerque, NM, September 1997.

130. Garcia, D., Orteu J. (2001). 3D deformation measurement using stereo-correlation applied to experimental mechanics. In *Proceedings of the 10th FIG International Symposium on deformation measurements*, pp. 50-60, California, March 2001.

131. Law, L. (2002). Autostereo in the UK. In *Advanced Imaging Europe*, published by Cygnus, pp. 20-24, January 2002.

APPENDIX A

Following are presented raw data from all operators during the evaluation study of the new interface's accuracy and precision.

Operator 1												
pt	3D (10x)			Bins (10x)			3D (16x)			Bins (16x)		
mm	x	y	z	x	y	z	x	y	z	x	y	z
1	76.95	134.05	367.5	77.1	134.05	367.45	77.1	134.05	367.4	77	134.05	367.55
2	77.75	135.6	365.4	77.8	135.5	365.45	77.9	135.6	365.35	77.75	135.6	365.45
3	78.6	137.65	362.45	78.6	137.55	362.4	78.55	137.45	362.6	78.5	137.5	362.6
1	76.95	133.95	367.6	77.1	134	367.5	77.05	134.1	367.5	77	133.95	367.6
2	77.7	135.55	365.55	77.85	135.55	365.4	77.75	135.6	365.4	77.9	135.55	365.35
3	78.45	137.45	362.6	78.6	137.5	362.55	78.55	137.5	362.5	78.6	137.5	362.6
1	76.95	134.05	367.5	77.05	134.05	367.5	77.1	134.1	367.4	77.05	134.05	367.5
2	77.6	135.45	365.7	77.85	135.7	365.3	77.65	135.5	365.5	77.7	135.45	365.5
3	78.5	137.65	362.5	78.6	137.5	362.55	78.75	137.4	362.45	78.55	137.45	362.5
1	77.05	133.95	367.5	77.1	133.9	367.55	77	134.05	367.4	77.2	134	367.4
2	77.75	135.55	365.5	77.85	135.4	365.5	77.7	135.55	365.4	77.9	135.5	365.4
3	78.5	137.45	362.5	78.65	137.5	362.5	78.55	137.5	362.45	78.6	137.45	362.5
1	76.95	134.05	367.5	77.1	134	367.5	77	133.95	367.5	77.15	133.9	367.4
2	77.8	135.6	365.45	77.85	135.45	365.45	77.65	135.5	365.65	77.85	135.5	365.4
3	78.55	137.45	362.5	78.65	137.6	362.45	78.55	137.55	362.5	78.55	137.4	362.65

Operator 2												
pt	3D (10x)			Bins (10x)			3D (16x)			Bins (16x)		
mm	x	y	z	x	y	z	x	y	z	x	y	z
1	75.9	134.6	367.15	75.75	134.3	367.4	75.95	134.65	367.1	76.1	134.35	367.05
2	76.6	135.85	365.15	76.4	135.85	365.4	76.65	135.95	365.25	76.75	136.15	364.95
3	77.4	138	362.75	77.4	138	362.4	77.35	137.95	362.55	77.65	138.2	361.85
1	75.95	134.65	367.05	75.85	134.45	367.2	76.1	134.6	367	76.1	134.4	366.9
2	76.6	136	365.2	76.4	136.15	365.15	76.7	135.95	365.1	76.85	136.05	364.85
3	77.5	138	362.25	77.3	138.05	362.15	77.35	138.05	362.35	77.65	138.15	362
1	76	134.65	366.9	75.85	134.4	367.15	76	134.55	367	76.15	134.6	366.8
2	76.6	136.25	364.9	76.45	136.15	365.3	76.6	136.25	365	76.8	136.2	364.85
3	77.45	138.05	362.3	77.35	138.1	362.25	77.5	137.95	362.3	77.7	138.15	361.85
1	76.15	134.55	367	75.9	134.3	367.25	76	134.65	367	76.15	134.45	366.85
2	76.75	136.1	365	76.35	136.1	365.3	76.7	136.05	365.05	76.9	136.2	364.8
3	77.5	137.85	362.25	77.35	138.05	362.4	77.5	138.2	362.2	77.6	138.1	361.95
1	76.05	134.55	367.05	75.85	134.35	367.3	76	134.5	367.15	76.1	134.45	366.95
2	76.65	136.25	365.05	76.5	136.1	365.25	76.65	136.15	365.05	76.75	136.2	364.85
3	77.5	138.15	362.05	77.3	138	362.4	77.4	138.15	362.15	77.65	138.25	361.95

