Abstract

This thesis describes a technique to establish live feedback to the image-guided neurosurgery environment, in the form of stereoscopic video scenes from the operating room. The video scenes are superimposed on images presently available in image-guided neurosurgery. The effect of displaying these superimposed images stereoscopically is that the viewer's natural, binocular vision appears to penetrate the patient's head, and to reveal internal anatomy.

Brief histories of conventional stereotactic, and frameless image-guided neurosurgery are provided. Relevant capabilities and limitations of the neurosurgical environment into which video is introduced are described, as well as the development and description of the pertinent video technology. Next, the mathematics of projection imaging, and principles of stereoscopic imaging are discussed. The role of stereoscopic video in the operating room is reviewed, and a technique for merging magnetic resonance images (MRI) and video stereoscopic-pairs is presented. The performance of the developed system is evaluated using a phantom and human subjects.

Résumé

Cette thèse décrit une technique qui fournit au système de neurochirurgie guidée par images électroniques, un feed-back continuel, sous la forme de deux scènes vidéos en stéreo de la chambre d'opération. On supèrpose les scènes vidéos sur des images qu'on emploie présentement dans le système de neurochirurgies guidée par images électroniques. La superposition de ces deux genres d'images en stéreo donne l'effet que, la vision binoculaire naturelle de l'observateur pénétre la tête du patient et expose la structure anatomique intèrne.

On présente un bref recit de l'histoire de la neurochirurgie guidée par images électroniques, stéreotactique et sans cadre. On décrit les capacitées et limitations du système neurochirurgical, dans lequel on introduit les scénes vidéos, ainsi que la déscription et le dévelopement de la téchnologie de vidéo pértinente. Ensuite, on discute la formulation mathématque de l'imagerie en projection et les principes de stéreoscopie. On met en question le rôle des scénes vidéos en stéreo dans la chambre d'opération, et on présente une technique de fusioner des images de résonance magnetique (IRM) avec des scénes vidéos en stéreo. On mesure la performance du système dévelopé sur des données artificiélles et humaines.

Acknowledgements

I sincerely thank Dr. T. M. Peters for his constant encouragement and enduring patience throughout the course of this work. I trust his keen insight and deeply appreciate his infectious enthusiasm for the subject and for his students' achievements.

I am very grateful for the members of the NIL group whose comraderie and thought provoking interchange is already dearly missed. I especially thank Mr. Majd Bakar, Dr. Bruce Davey, and Mr. Nurruddin Bohari for teaching me to program. Very special thanks to Mr. Roch Comeau, the video model, for the use of his beautiful head, and for so generously sharing his code along with innumerable hours of his time.

I thank Dr. Chris Henri and Mr. Dave McDonald for their essential and timely collaboration.

I would like to acknowledge the support of ISG in providing the IAP environment, and to thank Dr. E. B. Podgorsak for the use of the OBT-stereotactic frame.

Finally, I would like to thank my family and my dear friends, Mr. Robert D. Flanz and Ms. Erlian Lu (my rock) for their support.

Preface

This work describes a visualization system that allows the superposition of video data, recorded during a neurosurgical procedure, with the 3-D rendered model of the brain. This can be seen as augmented reality, in terms of adding otherwise invisible anatomical information to a video image. Originally, the approach of this thesis was to derive 3-D information from stereoscopic-video [11], then to relate that information to a 3-D anatomical model. Subsequently an important advantage of independently relating each view of a stereoscopic pair to 3-D model-space has become apparent: no calibration information about the video cameras is needed. This is advantageous in that the cameras used in the system can be adjusted or switched without need for recalculating calibration parameters. By registering video images to image guided neurosurgery (IGNS) space, and displaying them stereoscopically, the orientations of visible objects such as surgical instruments are established relative to internal anatomy. In addition, planned craniotomies, drawn on the patient skin prior to surgery can be displayed relative to underlying anatomy. Also, the surgeon can identify within the 3-D model a planned resection so that while the procedure is in progress, the video allows visual comparison of the pre-surgical plan with the actual tissue removed.

A further application of stereoscopic video, is to alert the surgeon to significant distortions or shifts that may have occurred in the cortical surface during the surgical procedure.

Chapter One provides a perspective on the technologies relevant to this project, with special emphasis on the historical developments which have brought them to their present states. The second chapter describes the present application of these technologies at the Montreal Neurological Institute and the limitations of that implementation which are addressed by the incorporation of stereoscopic-video. Chapter Three discusses video technology today and applications of stereoscopic-video in industry and medicine. The geometry of different projection imaging schemes, and of stereoscopic imaging are explained in Chapter Four. Results of preliminary validation experiments using this system are described in Chapter Five, and concluding remarks are presented in Chapter Six.

Original contributions of this work include constructing a video fusion program which solves a set of linear equations for the transformation from a mechanical probe based coordinate system to a video camera centred system. Using this transformation, a graphical representation of the probe overlaid onto the video reflects the position of the mechanical probe in near real time. Results and discussion of an experiment which demonstrates how the effect of rendering by parallel projection, as compared to perspective projection, depends on imaging geometry are also included. And, using tools from an image manipulation and display software package, a computer interface was created which allows segmented MRI data to be manually aligned with a video image.

Contents

List of Figures

Chapter 1

Introduction

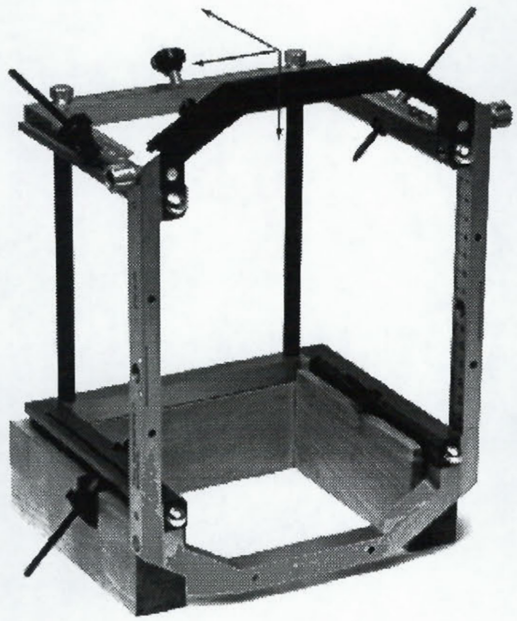
1.1 Conventional Stereotactic Surgery

Stereotaxy Defined

The term stereotaxy was first used by Horsley and Clarke [26] and refers to surgical techniques involving repeatable, tridimensional localization by means of a precision probe-guiding apparatus keyed to a brain atlas or image [15]. The word comes from the Greek, *stereos* meaning 'solid' or 'three-dimensional' and *-taxis* meaning 'arrangement' or 'order'. Stereotaxy has been used to:

- place electrodes within the brain for EEG stimulation or recording,
- approach deep seated brain lesions with a probe for biopsy, or estimate the volume of internal structures,
- Guide source placement for interstitial brachytherapy,
- localize targets for treatment by a high energy ionizing radiation beam.

The goals of stereotactic surgery are a high degree of accuracy, precision, and maximum flexibility of trajectory selection. Because stereotaxy gives the impression of



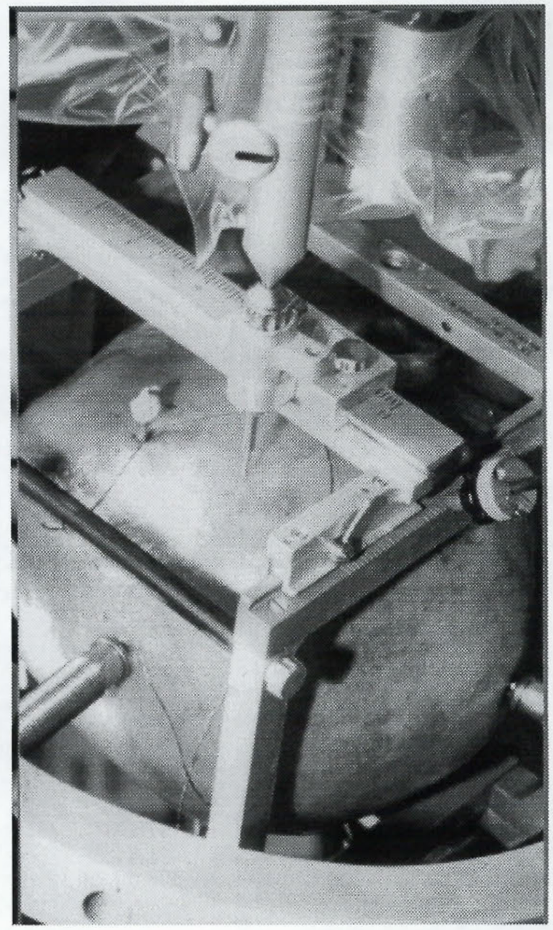


Figure 1.1: Left: the OBT-stereotactic frame.

Figure 1.2: Right: Stereotactic frame used for orientation relative to a patient head.

being able to touch brain structures, the adjective *stereotactic* from *stereos* and *-tactic*, 'to touch', is often used in place of *stereotaxic*.

Development

In 1873, Dittmar first wrote about using a device for neurophysiological orientation, but it was specialized for work on the medulla oblongata [19]. By 1908 Horsley and Clarke had developed what is considered the first stereotactic apparatus, to be used on monkeys, and introduced the concepts of a stereotactic atlas and electrolytic lesioning of nervous tissue [26]. Orientation in their system was based on skull landmarks.

Spiegel and Wycis developed a practical device for stereotaxy on humans in 1947. This apparatus, or stereotactic frame, was improved and used throughout the 1950s and early 1960s primarily for treating patients with Parkinson's disease. Ventriculograms were used to identify internal landmarks as ordinates to the patient anatomy by showing radiographic images of the individual patient's ventricular system. When an alternative treatment for Parkinson's patients was introduced, stereotactic surgery became less common. The introduction of computed tomography(CT) in 1973 by Hounsfield et al. [27] would lead to a resurgence in the popularity of stereotaxy, and Bergstrom [1] first combined CT technology with a stereotactic apparatus in 1976. This new aspect of stereotaxy increased its usefulness and prevalence, especially in intracranial mass lesion biopsy procedures [19]. Because the stereotactic frame is visible in the CT images, it was eventually used, rather than anatomical landmarks, to define the stereotactic workspace. More extensive reviews of the history of stereotaxy have been covered by Kelly [30], Maciunas [33], and Thomas [53].

1.2 Development of Frameless Image-guided Neurosurgery

In a frameless image-guided neurosurgical(**IGNS**) procedure, the space defined by the physical frame is replaced with a space defined by the motion of an articulated arm. Image-space is mapped to this surgical space by identifying anatomically relevant points on the patient with the articulated arm, and in the images, and then finding a transformation to relate the two spaces. These points of commonality are called **fiducial marks**. The mapping of the physical and image spaces is referred to as their **registration**. Using this approach, surgeons employ pre-recorded images of the patient anatomy for guidance during a procedure rather than orienting themselves relative to a stereotactic frame.

This system is useful when surgery targets are deep within the head, typically for depth-electrode positioning, tumor resection, and brain tissue removal for intractable epilepsy. In these procedures the approach path often passes close to important vessels or functional areas in the brain. The complications specific to neurosurgery which are simplified by frameless image-guidance include:

- portions of the skull must be identified for removal to allow adequate surgical access, without their extent becoming excessively large.
- often, critical functional areas or vessels limit the choice of trajectories to the target,
- brain tissue is very delicate so it is important to leave the vasculature supplying blood to healthy areas intact.
- tissue to be resected is not always visually distinguishable from the healthy tissue. [17]

History

Roberts [42] was one of the earliest investigators to propose a frameless image-guidance system for neurosurgery. His system integrated CT images with an operating microscope and was first presented in 1986. That image-guidance system has developed into a frameless stereotactic operating microscope where the improved version was described by Friets et al.[16] in 1989. This oriented surgeons with respect to internal anatomy by combining pre-recorded anatomical images with the normal microscope view. The system did not require a frame for image registration, neither did it include an interactive probe. Kato et al.[29] presented the CNS NAVIGATOR, a frameless, armless navigational system for use in image-guided neurosurgery. The CNS NAVIGATOR implements magnetic field modulation technology to track the position of a resin probe. The spatial data of the probe is translated onto the preoperative images using four fiducial markers. In 1992 Galloway [17] published results of frame-independent image-guidance experiments, also known as frameless stereotaxy.

1.3 Video in Medicine

The use of video as a medical imaging modality is growing in popularity and in variety of application. Clinical applications include both diagnostic and therapeutic modalities, and video is also exploited in medical research. Video endoscopy and video in x-ray fluoroscopy are the current clinical standards.

Diagnostic application of video includes video-colonoscopy, detection and distinction of minute lesions in video microscopy, and collaboration through video conferencing for remote hospitals. Video-EEG monitoring is also frequently used for observation of epilepsy patients during EEG recording [13, 52, 8].

Since 1990 videothoracoscopy (endoscopy in the thoracic cavity), sometimes combined with conventional thoracotomy, has helped to minimize trauma from surgical techniques [46]. By using video to avoid full thoracotomy, postoperative pulmonary function is less restricted, there is less pain, earlier mobilization is possible and the period of hospitalization is reduced. In the majority of procedures the operative time is reduced as compared with open surgery [44].

Improved light sources, lasers and video capability have made laparoscopy (abdominal endoscopy) advantageous to both the surgeon and the patient [5]. Abdomino-pelvic contents can be visualized directly through a laparotomy incision or indirectly with a laparoscopic video display to allow further target definition and minimize complications associated with previously blind procedures such as after hysterectomy or cholecystectomy [12].

In research, video is often applied to quantify aspects of a system, either of human kinematics, or in growth or morphological change, sometimes through a microscope. Eye movement or gait analysis are two examples of uses of video in human kinematics. Image-analyzers are used in research to quantify cyst size, vasoconstriction, and cell behavior and metabolism in conjunction with fluorescent dyes and microscopy. Research toward using computer stereographics integrated with stereoscopic video in a quantitative sense, has progressed in recent years. The influences of stereoscopic video image character and the use of a "virtual tape measure" on a user's spatial reasoning have been studied [35, 55].