Operator 3												
pt	3D (10x)			Bins (10x)			3D (16x)			Bins (16x)		
mm	x	y	z	x	y	z	x	y	z	x	y	z
1	77.1	134.1	368.15	77.25	134.45	367.5	77.25	134.2	367.75	77.55	134	367.65
2	78.05	135.75	365.75	78.2	135.75	365.5	78.1	135.7	365.65	78.35	135.8	365.5
3	78.9	137.7	362.9	79.05	137.8	362.6	78.85	137.55	362.95	79.15	137.4	362.75
1	77.15	134.35	367.8	77.2	134.1	367.8	77.35	134.15	367.9	77.65	134.15	367.6
2	77.8	135.85	365.85	78.05	135.65	365.65	78.05	135.7	365.85	78.15	135.85	365.55
3	78.6	137.55	363.1	78.95	137.5	362.65	18.8	137.6	362.9	78.9	137.75	362.7
1	77	133.75	368.15	77.45	134.25	367.65	77.5	134.05	367.85	77.45	134.2	367.75
2	77.8	135.95	365.8	78.05	135.65	365.7	77.75	135.9	365.75	78	135.7	365.7
3	78.6	137.5	362.95	79	137.95	362.35	78.6	137.7	362.9	79.1	137.4	362.75
1	77.2	134.25	367.9	77.35	134.15	367.65	77.2	134.3	367.75	77.5	134.25	367.7
2	77.95	135.85	365.75	78.1	135.8	365.6	77.95	135.8	365.75	78.1	135.65	365.8
3	78.85	137.85	362.75	78.9	137.8	362.5	78.55	137.65	362.85	78.9	137.65	362.8
1	77.05	134.2	368.1	77.4	134.2	367.7	77.3	134.15	367.9	77.4	134.15	367.65
2	77.9	135.65	365.8	78.1	135.7	365.6	77.9	135.75	365.95	78.15	135.6	365.55
3	78.65	137.75	362.85	78.85	137.75	362.5	78.75	137.5	363.05	78.95	137.65	362.6

Operator 4												
pt	3D (10x)			Bins (10x)			3D (16x)			Bins (16x)		
mm	x	y	z	x	y	z	x	y	z	x	y	z
1	76.2	134.4	367.7	76.1	134.55	367.7	76.75	133.9	367.7	76.5	134.2	367.75
2	76.65	135.95	365.85	76.85	136.05	365.85	77.3	135.65	365.9	77.05	135.45	365.9
3	77.65	137.95	363.05	77.5	138	363.05	78.25	137.45	362.95	78	137.35	363.15
1	76.15	134.4	367.8	75.9	134.55	367.55	76.65	134.15	367.6	76.35	133.8	368.1
2	76.8	135.9	365.85	76.65	136.05	365.75	77.3	135.55	365.85	77.05	135.5	366
3	77.55	138.25	362.85	77.55	138.1	362.9	77.95	137.4	363.05	78	137.4	363.25
1	75.85	134.2	367.75	76.05	134.5	367.7	76.7	134.2	367.6	76.5	134.1	367.9
2	76.75	135.75	365.85	16.55	136.15	365.75	77.15	135.6	365.75	77.1	135.5	366.1
3	77.35	137.9	362.85	77.55	137.85	363.1	78.05	137.65	362.95	77.95	137.6	363.05
1	76.05	134.45	367.7	76.1	134.7	367.65	76.55	133.9	367.75	76.5	134.1	368.05
2	76.85	135.85	365.9	76.65	136.25	365.75	77.15	135.6	365.85	77.15	135.65	365.95
3	77.45	137.85	363.1	77.7	137.95	363	78.1	137.55	362.9	78.05	137.5	363.1
1	76.2	134.55	367.75	76.1	134.65	367.55	76.8	134	367.8	76.55	134.1	367.8
2	76.7	135.85	365.95	76.7	136.05	365.8	77.3	135.4	365.95	77	135.4	366
3	77.55	138.2	362.7	77.6	137.9	363.05	78.15	137.45	362.95	77.8	137.5	362.95

Operator 5												
pt	3D (10x)			Bins (10x)			3D (16x)			Bins (16x)		
mm	x	y	z	x	y	z	x	y	z	x	y	z
1	76.95	133.5	367.75	77.1	133.5	367.95	76.75	133.45	368.3	76.95	133.5	36.1
2	77.5	135	365.9	77.75	135	366.05	77.2	165.05	366.45	77.45	135.25	366.2
3	78.25	136.75	363.1	78.7	136.85	363.15	18.05	137.05	363.45	78.45	136.8	363.55
1	77.05	133.6	368	76.85	133.25	368.1	76.7	133.5	368.35	76.95	133.8	368.05
2	77.75	135.35	365.8	77.5	134.9	366.2	77.15	135.15	366.3	77.25	135.2	366.1
3	78.5	137.05	363.1	78.2	136.95	363.35	78.1	137	363.85	78.2	136.95	363.25
1	77.25	133.55	367.9	76.8	133.35	368.15	76.65	133.45	368.25	77	133.35	368.3
2	77.6	135.2	366	77.45	135	366	77.05	135.15	366.45	77.3	135.5	366.05
3	78.3	137	363.2	78.35	136.85	363.25	78.3	137	363.4	78.15	136.65	36.7
1	77.05	133.6	367.9	76.8	133.35	368.3	76.75	133.65	368.15	76.55	133.85	138.3
2	77.45	134.9	366.15	77.4	134.95	366.05	77.25	135	366.3	77.45	134.75	366.5
3	78.5	136.9	363.3	78.5	136.95	363.25	78.05	136.85	363.5	78.15	137.45	363.4
1	77	133.6	367.95	77.15	133.5	368	76.7	133.45	368.3	16.9	133.45	368.2
2	77.55	134.95	366.1	77.65	134.8	366.05	77.15	135.1	366.2	77.65	135.2	366.25
3	78.5	136.85	363.25	78.65	136.85	363.15	78.15	136.9	363.25	78.45	137.25	363.4

Operator 6												
pt	3D (10x)			Bins (10x)			3D (16x)			Bins (16x)		
mm	x	y	z	x	y	z	x	y	z	x	y	z
1	76.25	134.15	367.65	76.5	134.2	367.05	76.1	134.25	367.55	76.8	134.1	367.35
2	76.85	136.05	365.45	77.15	135.85	365.3	77.2	135.85	365.45	77.55	135.75	365.3
3	77.5	137.6	363.15	77.8	137.9	36.2	77.75	137.7	362.8	78.1	137.95	362.25
1	76.2	134.45	367.35	76.55	134.15	367.3	76.1	134.4	367.55	76.75	134.2	367.2
2	76.8	135.95	365.6	76.95	135.8	365.1	76.95	135.9	365.65	77.3	135.8	365.2
3	77.6	137.6	363.2	78.25	137.75	362.5	77.8	137.8	362.7	78.15	137.9	362.25
1	76.1	134.25	367.45	76.65	134.2	367.15	76.3	134.15	367.55	76.65	134.35	367.05
2	76.8	135.8	365.55	77.3	135.7	365.25	77.2	135.85	365.45	77.25	135.75	365.2
3	77.5	137.45	363.35	78.25	137.85	362.35	77.8	137.95	362.65	78.1	137.95	362.35
1	76.65	134.1	367.25	76.85	133.95	367.25	76.25	134.55	367.35	76.9	134.1	367.2
2	76.8	135.8	365.55	77.15	135.9	365.15	76.9	135.9	365.55	77.3	135.85	365.1
3	77.5	137.75	363.2	78.3	137.9	362.25	77.9	137.7	362.65	78.2	137.9	362.25
1	76.2	133.95	367.75	76.6	134.35	366.9	76.3	134.25	367.55	76.65	134.15	367.1
2	77.25	135.75	365.6	77.15	135.85	365.2	77.2	135.9	365.65	77.25	135.8	365.2
3	77.5	137.45	363.4	78.3	137.75	362.4	77.8	137.9	362.6	78.4	137.7	362.4