1.4 Thesis Objectives

The purpose of this work is to apply video, which is already used in many areas of medicine, in the IGNS environment. This requires building an interface for orientation of tomographic data to match a video view of patient anatomy. Once this is accomplished an additional goal is to demonstrate the usefulness and practicality of including video in IGNS to address the problems of:

- confirming that the accuracy of IGNS probe registration is maintained, by displaying the stereotactic probe overlaid with the video image of the same probe. Because video can show the patient head and the probe, the accurate relation between the two in the video image may be compared with their IGNS representations.

- displaying the three dimensional orientation of operating instruments relative to tomographic renderings. The fused images will include stereoscopic views of the operating tools, the surgeon's hands, and other objects not normally included in the IGNS, shown relative to the internal anatomy.

- allowing markings on the patient skin to be used as guides in IGNS. Markings can be used to help register the video to the IGNS space, and to demonstrate a planned procedure, such as a craniotomy relative to structures beneath the skin.

- emphasizing morphological discrepancy between tomographic and real patient anatomy by superimposing images of the two. The pre-recorded medical images may initially correspond well, and subsequently become misaligned with the video images. The skin in the medical images may also register well with the video, but the cortex may distort when the cranium is opened.

and to assess the factors contributing to accuracy limitations of aligning the video to the tomographic data.

Chapter 2

Image-guided Neurosurgery at the MNI

This chapter describes the IGNS environment at the Montreal Neurological Institute. First, the ways in which IGNS is useful are discussed in general terms, then specific applications are presented. The environment in the laboratory, including computing facilities and technologies used, which are specific to IGNS are described next. A description of the imaging modalities available in the IGNS system, how they are combined with one another, and display formats offered follows. Finally, some of the limitations of the IGNS system are presented.

Neurosurgery at the MNI is performed with a workstation-based three dimensional imaging system which correlates multi-modality images, and is used both in neurosurgical treatment planning, and for guidance during surgery. The image-guidance system used is based on the Viewing Wand (**VW**), an interactive, frameless, stereotaxy environment developed by **ISG Technologies**[1], although video would be equally useful in a frame-based system. In describing the IGNS environment, attention is given to discriminating between standard VW functions and those developed at the MNI within the neuro-imaging laboratory.

[1]I.S.G. Technologies Inc., Mississauga, Ontario, Can.

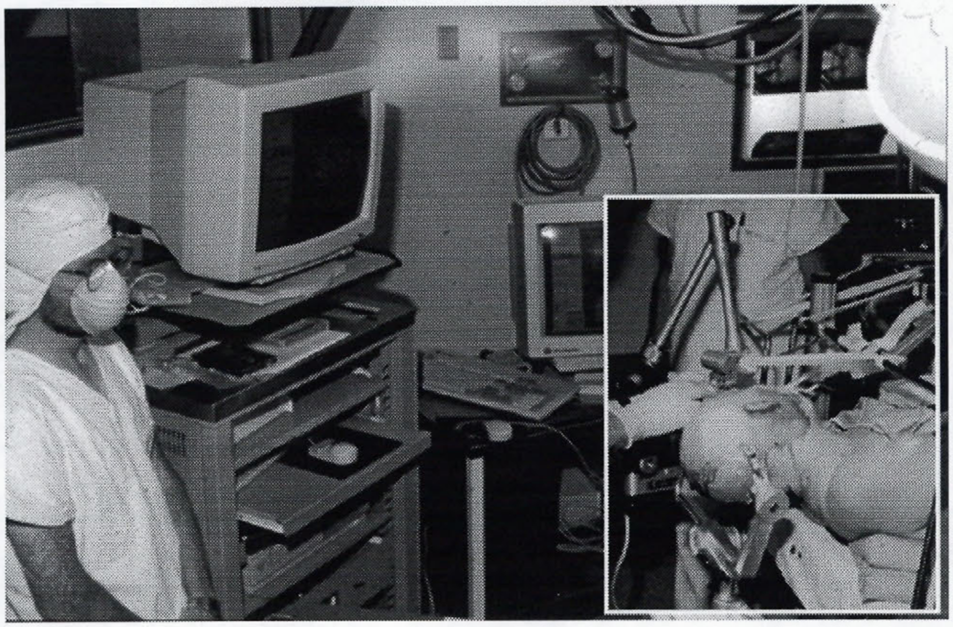


Figure 2.1: Left: the Viewing Wand probe being used in the OR. Right: the FARO arm in use in neurosurgery.

The VW system includes a hand-held probe whose spatial position is tracked during a procedure and displayed within the three dimensional imaging system. It is 'interactive' in that the surgeons can manipulate the probe whose position is continually updated and displayed on the workstation screen in the operating room. The adjective 'frameless' signifies that the relation of the viewing wand to the prerecorded diagnostic images may be established with or without reference to an invasive frame mounted on the patient's head. The environment is 'stereotactic' in that the surgeons use images to orient and guide themselves in three dimensions during a procedure. The relevant operating instruments and operating setup are shown in Figure 2.1, and it is into this environment that video will be introduced.

2.1 Applications of Image-guided Neurosurgery

Although CT and MR images are most commonly used and are standard to the VW system, additional modalities (PET, DSA, MRA, EEG) have also been integrated at the MNI [39, 38, 40, 24], and this thesis deals with the integration of video.

At the MNI, when the VW probe is used the surgeons claim that it continually improves and refreshes their knowledge of neuro-anatomy. It helps them to better recognize anatomy, and confirms their anatomical knowledge by increasing their awareness of the brain orientation. With the probe, surgeons can more confidently locate anatomical features when even major landmarks (e.g. central sulcus) are not clear from the exposed patient anatomy. In helping to identify landmarks, the probe is used in planning and confirming craniotomy boundaries.

Prior to epilepsy surgery, electrodes are implanted deep in the patient head to monitor brain electrical activity for the identification of seizure foci. The probe is most commonly used to guide this depth-electrode positioning. In treating intractable epilepsy, full or partial temporal lobe resection is frequently assisted by the VW system. It is also used to help surgeons in callosotomy and hippocampectomy procedures. Being able to visualize the probe trajectory for EEG depth-electrode implantation, cyst aspiration, and biopsies allows surgeons to carefully plan an approach path, and to judge the depth of penetration. Finally, tumor resection procedures are enhanced by using probe guidance.

2.2 Laboratory Environment

2.2.1 Programming Environment

The laboratory has an agreement with ISG Technologies, whereby ISG employees work part-time in the lab. This affords interchange of ideas between the clinical, research, and commercial environments. Clinical experience is gained by lab researchers by providing technical assistance, relating to the VW, to the surgeons in the OR. In the laboratory at the MNI both the VW and medical imaging applications from within the neuro-imaging laboratory run on Silicon Graphics Indigo workstations.

These programs are all written using the Image Applications Platform(**IAP**)[2], an object-oriented, image manipulation and display environment developed by ISG. This programming platform offers many image manipulation tools which can be combined to aid in building applications. Programming is performed in C on a UNIX operating system.

2.2.2 Mechanical Position Tracker and Probe

The **FARO arm**[3] is a mechanical pointing device (see inset of Figure 2.1), which is mounted on the operating table and connected to a probe [43]. It has six articulated joints, equipped with precision potentiometers which express the joint angles as voltages, which are then converted to digital signals. From these signals, six rotation matrices are computed, which are used to derive the position of the probe in space relative to its base. This space defined by the internal Cartesian coordinate system of the FARO arm-probe system is referred to as probe-space or patient-space.

Probe-Patient-Workstation Connection

The IGNS hand-held probe is displayed relative to imaged internal structures. The computer workstation monitors the probe position and displays it relative to the image space volume. When the probe is used in a neurosurgical procedure, the patient is rigidly fixed to the operating table by a Mayfield clamp, as shown beneath the patient head in Figure 2.1, or by the OBT-stereotactic frame. The probe-space is registered to the image-space via the patient; natural landmarks are identified on the patient skin or on the frame, and in the tomographic images, and input to a least-squares fitting algorithm. The homologous point registration is checked visually and if registration parameter accuracy is unsatisfactory, an optional surface fit determines the registration with a precision of typically better than ±2-4mm [57].

---

[2]Image Applications Platform

[3]FARO Medical Technologies, Orlando, Florida.

2.3 Image Modalities Available

Within this system, many image modalities are available to the surgeon during an operation. Magnetic resonance images (**MRI**), and computed tomography (**CT**) give anatomical information; magnetic resonance angiography (**MRA**), and digital subtraction angiography (**DSA**) show exclusively the cerebral vasculature, while **Ga-DTPA enhanced MRI** shows the vasculature contrasted within brain tissue; positron emission tomography (**PET**) and electro encephalography (**EEG**) offer functional information. CT, MRI and PET acquisitions are stored as volumes of 1mm×1mm×1mm voxels, and are displayed in multi-planar reformats, by volume rendering, or by slice segmentation and subsequent surface-rendering [7, 6]. The user controls which of these data-sets to include, and the relative contribution of each data-set in the final image.

2.3.1 Vasculature

MRI and MRA data-sets can be combined and displayed in two and three dimensions as will be discussed further, later in this section. Ga-DTPA enhanced MRI scans are displayed as surface rendered views where the vessels are segmented as separate objects. Registration of vascular images to other anatomical images allows one to compare the expected relation of vasculature and anatomy to the the correct relation, as depicted by the registered vascular images.

2.3.2 Importance of Displaying Functional Areas

PET images show brain activation in different areas of the brain. A PET study can be designed to define a functional region near to the planned operation site. If a procedure takes the surgeon close to an important functional area, a relevant PET study may be superimposed on the MRI data. Because PET acquisitions have inherently lower resolution than MRI, the volumetric PET data must first be interpolated to

Figure 2.2: The Viewing Wand image manipulation, user interface.

1mm×1mm×1mm voxels.

When a patient EEG is recorded, the electrode positions are noted relative to the patient head. Their position is also marked relative to the skin surface which is segmented in the MRI data-set. The patient's head is first modeled as a wire mesh sphere, which is altered to conform to the shape of the segmented skin from an MRI scan. The markers are visible in the MRI on this tessellated model of the head as they were positioned on the patient. Then the EEG signals are interpolated between electrodes. A series of images can be constructed from an EEG of a patient experiencing an epileptic seizure. The series can be played to surgeons in a short cine-loop so that when operating, electrical activity indicating an epileptic focus can be displayed relative to the standard stereotactic information.

2.4 Display Format Options

Features available in the VW program include:

- volumetric data may be reformatted and viewed in two-dimensional planes (axial, sagittal, coronal, parallel or perpendicular to the probe),
- segmented objects may be surface-rendered in three-dimensions,
- objects may be viewed as semi-opaque to allow internal structures to be viewed relative to an unbroken external surface, or as opaque and cut,
- registration may be achieved by homologous point matching and by surface matching of image and patient,
- volumetric data can be viewed irrespective of the probe position or indicating the current position of the localizing device.

The patient's CT/MR/PET scans are loaded into computer memory and a three-dimensional shaded surface rendering of the patient's head is displayed simultaneously with scan images on the intra-operative computer screen as shown in Figure 2.2. The arm's endpoint location and its directional vector are shown as cursors on the relevant scan slices in the upper and left three frames of the VW interface, and change continuously as the surgeon moves the probe. These three frames sift through planes of the data-sets to display the corresponding scan slice as the probe tip moves around the patient's head. Note however, that DSA images can only be displayed from the orientation at which they were acquired as they are not fundamentally volumetric data-sets.

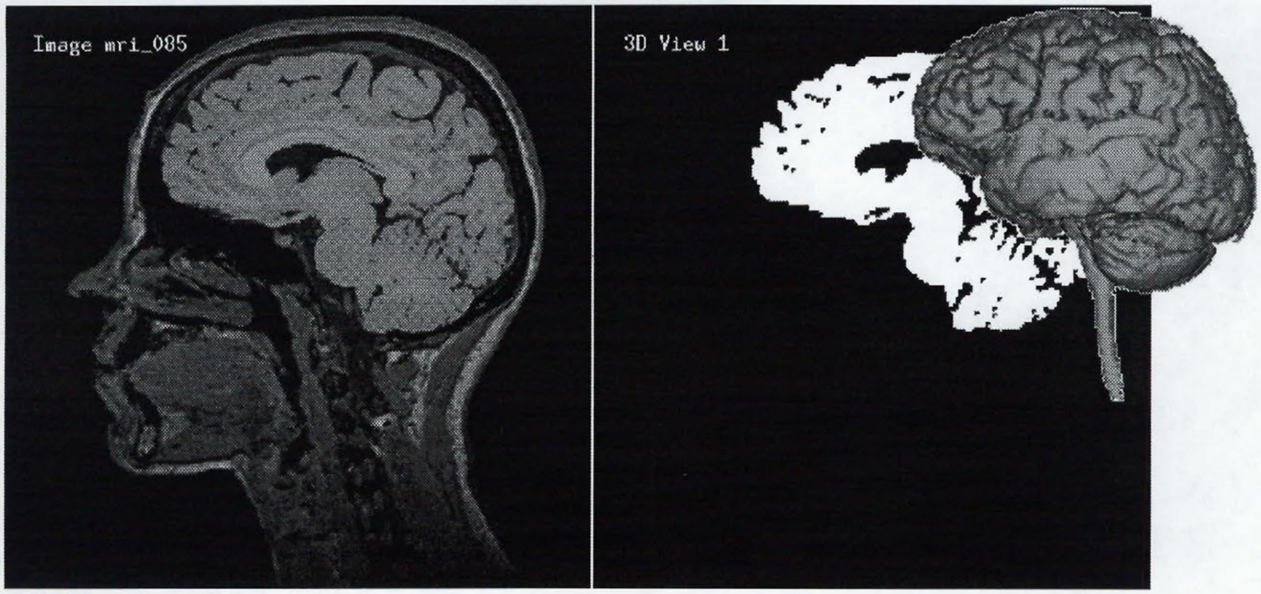


Figure 2.3: Slice segmentation for subsequent surface rendering.

2.4.1 3D Rendering

Surface rendering and Slice Segmentation

The display of surfaces of a 3-D image implies that the surfaces must be identified explicitly within the data volume. The **Allegro** is a workstation at the MNI that accepts volumetric MRI or CT data and provides tools to segment, display and analyze 3-D structures. A user places a cursor within an area of interest and clicks on a pixel, thus creating a "seed". All adjacent pixels within a specified grey scale window around the value of the seed pixel are automatically selected to create a slice object, shown in Figure 2.3. The seed and selection parameters can be automatically cascaded to the underlying slices and the image segments selected can then be refined manually. The slice segments are then stacked together to create the three dimensional segmented object or solid model of the segmented anatomy, which can be seen as surface rendered views with applied lighting models [4].