Operator 7												
pt	3D (10x)			Bins (10x)			3D (16x)			Bins (16x)		
mm	x	y	z	x	y	z	x	y	z	x	y	z
1	75.9	133.95	367.55	76.15	134.75	368.2	75.95	134.45	367.75	76.25	134.45	367.6
2	76.35	135.45	365.6	76.7	135.9	366.55	76.6	135.95	366.15	76.75	135.95	366.05
3	77.4	137.65	362.8	77.55	138	363.15	77.45	137.8	362.95	77.55	137.75	363
1	76	134.3	367.5	76.25	134.6	367.85	76	134.2	367.95	76.05	134.25	367.8
2	76.35	135.55	365.65	76.95	135.95	365.7	76.6	135.7	366	76.7	135.8	365.9
3	77.3	137.45	362.85	77.7	138.05	363	77.2	137.75	363.15	77.4	137.8	363.3
1	75.95	134.65	368.55	76.05	134.45	368.85	76.15	134.4	367.65	76.3	134.45	367.85
2	76.65	136.1	366.5	76.75	136.05	366.2	77.4	137.85	363.6	76.65	135.8	365.9
3	77.6	138.1	363.25	77.65	138.05	363.1	77.6	137.9	362.9	77.45	137.9	363.15
1	76.1	134.7	367.8	76.2	134.5	367.85	76	134.3	367.85	76.1	134.35	367.85
2	76.9	136.05	365.7	76.85	136.1	365.95	76.8	135.85	366	76.7	135.85	366
3	77.65	138	363.2	77.7	137.95	363.2	77.6	137.8	362.95	77.45	137.65	362.95
1	76.3	134.5	367.45	76.35	134.5	367.75	76.1	134.4	367.7	76.1	134.1	368.05
2	76.7	135.9	365.7	76.75	135.9	365.95	76.7	135.8	365.9	76.65	135.75	365.85
3	77.4	137.95	363.9	77.9	138.05	363.35	77.65	137.7	363.15	77.6	137.75	362.85

APPENDIX B

Following are the Viewing Comfort and DOF matrices from which a Kruskall-Wallis 1-way ANOVA test examines the effectiveness of the viewing interface.

Response / Disparity	Viewing Comfort (Series I) Much less comfortable	Less comfortable	Comfortable	More comfortable	Much more comfortable
ux_1.5mm	9	1	0	0	0
ux_1.2mm	4	5	1	0	0
ux_0.9mm	1	5	2	1	1
ux_0.6mm	0	2	3	2	3
ux_0.3mm	0	1	2	6	1
default	0	0	10	0	0
x_0.3mm	0	3	4	2	1
x_0.6mm	2	5	2	1	0
x_0.9mm	6	3	1	0	0
x_1.2mm	7	2	0	1	0
x_1.5mm	8	1	1	0	0

Response / Disparity	Viewing Comfort (Series II) Much less comfortable	Less comfortable	Comfortable	More comfortable	Much more comfortable
ux_1.5mm	5	0	2	1	2
ux_1.2mm	1	3	1	2	3
ux_0.9mm	0	3	2	2	3
ux_0.6mm	0	2	2	2	4
ux_0.3mm	0	3	3	3	1
default	0	0	10	0	0
x_0.3mm	0	2	5	1	2
x_0.6mm	0	3	3	2	2
x_0.9mm	3	4	0	3	0
x_1.2mm	3	5	1	1	0
x_1.5mm	6	3	0	1	0

Response	DOF (Series I)				
Disparity	Very annoying	Annoying	Normal	Better	Much better
ux_1.5mm	7	3	0	0	0
ux_1.2mm	2	8	0	0	0
ux_0.9mm	2	7	1	0	0
ux_0.6mm	0	1	6	1	2
ux_0.3mm	0	1	5	4	0
default	0	0	10	0	0
x_0.3mm	0	3	3	4	0
x_0.6mm	3	2	3	2	0
x_0.9mm	5	1	2	2	0
x_1.2mm	6	0	1	3	0
x_1.5mm	7	1	1	1	0

Response	DOF (Series II)				
Disparity	very annoying	annoying	normal	better	much better
ux_1.23mm	2	4	1	1	2
ux_0.98mm	0	2	1	4	3
ux_0.74mm	0	2	1	4	3
ux_0.49mm	0	4	2	2	2
ux_0.24mm	0	2	2	5	1
default	0	0	10	0	0
x_0.24mm	0	1	3	4	2
x_0.49mm	0	2	2	4	2
x_0.74mm	2	1	2	4	1
x_0.98mm	4	0	1	5	0
x_1.23mm	4	0	1	5	0

APPENDIX C

NETWORKS

Networks are normally categorised as communicative and distributive. For example, a standard telephony network is a narrowband interactive service while cable television is a broadband distributive service. The transmission of broadband television services is called *broadcasting*. Television networks use broadcasting to transmit live colour pictures to millions end users. Broadcasting channelling can be single-mode or multi-mode, permitting more than one image to be transmitted simultaneously. More demanding systems, require data exchange and image display capabilities between different endpoints. When every endpoint can reach each of the others the system is called *network*. To be maximally useful, each end must be able to provide data to any other end.

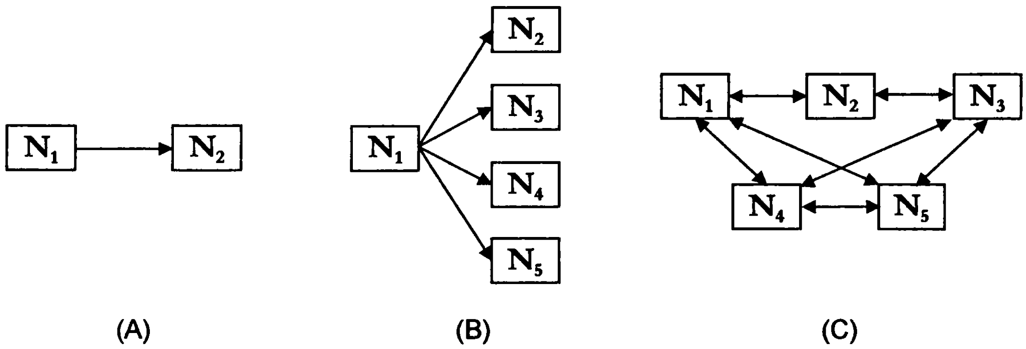


Figure C.1: Single link system (A), Multipoint system (B) and Network system (C).

Network Design

The designing considerations of a service network include theoretical and practical aspects of signal transmission. The most important ones are attenuation loss and usable bandwidth. Other parameters influencing the system are defined through the type of application (*e.g.* indoor or outdoor, buried, aerial or sub-terrestrial), the transmission properties (*e.g.* single or multi-channel transmission), the climatic conditions (*e.g.* temperature, pressure, tension, etc.) and the local environmental restrictions (*e.g.* usage of jacketed cable).

The network architecture is subject to the significance of the network service. There are three fundamental architectural designs, the *star*, the *double-star* and the *bus*. In the star design, a single cable is run from the central location to each terminal point. This is most suitable for high-capacity connections since the bandwidth of the cable is not shared with any other user. The presence of 'dedicated' links allows efficient delivery of switching services. The double-star design uses Remote Distribution Units (RDU) with switching techniques to reduce the amount of cabling by sharing this plant over more users. Further out from the RDU, the design is identical to the star one. Such a design is very convenient for mixed services, because the shared cable carries the distributive information to the RDU. Finally, in the bus design a series of switchers is used to distribute services as each user receives the same signal.