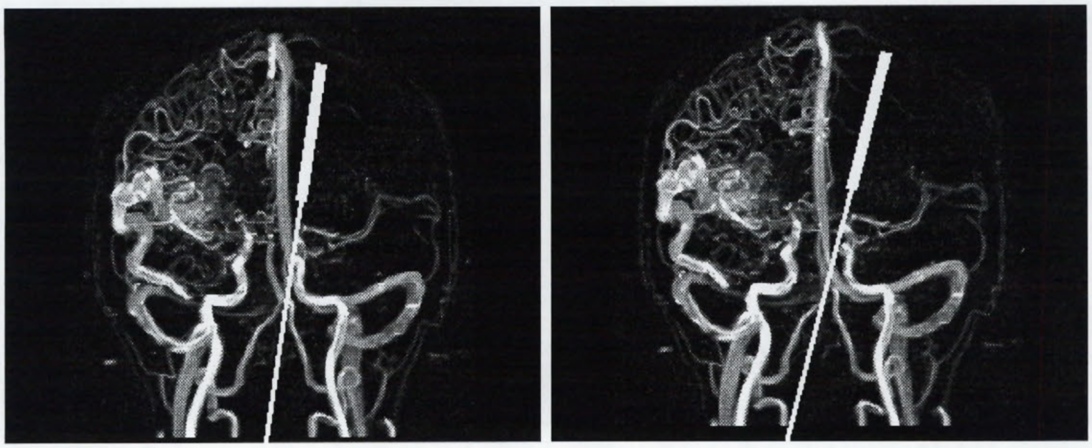


Figure 2.4: Stereoscopic maximum intensity projection images of an MRA data-set. Arranged for cross-eye viewing. The white line in the centre of each image is a representation of the VW probe.

Maximum Intensity Projections

Vascular information can be acquired using an MR scanner, and data displayed similarly, although instead of being segmented, MRA is commonly displayed as a Maximum Intensity Projection (**MIP**) image. MIP images are created by representing the maximum intensity encountered by a ray cast from the rendering perspective through the data volume on its path to the image plane [14] as in Figure 2.4, where the white line shown near the brain mid line is the probe. When images of this kind are viewed stereoscopically, the three dimensional orientation of the probe relative to this vasculature is obvious.

2.4.2 Stereoscopic Viewing

At the MNI, stereoscopic viewing of radiographs has been utilized since 1935, and it is standard practice to acquire and view all radiographic angiograms stereoscopically. The option of viewing images stereoscopically provides depth perception, useful for 3-D interpretation of images, especially angiograms [6, 7, 25, 37] as demonstrated by Figure 2.4. To view these images stereoscopically an observer must cross his or her eyes while keeping the images in focus. The left eye should centre on the image on the right side of the page and the right eye should look at the image on the left. In correct viewing, three images will be perceived and the middle image will appear in stereo.

DSA

Surgeons have been using stereoscopic DSA for planning neurosurgery and radiologists have been using it for diagnosis of vascular problems. DSA images are acquired and displayed stereoscopically. The surgeons affirm that stereoscopic DSA is especially useful because, with their knowledge of brain anatomy, they are able to envision the cortical surface just from the vasculature shown.

2.5 System Limitations

Viewing other instruments

During a neurosurgical procedure, the probe position is displayed on the VW workstation. A major limitation of interactive image guided surgery, however, is that surgeons usually work with instruments other than the probe, which are not presently displayed in image-space. They orient themselves using the IGNS probe, then have to lay it aside and return to using conventional instruments which are not shown relative to internal structures. In addition, marks drawn on the patient skin are not displayed in a standard IGNS system. Also, the surgeon has no way to compare a pre-segmented, planned resection as treatment verification of the accuracy or consistency of the plan during a procedure.

Confirming probe registration

The standard check on probe registration is a visual comparison of the position of known landmarks on the patient, and in tomographic data. The probe tip is positioned at an identifiable landmark on the patient, and the stereotactic representation is compared to that seen relative to the patient.

Brain distortion or shift

Another major limitation of interactive image guided surgery is the lack of live feedback to indicate to the surgeon, the current state of the brain or position of the probe and other surgical instruments relative to the actual position of brain structures during a procedure. Probe registration works very well assuming the relationship between image and probe-space remains constant during a procedure. Currently, if the positions of objects represented in prerecorded images changes significantly, the information represented in the stereotactic environment is no longer accurate, and may be misleading. When a craniotomy is performed, the cortex often distorts. This change may be a deflation, or inflation combined with pulsation when the brain is exposed. There is controversy as to the magnitude of brain motion, but it is generally accepted to be a few millimeters. While there is presently no way to quantify or correct for this phenomenon, we view the use of integrated video as a step towards addressing this problem.

Chapter 3

Video

This chapter introduces video in terms of the role it can play in improving IGNS. A basic description of the hardware involved in generating a video signal, and the signal components follows. Video signal limitations, and sources of image distortions are described next, followed by the definition of a camera space. The last section compares binocular with monoscopic viewing.

3.1 Video and IGNS limitations

Video is potentially a rich source of live visual feedback to the IGNS system. Objects included in the video scene are not limited to those in the pre-recorded diagnostic images and the probe. Indeed, once the video images have been registered so that their relations to both the probe space, and image space are known, other objects such as surgical instruments may be visualized within patient-image space. In order to visualize the 3-D space effectively, these other objects may be viewed stereoscopically, relative to the internal anatomical structures represented in the stereotactic environment. In this way, stereoscopic video connects instruments in the operating room, other than the VW probe, to tomographic data.

In addition to showing surgical instruments, a video image offers valuable information by including visible marks drawn on the patient skin. For example, a surgeon may draw a planned craniotomy onto the patient head, and it is immediately displayed in the video image that has been previously registered and integrated with the 3-D anatomy. This enables the surgeons to see which structures lie below the plan drawn on the skin surface. Also, a surgeon can pre-segment a planned resection from the tomographs so that when a procedure has been performed, comparing the video image of the operating scene with the planned resection can indicate the accuracy or consistency of the procedure with the plan.

Once the stereoscopic video and stereotactic spaces are registered, both will contain representations of the probe. The probe, seen stereoscopically in the video images can act as a constant check on its VW registration relative to the stereotactic patient head. The stereotactic probe and patient representations from the VW application should both overlap with the actual probe and patient shown in the video. The positions of these two probes relative to the patient and each other will indicate the validity of the registration of the stereotactic probe both initially and throughout the surgical procedure. If the patient and/or the probe in stereotactic space do not correspond, this indicates either video or probe misregistration. However, if the probe orientations initially correspond, the integrated video should serve as a check that any subsequent mis-alignment of the video and stereotactic anatomical representations is due to morphological change in the patient, and not video or probe mis-registration.

By introducing video into the operating room for open craniotomy procedures, one can detect and eventually monitor superficial cortical distortions or shifts during surgery. Displaying the video on both mono and stereoscopic data-sets aids in highlighting any such distortions. Once they are exposed and visible, cortical landmarks can be identified in the video image to observe if the brain orientation in the skull is accurately represented in the tomographic images, or whether it has changed since their acquisition, alerting the surgeons to the fact if there has been a change. Another option for using video to monitor changes in the patient anatomy once objects of interest have been exposed, is by subtracting a captured video image from one acquired

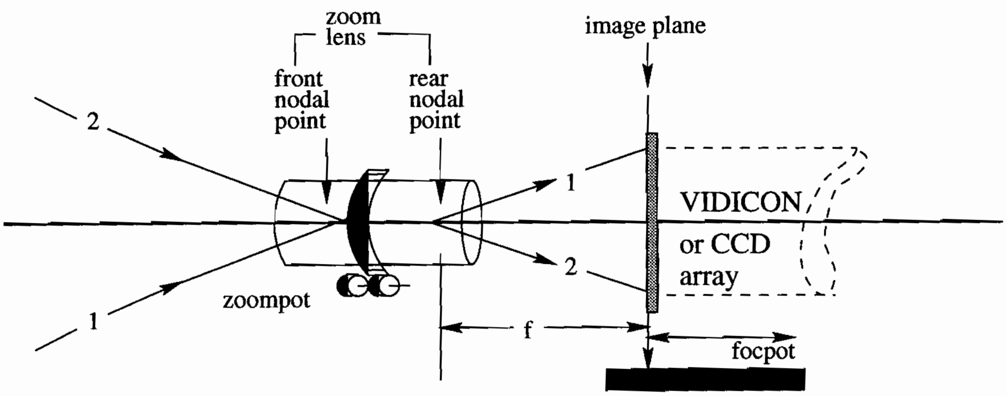


Figure 3.1: Video camera interior showing the relation of zoom lenses to a camera optical centre.

subsequently. This difference image will then emphasize discrepancies between the video scenes of the first and second images.

3.2 Signal Description

A video camera converts a visual image or scene into an electrical signal representation. This conversion occurs in the pick-up tube of the camera, often a vidicon, which consists of a photocathode, to convert an incident light image into a charge pattern, and an electron scanning beam which reads the charge pattern from the photocathode plate. The pick-up tube electron beam generates the video signal by scanning the photoconducting target. The current state-of-the-art video signal generator is the charge-coupled device (**CCD**), a semiconductor device that is used primarily as an optical sensor and that stores charge before transferring it sequentially to an amplifier and then a detector. A diagram of the internal hardware of a modern video camera is given in Figure 3.1.

Video cameras output images at 30 frames per second, where one waveform consists of a colour burst, then the field information followed by a synch or refresh pulse (see Figure 3.2). In addition to luminance, chrominance and audio, the video signal carries vertical and horizontal synchronization information, and a colour phase

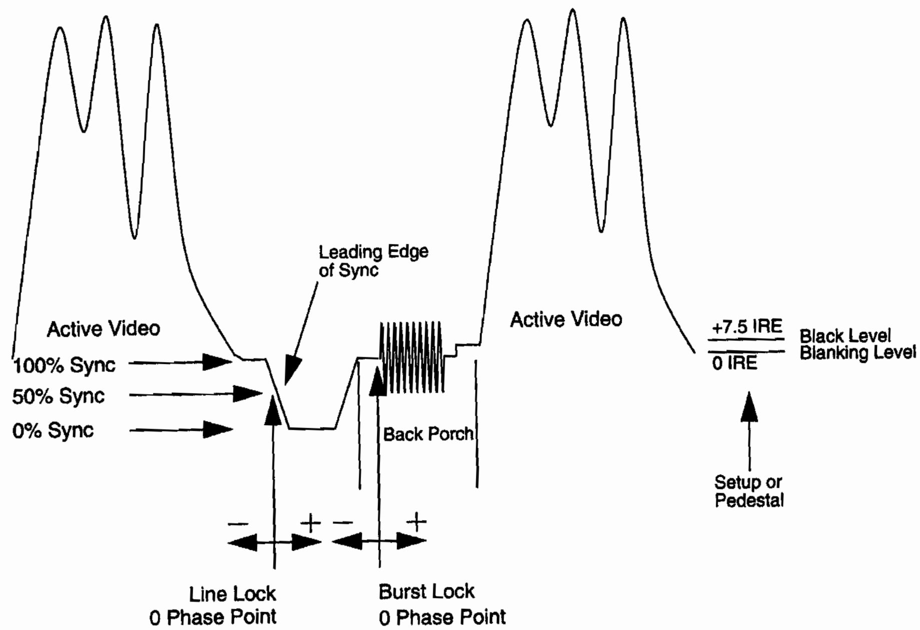


Figure 3.2: Video signal diagram.

reference.

When an eye is stimulated at a low rate the stimulus may appear to flicker [35]. The eye can resolve up to 50 flashes per second; this perception parameter is called the critical fusion frequency. The critical frequency depends on the duration of the light stimulus or image, the image character (contrast, brightness, etc.), and is also viewer dependent. To avoid displaying a flickering image the video frame is split into odd and even line fields and alternate fields are displayed at 60 fields per second.

3.3 Sources of Distortion

In the camera specifications, the SNR is specified to be >44dB. The required minimum illumination is 0.3 footcandle (3 lux) with a f1.4 lens used under incandescent light. Bandwidth, or bandpass (the frequency range that the electronic components of the video system are designed to transmit) determines the horizontal resolution potential of the signal [23]. The video interface used converts the analogue video signal into a 3×8 bit RGB (24 bit) colour scale which dictates the intensity resolution limitation of the image.

3.4 Video Coordinate System Definition

The image and camera coordinate systems shown in Figure 3.3 are Cartesian and denoted as:

Image (2-D): $\vec{\mu}$,$\vec{\nu}$

Camera (3-D): $\vec{x}$,$\vec{y}$,$\vec{z}$

Coordinates in image space are determined by the horizontal and vertical displacements of a pixel, within the image plane, from the image centre. The coordinate $\vec{\mu}$ increases to the right while $\vec{\nu}$ increases upward. The image acquisition plane is physically positioned so that the cross product $\vec{\mu} \times \vec{\nu} = \vec{w}$ lies along the **principal ray**. The principal ray of an image is the line perpendicular to the imaging plane which passes through the **optical centre**, centre of projection (defined in Section 4.2.2), or point of convergence for parallel-ray input of a video camera. The optical centre and principal ray are shown diagramatically in Figure 3.3.

The origin for the camera coordinate system is defined by the optical centre of the camera. The 'left handed' camera coordinate system can be defined so that the $\hat{z}$-axis also follows the principal ray and the $\hat{x}$-axis and $\hat{y}$-axis follow the scan line and pixel number respectively, making them parallel to $\vec{\mu}$ and $\vec{\nu}$ respectively.