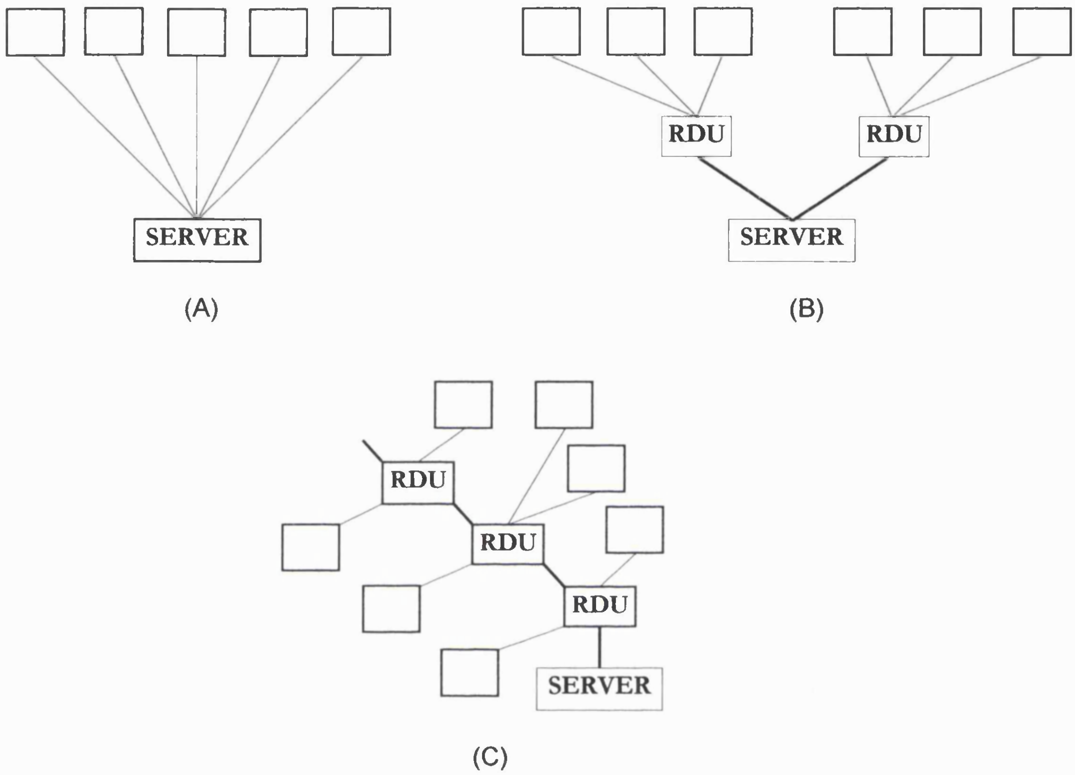


Figure C.2: (A) Star Design, (B) Double-Star Design and (C) Bus Design.

With respect to designing an analogue or digital network one has to look closely at the physical characteristics of the signal transmitted. The transmission of a live video image can be either done in an analogue form, with the three colours transmitted separately, or in a digital form after the analogue signal has been converted to the CCIR 601 format. In general, analogue systems are only attainable for low quality images. This is due to the fact that the active bandwidth and the linearity of the transmission are very poor. On the other hand, digital transmission designs only suffer from high permissible attenuation between the transmitter and the receiver and the average received power necessary to achieve a certain bit error rate. High quality compressed live video data normally occupy up to 40 Mbps (for a 50:1 compression ratio), thus requiring a broad enough bandwidth of transmission. Similar non-coded data can occupy up to 270 Megabytes per second (MBps).

Network designs for systems that perform high quality, live and real time, video transmission usually tend to use fibre optics. For transmissions up to 140 Mbps, graded-index fibres are used, while for images exceeding 140 Mbps, single-mode fibres are used as they provide the bandwidth required for the regenerators. Additionally, laser diodes are used in order to keep the spectral bandwidth low. In general, network designs are usually justified upon their *openness.* An open system is accounted for its ease to use, ease to change and ability to serve a wide variety of resources that are in use concurrently.

Information Technology has entered in numerous working environments, and so has in medicine. The need for direct communication between General Practitioners, hospital personnel and consultants is inevitable. Exchange of diagnostic data and therapeutic techniques has to be reliably carried out between all parties involved with the patient's treatment.

At present, telecommunication data transfer relies on a network with given capabilities. Networks are divided in three major categories: *Local Area Networks* (LAN), *Metropolitan Area Networks* (MAN) and *Wide Area Networks* (WAN). WAN and MAN are used for long distance transmissions while LANs are typically used for small number of terminals. There are different systems, with different capacities and characteristics, that can build a LAN, MAN or WAN. The main ones used are listed in the following table:

Ethernet	10 Mbps	LAN
Fast Ethernet	100 Mbps	LAN
FDDI	100 Mbps	LAN
ATM	64 kbps - 622 Mbps	LAN / WAN
ISDN	64 kbps - 2 Mbps	WAN
xDSL	2 Mbps	MAN
Satellite	40 Mbps (broadcast)	MAN / WAN
Copper Networks	40 Mbps (broadcast)	MAN
Fibre Optics	2 Gbps	LAN / WAN

Table C.1: Network speed rates.

The Internet is a global communications network for interconnecting computers by allowing them to communicate even when they reside on different types of sub-networks. One of the application services is the World Wide Web (WWW). In conjunction with other services such as email, file transfer and news, the WWW allows users to access data and services on computers throughout the Internet. Information in the WWW are stored in WWW servers. Data and applications are extracted from every connected server with the use of specific Internet protocols. Information is transmitted in Transmission Control Protocol (TCP) packets, ensuring the transmission of a stream of bits form the transmitter to the receiver. The bit stream is delivered at the receiver in the same order that they were transmitted.