3.4.1 Projected point coordinates

The video imaging geometry can be modeled after a pin-hole camera. The two dimensional image coordinates of a tridimensional point in terms of the defined geometry are given by similar triangles:

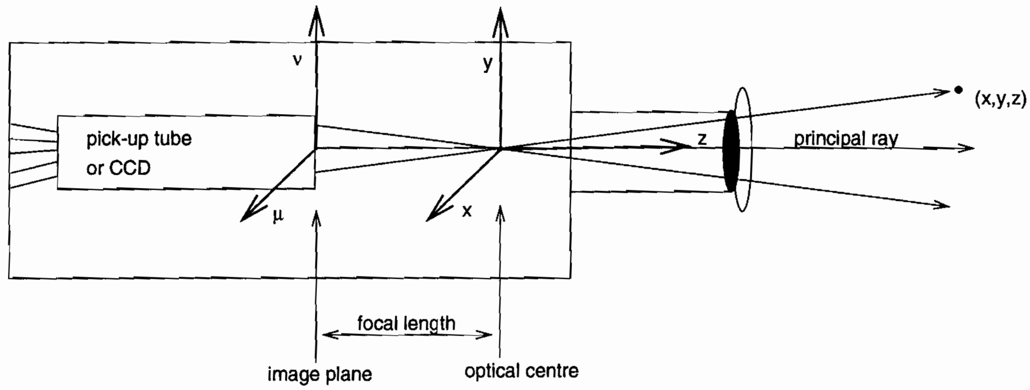


Figure 3.3: Diagram of Image and Camera Coordinate Systems

$$\sigma = \frac{f \cdot x}{z} \tag{3.1}$$

and

$$\eta = \frac{f \cdot y}{z} \tag{3.2}$$

where f is the camera focal length [34]. These image coordinates can be scaled to the video focal length, giving

$$\mu = \frac{\sigma}{f} = \frac{x}{z} \tag{3.3}$$

and

$$\nu = \frac{\eta}{f} = \frac{y}{z} \tag{3.4}$$

The scale of the camera coordinate system is set by assigning f to be unity and the units of $\vec{x}$ and $\vec{y}$ to be equal to those of $\vec{\mu}$ and $\vec{\nu}$. This also sets the image plane to be one unit of distance from the camera optical centre along the $\hat{z}$ axis (Figure 3.3).

3.5 Mono vs. Stereoscopic Video

Although it is commonly known that binocular vision aids in depth perception, monocular depth cues are also very strong depth indicators. Examples of monocular depth cues are:

- size- nearer objects appear larger than farther objects
- occlusion or overlapping contours- near objects overlap distant objects
- perspective- the appearance to the eye of objects with respect to their distances and positions relative to a particular vantage point
- shading- the change in intensity within outlines which suggests three-dimensionality, shadow, or degrees of light and dark in an image
- air haze- water vapor, dust, and smoke in the atmosphere absorb and reflect light, casting a veil over distant objects and causing them to appear less sharp.

Extensive research into the success of performing teleoperation tasks in three dimensions, using monoscopic versus stereoscopic viewing has been carried out at the University of Toronto [35, 36, 47, 48, 49, 56]. The results of this work indicate that the benefits of perfoming a task using stereoscopic video rather than monoscopic video last longer as the dependence of the task on stereoscopic depth cues increases [9]. The effect of motion parallax has also been investigated by this group and they have found it to be beneficial, through assessment of task completion success, using a rotating wireframe angiogram model [49].

Although stereoscopic images give information about the relative distances of objects displayed, absolute depths can only be determined by knowing parameters external to the images themselves. Examples of parameters which allow absolute distances to be determined from binocular images are, the focal lengths of the imaging cameras, the absolute dimensions of an imaged object, or the distance separating two views. Without such information, if the distance of the imager from the subject and the size of the subject are increased proportionally, any number of identical images can be acquired while the absolute distances vary. In this way, the scale of an entire stereoscopic imaging system can change and produce identical images with different setup geometries.

Figure 3.4: Stereoscopic video images of subject arranged for cross-eye viewing

Binocular depth perception, or stereopsis, depends on the ability of the brain to receive discrepant images from each eye, and to fuse them into a single image which has depth. One way that images in a stereoscopic pair differ is in disparities, or relative displacements of corresponding left and right image points. Stereopsis uses clues like disparity to give relative depth information, but not absolute depths. It allows a viewer to qualify which of a group of objects is closer to the image receptors, but not to quantify distances. However, with very little external information about the imaging system, or the imaged subject, stereoscopic images can be used to create an absolute depth map.

Corresponding image points are also referred to as homologous or conjugate points. The disparity of **parallax** is the distance between left and right homologous image points, and this parameter actively contributes to depth perception. Horizontal screen parallax, (the effect which parallax has on a stereoscopic screen display) is generally comfortable for disparities up to 1/2 inch, depending on viewer distance from the screen, the nature of the image, and the particular viewer. Vertical parallax is the vertical displacement of homologous image points in a stereoscopic pair due to their distances from the viewer [50] or due to asymmetrical image magnification of the stereoscopic pair.

Convergence and **accommodation**, or focusing, are involved in the mechanics of producing **fusible** discrepant images with the two eyes, but not in the active perception of depth. Viewing stereoscopic images on a computer screen creates an unnatural situation which forces the viewer to decouple their accomodation and convergence functions. In the real world, when a point is observed, the eyes centre on, and focus at, a single point. In stereoscopic viewing on a computer screen, the accommodation/convergence ratio is not unity when disparity is present due to parallax. Keeping this relation close to its natural value of unity makes stereoscopic fusion easier, but it is the point position disparities that give the perception of depth.

Other ways to make stereoscopic fusion easier, using stereoscopic video include avoiding [32]:

- Illumination asymmetry- differing point-for-point illumination intensity in an image pair
- Aberration asymmetry- differing optics, focus, or vidicon or CCD character
- Geometrical asymmetry- differing focal length or magnification, image linearity, convergence and coplanarity.

When combined with stereoscopic graphics, geometrical distortion becomes an especially significant factor in a viewer's success in stereoscopic fusion [35]. By having the video cameras in a position which mimics human eyes, the images acquired are easily fusible. Matching the graphics to real video images resolves many of these concerns by generating a pair of naturally fusible stereoscopic images. In combining graphics with stereoscopic DSA images, the potential pitfalls and advantages of different colour combinations have been investigated [3]. Experiment indicates that a user's perception of the depth of a graphical cursor passing over stereoscopic greyscale images is most reliable when luminance is consistent between the cursor and the images and colour offers high contrast.

Chapter 4

3-D Video / MR Integration

This chapter begins with a description of the programming and laboratory environments. Next, issues involved with achieving the fusion between video and 3-D models are introduced, namely: how VW functionality is used by the video fusion application; how IAP and video create images of a 3-D subject; a potential method to approximate a video image using the IAP rendering capabilities; and finally, the two main display techniques used. A key component of the fusion procedure is presented in detail in the third section, and the fourth section details the application program. This is followed by a discussion of video fusion with a 3-D model including the determination of the relation between two dimensional image coordinates of features from a video image, and the corresponding features on the actual patient as identified by the probe. Finally, once the three dimensional relation of each stereoscopic video image to probe-space is established, appropriate stereoscopic views of the fused 3-D model and video are generated and a graphical probe representation is displayed, invoking the registration from probe to video space.

4.1 Laboratory Setup and Equipment

A VW probe designated solely for research, is shown in the centre of Figure 4.1.

4.1.1 Programming Environment

The Image Applications Platform (IAP) [22, 21] is a high level object oriented programming environment which provides 2-D and 3-D image processing functionality. It is principally used for interactive medical image visualisation, but is designed to be more generally applicable. IAP comprises three distinct parts, namely the processing, database, and hardcopy servers. It is built in a client/server style of architecture with the three servers and their associated client-side toolkits. In operation, the system 'server' performs computation requested by a 'client' or the application code. A single server may respond to multiple clients as in this case where the processing server is being accessed by both the VW application, and by the video fusion code. The processing server communicates with both a graphics library and a windowing server (for example X11).

In the context of IAP, an **object** is a package of functionality that facilitates interaction with the contained functions. When an application is executed, instances of objects are created, and their functionality is implemented on the server side, while objects created on the client side are 'shadows' of that functionality. Each object can be accessed by an 'external handle' which is common to both the client and server and is mapped to the server handle within the server and to the client handle within the client. Examples of objects in IAP which will be mentioned in this work, and brief descriptions of their functions are as follows:

- "Fusion" object allows 'overlay' display or fusion of two displayable rasters.
- "Geom2" object manages a list of 2-D graphical geometry specifications.
- "Raster" object manages the internal storage of a 2-D image raster within an IAP processing server.

- "Tx3" object manages a set of 3-D transformation parameters.

- "Solid" object manages a group of different objects which taken together represent a 3-D object (or group) which can than be displayed.

- "Solid view" object takes as input 3-D object descriptors and various other 3-D render parameters to produce a raster image which is the rendering of the scene.

In order to execute an application, a client sends messages to the server requesting it to create instances of whatever objects are required to perform the specific task and to connect these instances together in a data-flow pipeline. A pipeline exists entirely within the server and links instances of objects together as a message channel. If an object is disconnected from the display pipeline, it will no longer be rendered. When two applications are running simultaneously on a server, their pipelines can be connected by a 'watch' mechanism. The pipeline network of the video fusion application can be connected to the VW application pipeline through such a 'watch' mechanism since both applications communicate with a common processing server. When a particular object in a pipeline is connected to the pipeline of another application, a 'watch has been taken' on that object. In this way the functionality of the object established in one application can be imported by another.

Finally, there is a general coordinate system built into the processing server. Coordinates are specified in one of four spaces:

- Image space - the origin is the top left corner and points within a raster are addressed as (*row, column*) in pixel units

- Slice stack space - A slice stack is a set of 2D rasters stacked on top of one another to form a 3D volume. Pixels in the slice stack must be addressed by means of a $(X,\ Y,\ Z)$ and the origin of the slice stack may be at any arbitrary location

- Volume space - Binary volumes within IAP exist within a cubic volume comprising 134,217,728 voxels i.e. dimensions of $512\times512\times512$. The origin of this

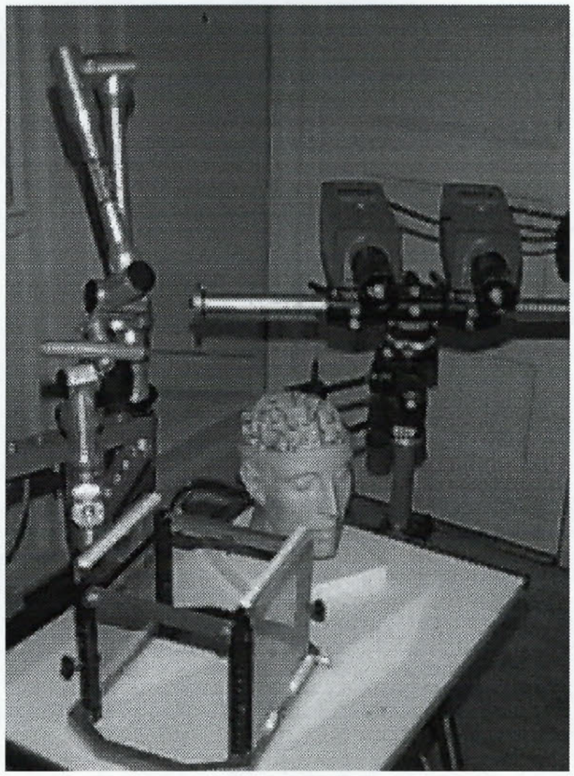


Figure 4.1: Mechanical arm, stereotactic frame, phantom, and stereoscopic video cameras.

volume is the top, left, front corner. voxels must be addressed by values from (0,0,0) to (512,512,512).

- Screen space - A point is addressed on the image in viewport coordinates where the top left corner of the viewport is (0,0) and the bottom right corner is (1,1)

4.1.2 Video Sources

The closed-circuit video system in the laboratory is comprised of two Panasonic[1] WV-1500 composite cameras, connected to a Silicon Graphics, Indigo workstation via the **Galileo Video**[2] board. The cameras have f1.6 (16mm) C-mounted lenses.

[1]Panasonic Industrial Company, Division of Matsushita Electric Corporation of America.

[2]Silicon Graphics Inc.

The camera pick-up tube is a vidicon with 2/3" wire mesh target. This pick-up tube target defines the imaging plane of the system. The video output level is $1V_{p-p}/75\Omega$. Video images are acquired by grabbing a single frame from each of the cameras shown on the right in Figure 4.1. The 525 line (262 line-pair) analogue signal, with 650 lines of horizontal resolution, is fed through cables to the Galilleo board, which digitizes them, loads them into buffers, and transfers them to the computer. The Galilleo board also generates pulses to synchronize the camera scanning beams, and to coordinate field acquisition by the Galilleo board. This signal is passed through a signal amplifier before being split and fed to the VS/HD input of the composite cameras. The video can be shown live, or as single frame-grabs. The time at which a frame is grabbed is controlled by the user via a button on the software interface. The single video frame acquisition and display takes 10 seconds because the pixel values are recorded into a buffer. Careful attention must be paid to the camera adjustments to ensure comfortable binocular fusion (see Section 3.5). Vertical skew of the camera poses, and different image zooms were the most difficult asymmetries to avoid. The camera mounts were shimmed to achieve the desired relative poses. Once the relative orientation of the cameras was accepted, they could be moved as one unit by adjusting the tripod. It is envisioned that miniature cameras will ultimately be mounted in the operating room, probably integrated with an operating microscope.

4.2 Merging Tomographic and Video Media

Fiducial coordinates from the 3-D model data points can be registered to the stereoscopic pair through a transformation, and then a projection. Once the two sets of stereoscopic video coordinates are fitted, each transformation can be applied to the 3-D model to attain the appropriate rendering of the MRI data.

In the operating room, before draping the surgical site, or making any incisions, one can capture a video image of the surgical scene to computer memory. This image can be displayed on the computer screen simultaneously with a monocular representation of the surface of the cortex and related vasculature, as obtained by rendering a surface of the appropriately segmented tomographic data-sets. With the use of stereoscopic pairs of video images, typically of the patient's skin or exposed cortex, live video offers more sophisticated possibilities which require depth information not apparent from monoscopic views.

As will be shown in section 4.3.2 identification of six or more object points or features allows the relative orientation of the two projections to the 3-D model to be calculated. Examples of natural landmarks or features which are recognizable in both the video and stereotactic images are facial structures and any visible markings on the patient skin which can be identified using the probe. The steps in displaying the 3-D model in the correct orientation are as follows:

- Run VW software and load patient data-set
- Register VW probe to MR data-set
- Compute transformation from probe-space to video camera-space
- Display probe as graphical object continually updated from probe output
- Decompose transformation matrix and generate 3-D transformation components to orient tomographic data

The first two of the above points are standard VW functions; the latter three are discussed in detail in the remainder of this chapter.