Audiovisual data that contain heavy amounts of information are very difficult to move around the Internet and usually cause traffic. Transmitting traffic form one server to many terminals is known as *muticasting.* Even after large image compression and video coding, audiovisual data generate traffic of up to 500 Kbps during broadcasting and occupy most of the bandwidth. Resulting live video broadcasts display poor image quality and frame rate.

In a hospital environment, a network would be used to transmit diagnostic images from one department to another, live video images from the operating theatre to the lecture theatre, DICOM 3 information, other computer data, as well as audio data transfer between medical specialists. To avoid traffic, experts tend to use fibre optic cables for data transfer.

- FIBRE OPTICS NETWORK DESIGN - ROYAL NATIONAL ENT HOSPITAL, LONDON –

The author designed an optical fibre network system for the complete communicative and broadcasting demands of augmented reality microsurgery in the Royal National ENT Hospital, London. Single-mode fibre links were dedicated for the transmission of the live optical and augmented video data. Separate links were dedicated for the transmission of the mouse, keyboard and tracking communicative components. The author interconnected three major sites within the hospital: the operating theatre, the Institute of Laryngology & Otology computer laboratory and the hospital's lecture theatre. The traditional "star" design for which all the end points are interconnected through a single location (*i.e.* the server) was used. A direct audio and composite video link between the operating theatres and the lecture theatre was also created. The design dedicated two links, one for transmitting two composite video signals acquired from video cameras inside the operating theatre and a second one, for the bi-directional transmission of two intercom audio connections between the surgeon in the operating theatre and the attendants in the lecture theatre. This work was carried out independently from the augmented reality microsurgery project but the configuration allows possible future usage of the audio and composite video distributive data.

The author used a 24-core fibre optic cable for the connections between the operating theatres, the server and the computer laboratories. In addition, a 16-core fibre optic for the connection between the lecture theatre and the server was used. The lecture theatre connection has two endpoints: one at the audiovisual control room (otherwise called "the booth") where the audio/video console is found, and a second one at the lecture theatre's main stage where the co-ordinator can manage or present the whole system operation. A 12-core fibre optic cable was dedicated for the connection to the booth and a 16-core for the connection to the stage. With this configuration, the speaker can access directly software applications and viewing conditions. It is believed that by doing that the whole procedure becomes more interactive, with the audience actively participating in the surgical procedure.

The operating theatres are all connected to the surgeons main office, found on the same floor and at a distance of 15 metres from each theatre. The designer used additional 24-core fibre optic cable from the office to each theatre. The office has a fibre optic '*hub*' which allows each theatre to be connected to the computer mainframe. The hub allows only one theatre at the time to be connected to the SGI Onyx2 server. All three fibre optic cables run over the false ceiling found on the theatres floor, together with the main fibre optic cable that runs from the operating theatres floor to the computer mainframe. The following diagrams illustrate the cable configuration at the theatre floor and in the lecture theatre.

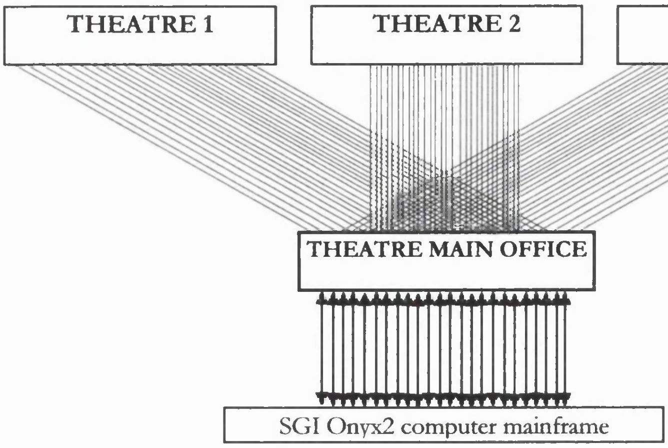


Figure C.3: Operating theatres fibre optic network design.

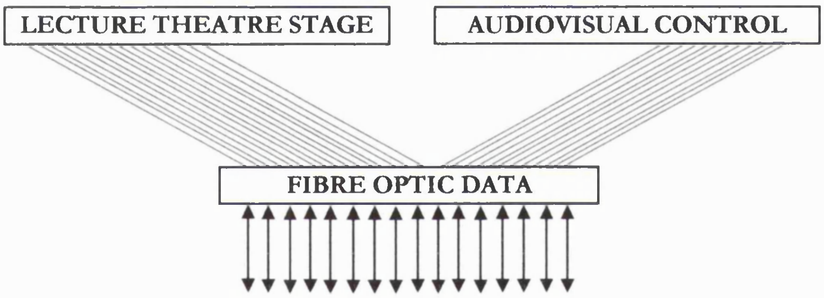


Figure C.4: Lecture theatre fibre optic network design.