4.2.1 Inputting MRI Objects from the Viewing Wand

The image display software platform is used to combine the video frame and tomographic data. In order to view an anatomical object with the video, the fusion application must put a 'watch' on the object which exists in the VW application. Attributes of that object such as colour are brought with it when it is imported into the video software. It can also be 'hidden' in the VW, and subsequently hidden in the video software by disconnecting the object from the data flow pipeline (Section 4.1.1). The relative orientations of the solid objects imported from the VW are also imported, thus VW probe registration aligns the probe object with the patient data in the video fusion application as well as in the VW.

4.2.2 Perspective and Parallel-ray Projection Geometries

The imaging geometry of a perspective, or diverging-ray system, as compared to that of a parallel-ray, or orthographic projection system is shown in Figure 4.2. This figure shows two projections of a line in 3-D, formed by straight projection rays emanating from a centre of projection(**CoP**), passing through each point of the line, and intersecting a projection plane. Both systems can be considered as having a CoP, where the parallel-ray CoP is at infinity, giving rise to an orthogonal projection. [14]

Once the 3-D model is appropriately oriented in 3-D camera-space to correlate it with the patient shown in the video, their images differ in that the tomographic data are rendered by orthogonal projection while the video imaging system creates perspective projections. The parallel-ray projection method is commonly used in 3-D graphics rendering. The diverging-ray projection can be used to model a pin-hole imaging system because if the imaging plane is reflected about the CoP (which is the

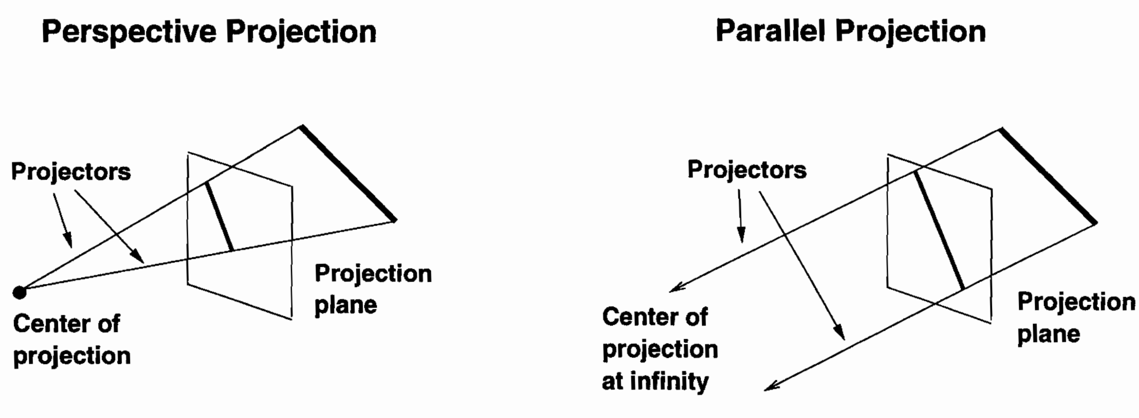


Figure 4.2: Projection geometry compared to parallel-ray projection.

optical centre of the camera imaging system) along the principal ray, into position for a perspective projection (see Figure 4.2), the resultant image would be identical to one with the projection plane in the simple pin-hole arrangement.

Both imaging systems have inherent ambiguity in absolute distances derived from their projections. In a parallel-ray projection scheme, the perpendicular distance of an object from the image plane has no effect on the resultant image. It is therefore impossible to determine distances perpendicular to the imaging plane from a single parallel-ray projected image. Perspective projections incorporate a phenomenon called **perspective fore-shortening**, where the projection size of an object is inversely proportional to its distance from the CoP. For a diverging-ray system, only the ratio of the displacements from the **piercing point** of the principal ray, parallel and perpendicular to the image plane is determined. The absolute distance along the principal ray, and the absolute perpendicular distance from that ray, i.e. the scale of the system, are indeterminable.

Consider the following expressions which relate 2D image coordinates to 3D fiducial coordinates in video space, for a perspective, or true projection as in Section 3.3, and for an orthogonal, or parallel projection, respectively

$$\mu_T = \frac{x_T}{z_T} \tag{4.1}$$

and

$$\mu_P = \frac{x_T}{z_{avg}} \tag{4.2}$$

where x_T, y_T, and z_T are the fiducial coordinates in video space and z_{avg} is the average z_T among a cloud of points. The error introduced by orthogonal projection is represented as the separation between the ideal image coordinate, μ_T, and the parallel projected coordinate, μ_P, i.e.

$$\mu_T - \mu_P = \frac{x_T}{z_T} - \frac{x_T}{z_{avg}} = \frac{x_T(z_{avg} - z_T)}{z_{avg} z_T} \tag{4.3}$$

Because this behavior is radially symmetric about the principal ray of the imaging system, analysis along one axis is sufficient. For a point at a given distance from the principal ray, represented by x_T, the error due to orthogonal projection increases linearly with the depth variation from z_{avg}. The error also increases as z_T decreases, or as the points approach the projection source (in this case, the video camera).

4.2.3 Truncated Pyramids

On certain systems such as the PIXAR imaging computer, a correction can be applied to transform a 3-D object, so that the parallel projection of the transformed object is the same as the perspective projection of the untransformed object [24]. This correction involves scaling successive voxel planes, parallel to the image plane, according to the magnification they experience during the video projection. The volume of a cube with one face parallel to the image plane, rescaled in this way, would create the shape of a pyramid with the point chopped off, or a truncated pyramid. The resizing is centred along the principal ray, so the intersection of the principal ray with the volume must be determined. The scale factor can be determined either geometrically or by comparing line lengths within voxel planes, with their lengths in the perspective image. The voxel planes can thus be scaled to mimic a diverging-ray projection scheme which exhibits perspective fore-shortening.

4.2.4 Implementation of IAP

Stereoscopic Display

A Silicon Graphics workstation provides the stereoscopic viewing system into which the VW functionality is extended. This system sequentially displays right and left images at 120 fields per second by doubling the refresh rate and modifying the video input signal so that the left image is held in the upper half of the field and the right image is in the lower half. A pair of LCD shuttered glasses (**Crystal eyes**)[3] is synchronized with the workstation hardware so that each eye sees only the appropriate

[3]Stereographics Corporation, San Rafael, CA

image fields. The overall line scanning rate remains constant because, while the screen refresh rate is doubled from 60 to 120 fields per second, the number of lines per field is half that contained in a mono-video frame. Due to the reduced number of scan lines in each field in the SGI system, the vertical to horizontal aspect ratio is halved, and the vertical resolution is halved. The vertical-controlling picture circuit is therefore adjusted to double its normal deflection magnitude in order to restore the proper image proportion. Other stereoscopic displays which use a full field per eye view, and alternate at 60 fields per second generally offer the same (512 line) vertical resolution as this system.

A stereoscopic image-pair of the 3-D model is generated by rotating the model by a small angle: typically 5 degrees [51]. This corresponds to the views observed when looking at a scene with the cameras' principal rays converging at about 3m from the imaging planes, where the imaging planes are separated by .25m. The rotation method of generating a stereoscopic pair is acceptable because there is no problem of vertical parallax (see Section 3.5) since the objects are projected orthographically [50].

Differences between the angle of the video camera's field of view and that subtended by the display window cause the perceived object to drift away from the plane of the screen. The monitor would ideally subtend an angle matching that of the field of view of the image displayed. The angle subtended by the whole monitor screen is about 30 degrees, while the angle describing the field of view of each video camera varies with the zoom setting of the camera. This can vary the apparent size of an object by a factor of four on the Panasonic cameras used (see Figure 3.1).

Fused Image Display

As discussed in Section 4.1.1, the IAP platform can combine two 2-D images from two sources (i.e. the video and the 3-D model rendering) by various interleave fusion operations. IAP generates a raster as a set of pixels, which originate from one or another source. The pixels in the final raster alternate between the segmented tomographic data and video frame. In this application, the pixel sources switch every second pixel in the horizontal direction, with the source at the beginning of each line also alternating. Alternate pixels display either the video or the 3-D model image resulting in a checker-board pattern. A typical fused image is shown in Figure 4.3, which demonstrates the combination of a frame-grabbed video image of an head and the segmented MRI surface rendering of the same skin surface.

4.3 Finding the Probe Space Relation to Video Space

This section describes the mathematics of transforming coordinates between the tomographic data and video image systems. Input from the VW probe and the video image systems themselves are used to define this relation. Fiducial landmarks are identified in both the video and on the patient in the OR. In order to find the geometric relation between the 3-D model and video points, the location of points on the patient must be determined in three dimensions and appropriately transformed to match or correlate with the homologous features seen in the video. The procedure involved in this operation is described below.

4.3.1 Homologous features

As in the VW registration of the 3-D model to probe-space, the bases for registering the video with the 3-D model are homologous points. These are identified on the patient in the OR, or the intra-operative patient, via the probe, and in the video pair, using a cursor in the image window. The points chosen are features on the intra-operative patient skin, also visible in the video images. Marks drawn on the patient skin can serve as fiducials for this registration. Coordinates are returned by the probe relative to its own internal coordinate system, or in probe-space (Section 2.2.2). The coordinates attributed to points identified in the video pair are given according to

Figure 4.3: Stereoscopic images of a fused stereotactic and human head.

the video coordinate system defined in Section 3.4.

A least-squares fit is used to determine a transformation to convert probe-space coordinates into camera-space, so that they best project along the $\hat{z}$ axis onto their corresponding video coordinates. As has been described in Section 3.4, 2-D image coordinates can be expressed in terms of 3-D camera-centred coordinates as

$$\mu_i = \frac{x_i}{z_i} \tag{4.4}$$

and

$$\nu_i = \frac{y_i}{z_i} \tag{4.5}$$

where x_i, y_i, and z_i are the camera-centred coordinates of the i^{th} point as depicted in Figure 4.4, and μ_i, ν_i are the i^{th} point's image coordinates. The 3-D probe-space coordinates must be transformed into camera-space so that relations 4.4 and 4.5 are consistent between transformed 3-D, and image coordinates.

Coordinate system relations

Points on the patient skin are initially defined by the probe in terms of the probe coordinate system. As depicted in Figure 4.4 the registration of probe-space to camera-space determines the transformation which describes where a point in probe-space occurs relative to the camera-centred coordinate system. The three dimensional object coordinates are initially defined relative to the probe's own reference frame. These coordinates can be expressed relative to the video camera coordinate system via the following transformation;

$$\left\{ \begin{array}{c} x_i \\ y_i \\ z_i \end{array} \right\} = \mathbf{R} \left\{ \begin{array}{c} X_i \\ Y_i \\ Z_i \end{array} \right\} + \left\{ \begin{array}{c} t_x \\ t_y \\ t_z \end{array} \right\} \tag{4.6}$$

where X_i, Y_i, and Z_i are the coordinates of the probe tip relative to its internal coordinate system (Section 2.2.2). The column vector elements, t_x, t_y, and t_z represent the translation vector component of the transformation, where t_x is the translation

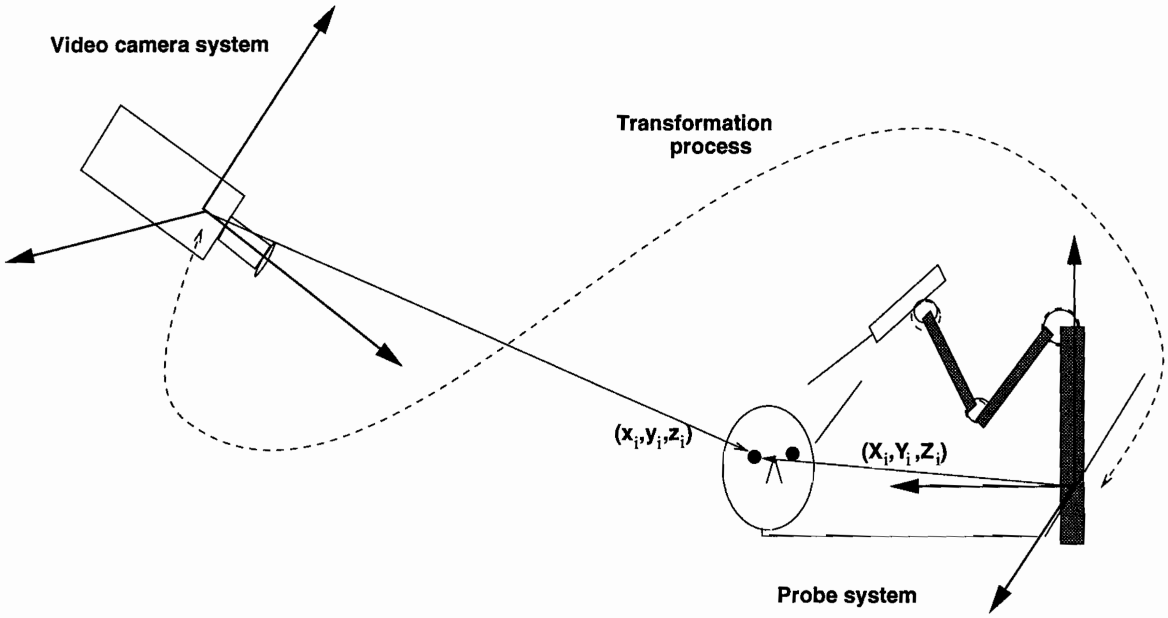


Figure 4.4: Diagram of the transformation of coordinate definitions between the Video camera system (x_i, y_i, z_i) and the Probe system (X_i, Y_i, Z_i). The transformation consists of rotation, translation and scale components.

in $\hat{x}$ between the origins of the probe and video camera coordinate systems, while t_y, and t_z are the corresponding translations in $\hat{y}$ and $\hat{z}$. The components of the rigid rotation matrix R can be given as

$$\mathbf{R} = \mathbf{\Theta\Psi\Omega} \tag{4.7}$$

where

$$\mathbf{\Theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{bmatrix} \tag{4.8}$$

$$\mathbf{\Psi} = \begin{bmatrix} cos\psi & 0 & sin\psi \\ 0 & 1 & 0 \\ -sin\psi & 0 & cos\psi \end{bmatrix} \tag{4.9}$$

$$\Omega = \begin{bmatrix} cos\omega & -sin\omega & 0 \\ 0 & sin\omega & cos\omega \\ 0 & 0 & 1 \end{bmatrix} \tag{4.10}$$

and where ω, ψ, and θ are clockwise rotation angles about the $\hat{x}$, $\hat{y}$, and $\hat{z}$ axes, respectively. These transformations can be expressed in terms of the homogeneous transformation matrix multiplication:

$$\begin{Bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{Bmatrix} = \begin{Bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{Bmatrix} \begin{Bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{Bmatrix} \tag{4.11}$$

where rotation matrix elements with a scaling component are represented as r_{11} through r_{33}, and the translation components are again in terms of video camera-space.

4.3.2 Relating inputs by transformation

Once the probe-space inputs are transformed into video-space they can be related to their homologous image inputs. The input image coordinates can then be expressed in terms of the transformed probe coordinates as

$$\mu_i = \frac{X_i r_{11} + Y_i r_{12} + Z_i r_{13} + t_x}{X_i r_{31} + Y_i r_{32} + Z_i r_{33} + t_z} \tag{4.12}$$

and

$$\nu_i = \frac{X_i r_{21} + Y_i r_{22} + Z_i r_{23} + t_y}{X_i r_{31} + Y_i r_{32} + Z_i r_{33} + t_z} \tag{4.13}$$

In an ideal case, the following relation would therefore be true:

$$(\mu_i - \frac{X_i r_{11} + Y_i r_{12} + Z_i r_{13} + t_x}{X_i r_{31} + Y_i r_{32} + Z_i r_{33} + t_z})^2 + (\nu_i - \frac{X_i r_{21} + Y_i r_{22} + Z_i r_{23} + t_y}{X_i r_{31} + Y_i r_{32} + Z_i r_{33} + t_z})^2 = 0 \tag{4.14}$$

Due to error introduced by identifying points with a hand-held probe, and with a cursor on the computer screen, the image and probe points will not relate perfectly through a straight rigid body transformation. Equation 4.14 is still quite useful, however, in that a transformation can be sought which most closely relates the probe and image coordinates. Manipulating this relation to be solvable in a linear least-squares sense, the following equation results:

$$(\mu_i[X_i r_{31} + Y_i r_{32} + Z_i r_{33} + t_z] - [X_i r_{11} + Y_i r_{12} + Z_i r_{13} + t_x])^2 + \qquad (4.15)$$

$$(\nu_i[X_i r_{31} + Y_i r_{32} + Z_i r_{33} + t_z] - [X_i r_{21} + Y_i r_{22} + Z_i r_{23} + t_y])^2 = 0 \qquad (4.16)$$

By taking the video image coordinates, μ_i, ν_i, of at least six features, and their corresponding probe coordinates, X_i, Y_i, and Z_i, the transformation matrix elements can be determined. An assumption is made that the function minimum corresponds to the minima with respect to each variable of the transformation $(r_{11}, r_{12}, r_{13} ... t_z)$. The solution of this equation is taken to be the least-squares approximation of the function minimum, found by setting the derivative with respect to each variable to zero. The set of equations corresponding to these minima can be expressed in terms of the following matrix equation:

$$\mathbf{A} \times \left\{ \begin{array}{cccccccccccc} r_{11} & r_{12} & r_{13} & t_x & r_{21} & r_{22} & r_{23} & t_y & r_{31} & r_{32} & r_{33} & t_z \end{array} \right\}^T = 0 \qquad (4.17)$$

where the symmetric matrix $\mathbf{A}$ is defined as follows:

$$\begin{array}{ccccccccccc}
X_i^2 & X_iY_i & X_iZ_i & X_i & 0 & 0 & 0 & 0 & -\mu_iX_i^2 & -\mu_iX_iY_i & -\mu_iX_iZ_i \\
Y_iX_i & Y_i^2 & Y_iZ_i & Y_i & 0 & 0 & 0 & 0 & -\mu_iY_iX_i & -\mu_iY_i^2 & -\mu_iY_iZ_i \\
Z_iX_i & Z_iY_i & Z_i^2 & Z_i & 0 & 0 & 0 & 0 & -\mu_iZ_iX_i & -\mu_iZ_iY_i & -\mu_iZ_i^2 \\
X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -\mu_iX_i & -\mu_iY_i & -\mu_iZ_i \\
0 & 0 & 0 & 0 & X_i^2 & X_iY_i & X_iZ_i & X_i & -\nu_iX_i^2 & -\nu_iX_iY_i & -\nu_iX_iZ_i \\
0 & 0 & 0 & 0 & Y_iX_i & Y_i^2 & Y_iZ_i & Y_i & -\nu_iY_iX_i & -\nu_iY_i^2 & -\nu_iY_iZ_i \\
0 & 0 & 0 & 0 & Z_iX_i & Z_iY_i & Z_i^2 & Z_i & -\nu_iZ_iX_i & -\nu_iZ_iY_i & -\nu_iZ_i^2 \\
0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -\nu_iX_i & -\nu_iY_i & -\nu_iZ_i \\
-\mu_iX_i^2 & -\mu_iX_iY_i & -\mu_iX_iZ_i & -\mu_iX_i & -\nu_iX_i^2 & -\nu_iX_iY_i & -\nu_iX_iZ_i & -\nu_iX_i & (\mu_i^2+\nu_i^2)X_i^2 & (\mu_i^2+\nu_i^2)X_iY_i & (\mu_i^2+\nu_i^2)X_iZ_i \\
-\mu_iY_iX_i & -\mu_iY_i^2 & -\mu_iY_iZ_i & -\mu_iY_i & -\nu_iY_iX_i & -\nu_iY_i^2 & -\nu_iY_iZ_i & -\nu_iY_i & (\mu_i^2+\nu_i^2)Y_iX_i & (\mu_i^2+\nu_i^2)Y_i^2 & (\mu_i^2+\nu_i^2)Y_iZ_i \\
-\mu_iZ_iX_i & -\mu_iZ_iY_i & -\mu_iZ_i^2 & -\mu_iZ_i & -\nu_iZ_iX_i & -\nu_iZ_iY_i & -\nu_iZ_i^2 & -\nu_iZ_i & (\mu_i^2+\nu_i^2)Z_iX_i & (\mu_i^2+\nu_i^2)Z_iY_i & (\mu_i^2+\nu_i^2)Z_i^2 \\
-\mu_iX_i & -\mu_iY_i & -\mu_iZ_i & -\mu_i & -\nu_iX_i & -\nu_iY_i & -\nu_iZ_i & -\nu_i & (\mu_i^2+\nu_i^2)X_i & (\mu_i^2+\nu_i^2)Y_i & (\mu_i^2+\nu_i^2)Z_i
\end{array} \qquad (4.18)$$

and where each successive row of $\mathbf{A}$ is the derivative of Equation 4.16 with respect to r_{11}, r_{12}...(etc.) In the twelve rows of $\mathbf{A}$ only two linearly independent relations are expressed. They are

$$\left\{ \begin{array}{cccccccccccc} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & \mu_i X_i & \mu_i Y_i & \mu_i Z_i & \mu_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & \nu_i X_i & \nu_i Y_i & \nu_i Z_i & \nu_i \end{array} \right\} \times \left\{ \begin{array}{c} r_{11} \\ r_{12} \\ r_{13} \\ t_x \\ r_{21} \\ r_{22} \\ r_{23} \\ t_y \\ r_{31} \\ r_{32} \\ r_{33} \\ t_z \end{array} \right\} = 0 \qquad (4.19)$$

It is because two linearly independent relations exist for each point that six points must be identified to determine the twelve unknowns. In order to exclude the obvious but meaningless possibility of setting all of the transform elements to zero, one element of the matrix is set to unity and the corresponding column is brought to the right side of the equation.

Projected probe display

Once this raw transformation matrix is determined, it may be used to generate a probe rendering. The probe-space parameters for the position and orientation of the VW probe are sent to the program and transformed by the newly acquired matrix. Image coordinates for the probe rendering are then generated using Equations 4.4 and 4.5. The result is a graphical representation of the VW probe (see Figure 4.5) which is continually updated.

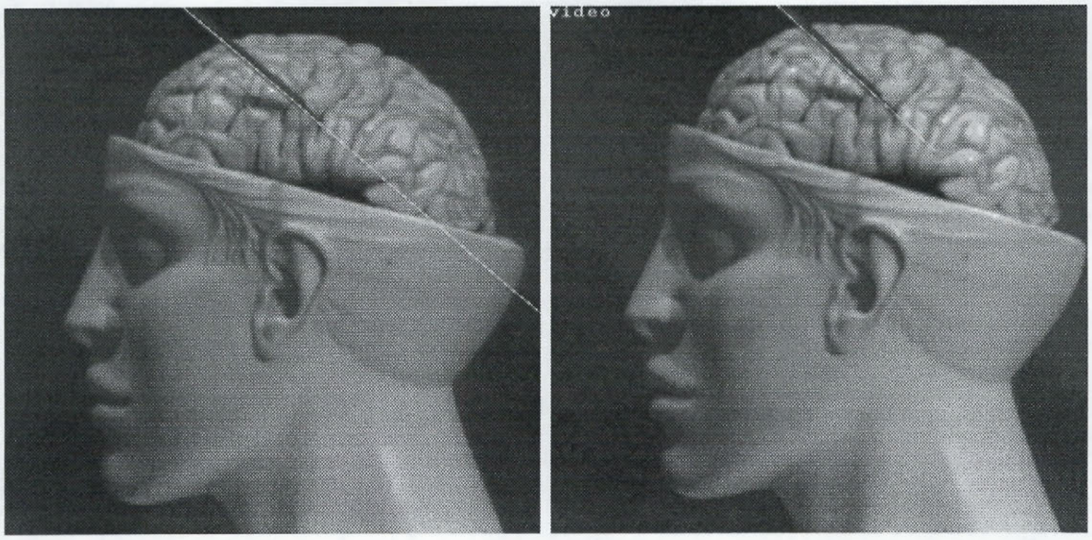


Figure 4.5: Stereoscopic images of the 2-D graphically generated probe (red) and the line of its trajectory (white) in proper alignment with the video probe.

4.3.3 Decomposing the transformation matrix

In order to implement the transformation in the IAP environment, it must be broken into its rotation, translation, and scale components which are loaded into a "tx3" object (Section 4.1.1). The rotation, scale, and translation are then applied by the "tx3" object to the 3-D model according to:

$$x' = \mathbf{R}x - \vec{t} \tag{4.20}$$

where $\mathbf{R}$ is the rotation matrix, with any scaling factor included, $\vec{t}$ is the translation vector, x represents the untransformed coordinate, and x' is the resultant transformed coordinate. The steps involved in separating the transformation components for input into the "tx3" object are shown diagramatically in Figure 4.6.

Rotation matrix components

Due to the unknown scale of the calculated transformation matrix, the rotation component of the transformation must be conditioned before it can be used in the IAP context. In addition to normalizing it, conditioning the 3×3 rotation component of the transformation also orthogonalizes the columns. This conditioned rotation matrix is applied to the 'cloud' of raw probe points.

Corrective scale factor

The centroids of the two homologous point clouds, in the video and probe spaces, are calculated. The rotated probe points are then translated to be centred about the origin of probe-space, and the video points are translated so that their centroid coincides with the image origin (Section 3.4). The ratio of the root mean square of the displacements of both sets of points from their respective centroids in the $\hat{x} - \hat{y}$ plane determines the scale difference between the probe and video coordinate systems. The difference of the video centroid and the rescaled probe system centroid is taken as the translation component of the transformation matrix.

Because one element of the original transformation matrix is set to unity (Section 4.2.2), an unknown scale factor is introduced. Only the *ratios* of the transformed object coordinates are related to the image coordinates, therefore only their magnitude relative to one another is required to be consistent for a solution. If the element which is initially set to unity is, in reality, close to zero, other larger matrix elements will blow up. To avoid this, the elements of the transformation matrix solution are examined, the largest is set to unity, and a solution is sought with this new constraint.

The image of the 3-D model on the screen is projected orthographically while the video image is the result of a diverging-ray projection (see Figure 4.2). In the case of the 3-D model, its distance from the image plane does not affect its appearance.

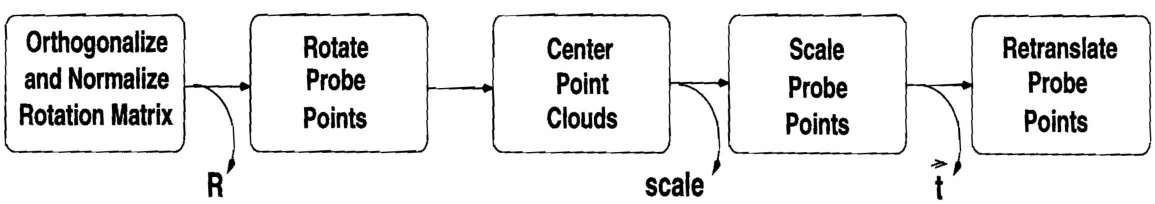


Figure 4.6: Steps in Decomposing Transformation Matrix into $\mathbf{R}$(rotation), scale, and $\vec{t}$(translation).

What matters, and what the described transformation aims to accomplish is that the objects within probe-space are rotated, translated and scaled to minimize the error resulting from the differing projection geometries. This is equivalent to scaling by the average z' value of the probe point cloud. Again, if the truncated pyramid approach for representing 3-D objects with converging, rather than parallel viewing geometry (Section 4.2.3) were possible, it would overcome the discrepancies due to the differing projection geometries.

4.4 The Video Fusion Program

The video fusion program is an adaptation of a separate software application which displays and performs three dimensional measurements on stereoscopic DSA projections. A diagram of the object pipeline for the video fusion application is provided in the Appendix. This section describes the function of the video fusion program. The steps in using the module which will be discussed can be summarized as follows:

- Starting Viewing Wand Software
- Importing Solid Objects
- Viewing Options
- Controlling Solid Object Orientation
- Registration Process
- Using utilities.

Figure 4.7 shows the basic user interface to the video fusion module. The VW should be running when the video fusion software is started. Before using either the VW or the video fusion software, the IAP processing server must already be activated.

Starting Viewing Wand Software

In preparation for running the video fusion application, a patient data-set must be loaded into the VW software. The patient data-set contains segmented objects such as the patient skin, and other anatomical structures of interest. These segmented objects and a representation of the probe functionality, are treated as Solid objects (Section 4.1.1) in the IAP environment. The standard registration of the probe to tomographic data is performed in order to display the VW probe accurately, relative to the other Solid objects. This relative orientation is then exported with the Solid objects into the video application.