- <u>CABLE ARRANGEMENT OF OPTICAL FIBRE PATCH LEADS FOR THE ONYX2-COMPUTER LABORATORY CONNECTION</u>

The Onyx2 Room

1. Serial Digital, Camera to Onyx2: Left Eye
2. Serial Digital, Camera to Onyx2: Right Eye
3. Serial Digital, Onyx2 to Camera: Left Eye
4. Serial Digital, Onyx2 to Camera: Right Eye
5. TTL Genlock, Onyx2 to Sync Generator: Left Eye
6. TTL Genlock, Onyx2 to Sync Generator: Right Eye
7. RS232 Tracking, Polhemus to Onyx2: Receiver
8. RS232 Tracking, Polhemus to Onyx2: Transmitter
9. RS232 Keyboard, Onyx2 to SGI: Receiver
10. RS232 Keyboard, Onyx2 to SGI: Transmitter
11. RS232 Mouse, Onyx2 to SGI: Receiver
12. RS232 Mouse, Onyx2 to SGI: Transmitter
13. RGB Monitor, Red to Screen: Transmitter
14. RGB Monitor, Green to Screen: Transmitter
15. RGB Monitor, Blue to Screen: Transmitter

In the diagram below, the arrangement of the fibre optics patch panels is configured in such way so that the panel at the top corresponds to the connection with the lecture theatre, the middle one to the computer room and the bottom one to the connection with the operating theatre.

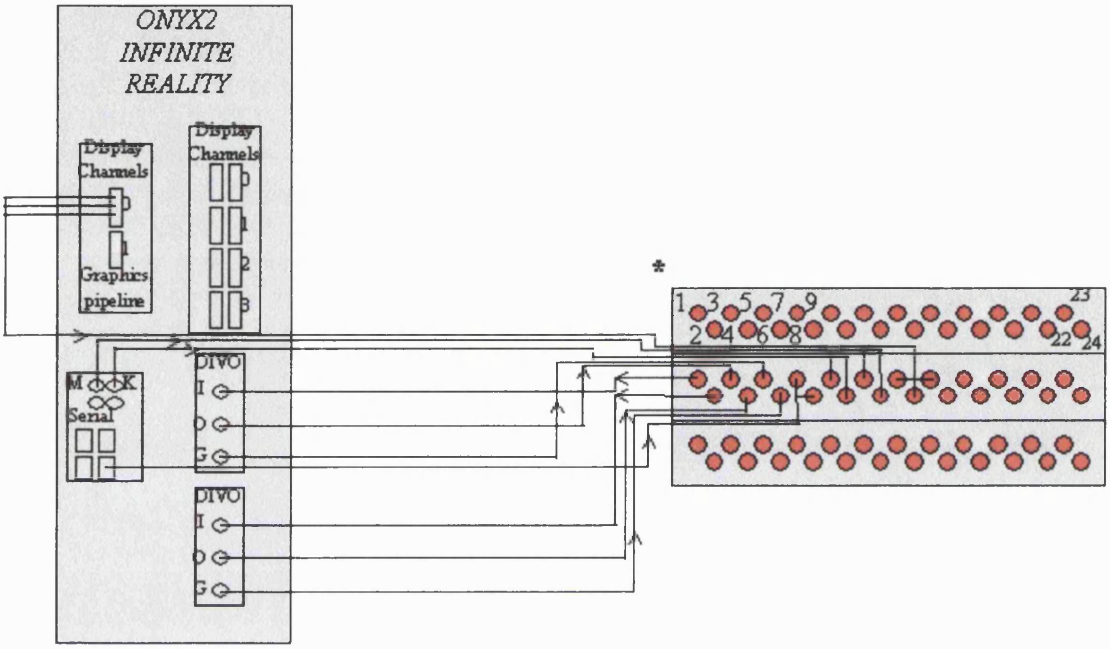


Figure C.5: I/O cabling arrangement of the server-panel connection.