Importing Solid Objects

In the call to open the initial video fusion display window, the software checks what Solid objects are being displayed in the VW application if it is running. When the VW is run and a patient data-set is loaded containing segmented structures, their solid objects are connected to a Solid View(Sv) object in the display pipeline. A file is generated with the external handles of all objects connected to the Sv. Solids representing patient anatomy in the video fusion application are all imported from the VW software. The video fusion software scans the "handles" file for connected Solids, then takes a "watch" on the relevant Solids, or creates its own client side Solids to reflect the server Solid objects shared by both applications (see Section 4.1.1). However, once a probe has been chosen in the VW the video fusion application must be explicitly requested to connect with the VW probe. Control of the display of these Solid objects is discussed below.

Viewing Options

In order to expose underlying structures like the corpus collosum, solids like the skin object can be disconnected from the VW viewing pipeline. Once they are disconnected, when the fusion application rescans the handles file, the Solids that have been disconnected in the VW application are absent and will be removed from the video fusion. The fusion application can update the Solid object connection list at any time. The user can also update the video image displayed by grabbing a new frame when necessary. In addition to removing an object from the display window altogether, parts of solid objects may be "cut" away along planes designated by the probe.

At any time, the video or the solid objects may be viewed independently, or the two can be fused in the manner described above (Section 4.2.4). When the user wishes, either the left or right view of the stereoscopic pair can be displayed. The pair can also be viewed simultaneously, either juxtaposed, one above the other, or in stereoscopic viewing mode (Section 4.2.4).

Solid Object Orientation Control

Once the Solids are loaded into the video application, the position of the probe Solid object moves with the physical probe relative to the anatomical Solid objects. The rendering perspective of the group of Solid object can be manually manipulated in three dimensions using sliders (see right panel in Figure 4.7). The disparity between the perspectives of the left and right views can also be controlled explicitly.

Registration Process

The orientation of the video imaging apparatus relative to probe-space is determined by activating a registration tool, shown in black in the upper left corner of Figure 4.7, then selecting twelve or more points in both video pairs. In order to select the fiducial points, the user moves a cursor, displayed on the the screen, to the desired image point, then presses a mouse button. A graphical circle is then superimposed on the video image,[4] and in order to centre the fiducial marker in the graphical circle, fine

[4]The priority of displaying the VW probe is set so high that it must be hidden during this part of the procedure in order to display the graphical circles.

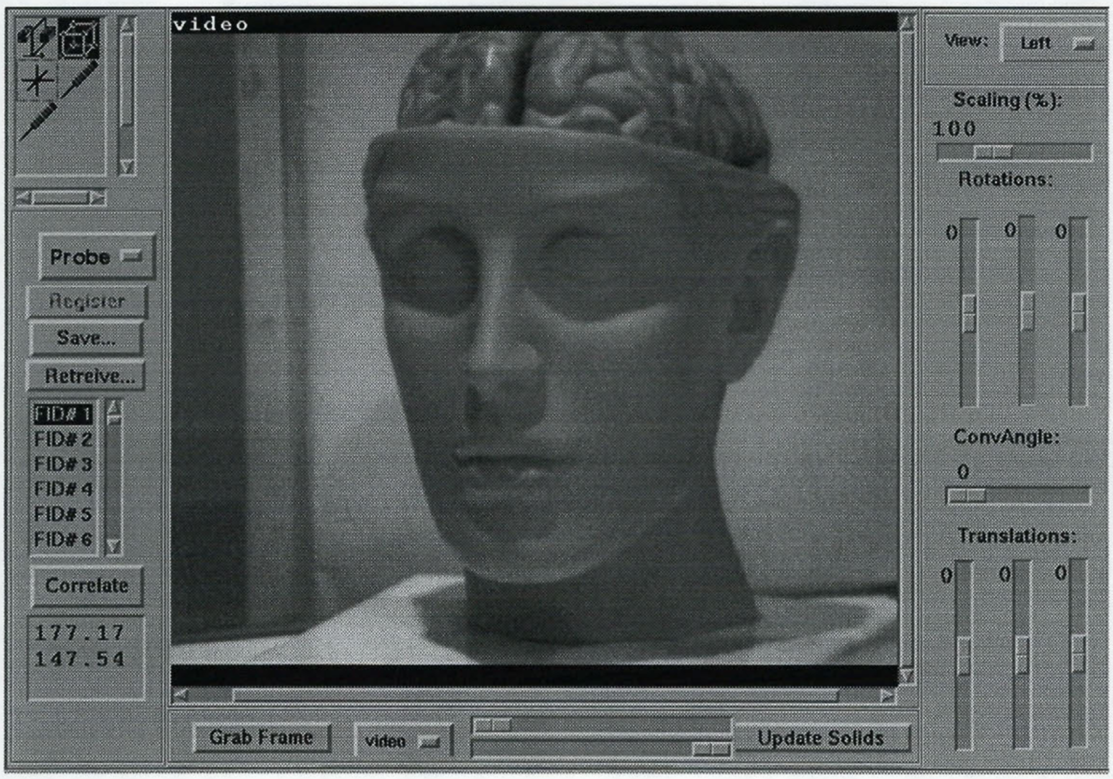


Figure 4.7: User interface of the video image fusion program.

adjustments can be made using arrow keys on the computer keyboard. The diameter of the graphical circle is six pixels, centred around the point selected by the user. The image coordinates of these points are stored in structures which also contain the number of points recorded in each of the stereoscopic image pair, and information about graphical marker representations. The physical features corresponding to these points are then identified on the patient with the probe, and the probe position at each landmark is recorded. The position of the probe pointing to each of these features is then correlated with the corresponding video coordinates, as described in Section 4.3.2 and a transformation is derived. At this point the user may display the graphical perspective representation of the probe if desired.

The Solids are reoriented by transforming first between tomographic data and probe-space, then from probe to camera-space. The transformation matrix representing the VW probe registration to the anatomical data is inverted and applied to the "Sv" through a "tx3" object. The newly calculated transformation parameters are loaded into a "tx3" object which is connected to the "Sv" subsequent to the inverted VW "tx3".

Utilities

The greyscale pixel values of the video images can be manipulated to emphasize features of interest adjusting the window width, and window level sliders. The video, and either the graphical or the Solid object probe can be shown simultaneously after registration. Currently the video is presented as a grey-scale image, while merged objects are depicted in colour, providing a convenient means of visually separating the images from the various sources.

Chapter 5

Results

During an operation, surgeons need as much freedom of movement as is possible. Statically mounted cameras are easily obscured and would restrict the movement of the surgeons while the stereoscopic video was in use. The optimal camera convergence angle may also change during an operation. It is therefore desirable to be able to reposition the cameras and to find their relation to stereotactic space without need for information external to the images. The following section describes experiments on phantoms, and simulated applications of fused video in the OR. Finally, the success of the fused video and tomographic data in these contexts is discussed.

5.1 Phantom Experiments

An experiment was performed to observe the differences due to parallel or perspective projection, in the tomographic data rendering and video imaging respectively. A phantom was constructed by inserting two transparent plates into two opposing sides of the OBT-stereotactic frame (see Figure 1.2). Each of the two plates has four dots arranged in a square which, in addition to landmarks on the frame itself, serve as fiducial markers for the experiment. The fiducial markers are distributed over a volume of 19cm×17cm×14cm, where the distance from the camera varies by 19cm within the volume. In a typical procedure, the depth of the volume of interest which is visible to the video will vary less than this.

The video cameras were set pointing horizontally with the phantom centred in each field of view. The vertical alignment was ensured by checking that the phantom image remained centred vertically without adjusting the cameras whether near or far from the phantom. As the cameras were repositioned, each stereoscopic view was again centred horizontally on the phantom. The physical configuration of the video cameras is shown in Figure 4.1. The distance from the centre of the camera to the centre of the phantom volume was measured and recorded. The zoom on the cameras was set as high as possible such that the fiducial markers used still fit vertically on the computer screen. The camera-subject separation varied to the limit of camera performance; the cameras were set as far and as near to the phantom as possible while the fiducials were reliably identifiable.

At this point, fiducial markers were identified in each video image and on the phantom using the VW probe, from which the video fusion application derived the transformation from probe-space to image-space. The probe coordinates are directly transformed before matrix decomposition, so that error introduced in the decomposition process is avoided. The perspective-projection of the image-space points is calculated from Equations 4.4 and 4.5 using the transformed probe coordinates. The root mean squared (RMS) deviation of the perspective and parallel-ray projected registration points from their input image coordinates was calculated. The orthogonal RMS values are taken from transformed x and y values, divided by the average z value of the group of fiducial markers. This is the effective result of the scaling calculation in the decomposition process (see Section 4.3.3).

Ten registrations were performed at each position, and the RMS deviations were averaged. A graph comparing perspective projection, and parallel projection RMS deviation vs. camera-subject separation is shown in Figure 5.1.

The parallel-projected coordinates are consistently less accurate than the perspective-projected coordinates. The perspective-projected coordinates do not become signifi-

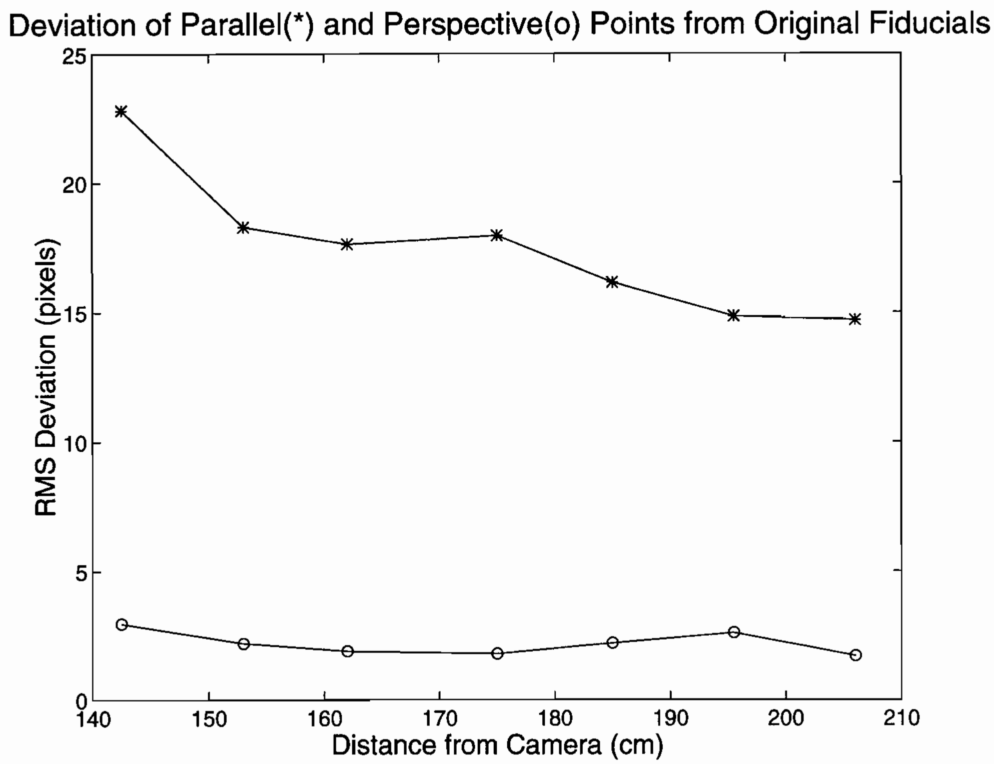


Figure 5.1: RMS Deviation of Parallel and Perspective Projected Points From Original Fiducial Markers.

cantly more or less accurate at different separations between the phantom and cameras. Positioning the graphical circles around the fiducial markers in the video image, and the location of the probe tip relative to the fiducial markers on the phantom are the two primary sources of error in the perspective-projected points. Orthogonal coordinates are less accurate as the cameras approach the phantom. The magnitude of the error of the parallel-projected points depends both on the distance of the fiducials from the camera and the fiducials' positions relative to one another. The fiducials in this experiment were arranged with an artifically large range in depth (19cm) so that a clear trend in the parallel-projection accuracy vs. distance from the camera was observable. This explains why the two curves in Figure 5.1 are so far apart. The separation demonstrates that the amount of error due to orthogonal projection is large relative to the error from imprecision in fiducial location. The two curves intersect where the distance from the camera is large enough that the method of projecting the points (orthogonal vs. perspective) becomes insignificant. In a surgical context where at most, half of a patient's head is visible to a camera at a time, a depth range of about 10cm is appropriate. The graphed data can be extrapolated to predict errors on the order of seven pixels according to Equation 4.3. In the situation where the depth range of the video image is 2cm, and when the registration fiducials are close to the plane of interest, the discrepancy may be reduced to 1-2 pixels.

The consistency of probe correspondence was checked, relative to the probe distance from the principal-ray. In order to accomplish this, a standard VW registration was performed, following which the probe was laid in some stationary position and the VW probe was 'frozen' or de-activated (see Section 4.1.1). The VW solids (including the probe representation) were manually aligned to the video. A video registration to probe space was then performed, and the transformation derived was used to generate the graphics probe. The position of either the graphics probe or the VW probe object can be displayed continuously, but the video probe will be updated only at image acquisition. Because of this, the correspondence was checked by again fixing the physical probe, then acquiring a current video frame, followed by switching between displaying the VW object or graphical probe for comparison with the video. While the 2-D graphical probe registration to the video is valid, the VW probe drifts from the video probe as the physical probe leaves the central imaging axis, or principal-ray. This trend is a result of the orthogonal projection of the VW objects, as predicted by Equations 4.1 and 4.2. The behavior would disappear if the tomographic data volume was first altered according to the truncated pyramids correction (Section 4.2.3).

5.2 Operating Room Simulations on Humans

Examples of video images fused with 3-D models of patient anatomy are given in Figures 5.2 and 5.3.

Figures 5.4 and 5.6 demonstrate that video can show objects used to perform surgery within the stereotactic image space. The stereoscopic viewing mode allows the orientation of the instruments in the video to be seen relative to the 3-D model of patient anatomy. This type of image can be used in the operating room to indicate where the VW probe is accurately registered, and whether a shift of one of the registered components has invalidated the VW probe registration (Figures 5.4 and 5.5).

An example of how a planned craniotomy may be displayed to indicate what anatomy lies directly beneath, is given in Figure 5.6. The integrated video can give feedback to surgeons even while they are drawing a contour. This, and other kinds of skin markings can be used to guide a procedure. Skin markings can also to serve in the initial video registration since they are visible in the video and can be located with the probe.

The current application successfully derives the transformation matrix, as demonstrated by the 2-D graphical probe display. Its position relative a subject, shown in the video concurs with the physical probe's relation to the subject. The concatenation of the IAP transformation objects (tx3's) fails, however. This results in inappropriate rendering of the anatomical solid objects. They are often moved out of the field of

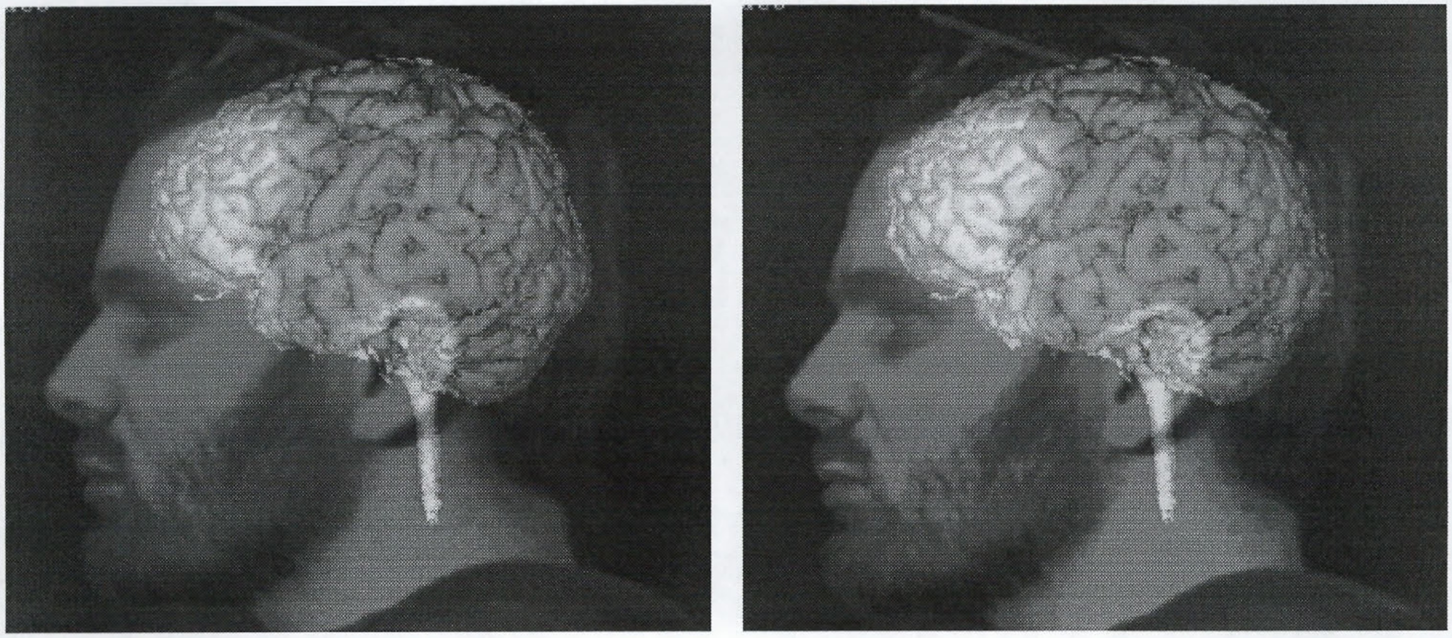


Figure 5.2: Stereoscopic pair of stereotactic cortex and subject skin in video.

Figure 5.3: Deep internal structures shown relative to skin in video.

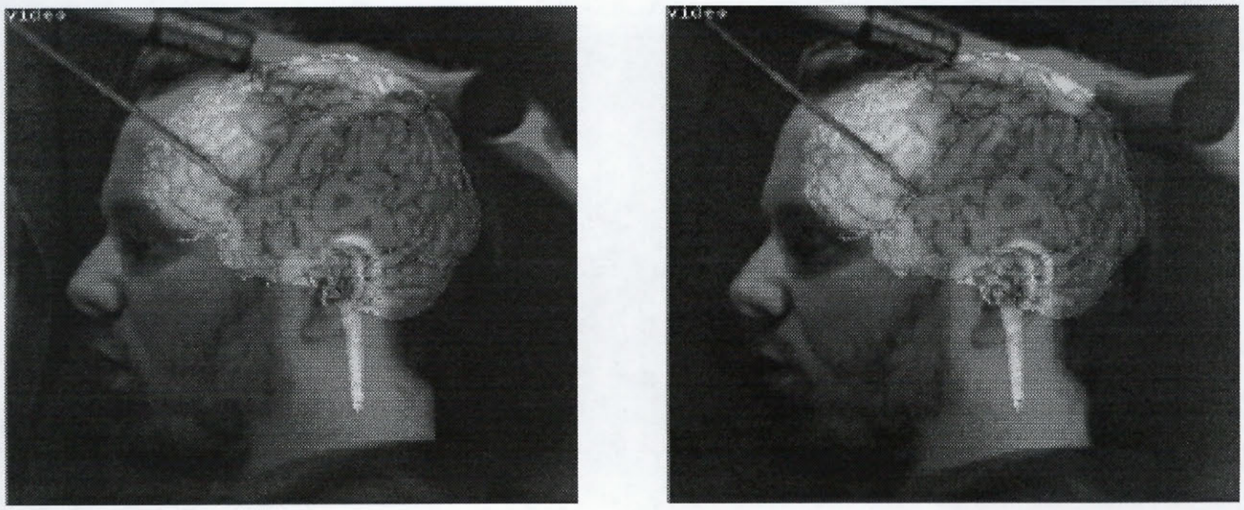


Figure 5.4: Good stereotactic and video probe alignment.

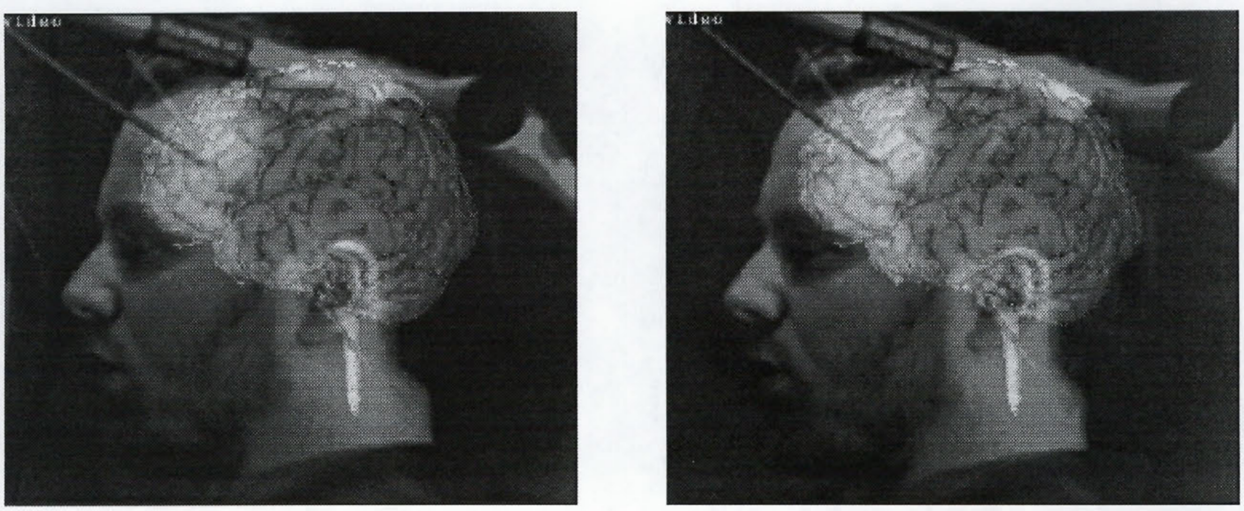


Figure 5.5: Bad stereotactic and video probe alignment.

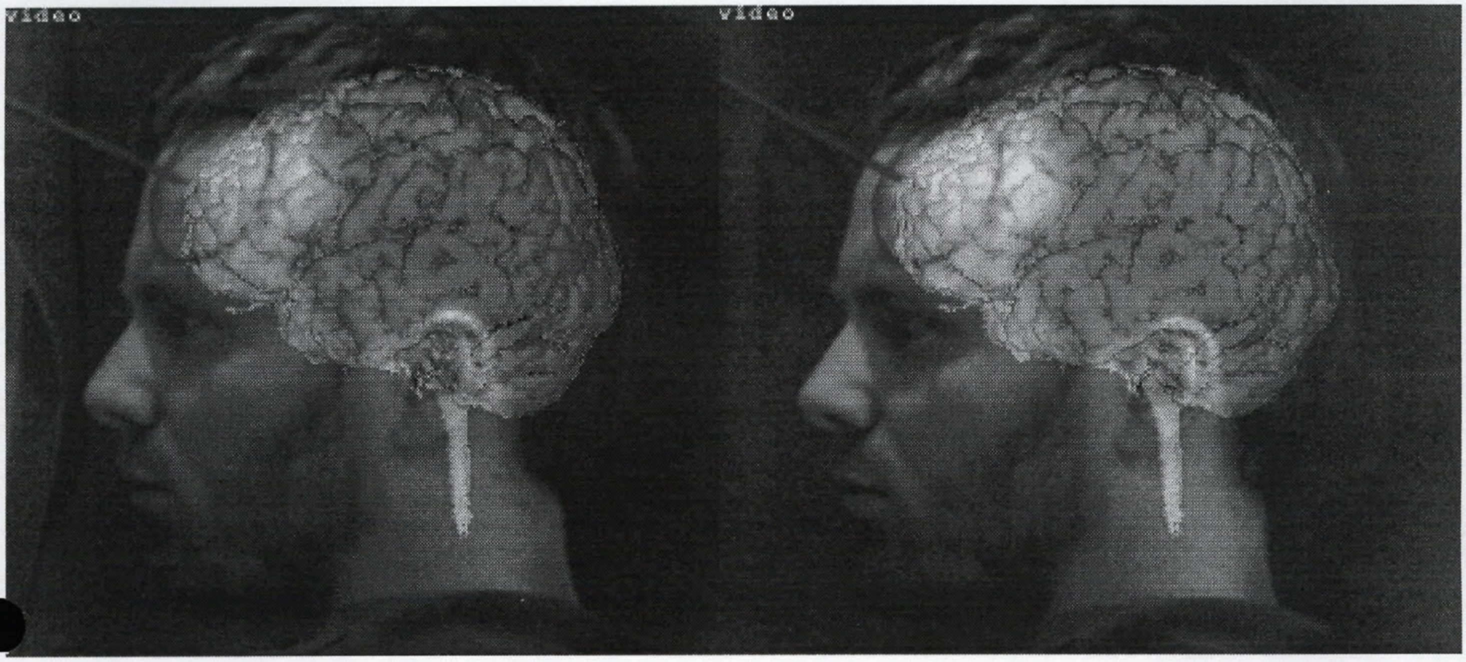


Figure 5.6: Planned craniotomy shown relative to cortex surface.

view completely. The potential reasons for this are that:

- The transformation parameters from the derived matrix are not correctly loaded into the tx3 object.
- The VW tx3 to register the probe object to the anatomical solid objects is either not inverted or not concatenated correctly.
- The IAP rendering geometry does not ensure that objects centred about the z axis appear centred in the rendered view.

Chapter 6

Conclusions

6.1 Summary

Stereotactic neurosurgery in medicine has been evolving since 1908 with its introduction by Horsley and Clarke. With the improvement of imaging techniques and the engineering of trackable pointers, the frame space of stereotaxy has to some extent been replaced by the imaged volumes and the tracking space of various hand-held pointers. The more recently developed orientation approach is called frameless IGNS. The display of these pointers or probes relative to the neuroanatomical images has established live feedback to, or interaction with, the stereotactic environment from the OR. The IGNS system at the MNI already has enhancements built upon the standard VW functionality, and this non-standard functionality is available to neurosurgeons through the availability of additional image modalities and modality combinations.

This thesis demonstrates the incorporation of stereoscopic video into IGNS. The full value of adding such a rich source of feedback to the IGNS environment will only be fully appreciated after it has been evaluated in use in the OR. Although the IGNS environment supplied by the VW already has the live update on the probe position, this information is expanded with the introduction of stereoscopic video. Video offers viewing of operating instruments and markings on the skin, confirmation of probe registration, and direct demonstration of brain distortion or shift.

The registration of the probe is currently checked by setting its tip on easily identifiable landmarks in the OR, and comparing that known position with the stereotactic representation. The addition of video allows the neurosurgeon to continually analyze the accuracy of the probe display, by directly comparing its image superimposed on the video display on the computer monitor. Homogeneous transformation matrices are used to display a perspective projection of a graphically generated probe (as opposed to the orthographically projected probe supplied by the VW), for consistent and accurate comparison of the probe feedback with the video.

The most useful applications of video derive from its inclusiveness as an imaging medium. Any visual events shown in the video are inherently registered to stereotactic space. Markings and objects in the OR are automatically shown relative to the 3-D stereotactic models. Although the VW probe already brings live interaction with stereotactic space into the OR, the probe itself is not used to perform surgery. With stereoscopic video, actual operating instruments can be followed in stereotactic space while a procedure is performed.

The results of comparing accurate perspective projection with the corrected parallel projection technique show the superiority of the diverging-ray projection but that for geometries likely to be found in the IGNS context parallel-ray projection gives a reasonable approximation to the video projection. The video cameras can be removed from the patient by more than 1.5m, at which distance differences in the final pixel values of parallel-projected points are not visually observable.

The technique presented for combining tomographic and projection image data involves rendering the tomographic volume to yield a view that matches the video projection, allowing it to be overlaid. Stereoscopic viewing of these overlaid images gives 3-D information about objects represented in the video and the stereotactic environment. MRI is the modality of choice because it is already correlated to other modalities and most commonly used in the operating room.

Another approach to check the accuracy of the work-station images is to manually align a video camera so that its image can be superimposed on the stereotactic images with a video-mixer. This simple integration of video has already been demonstrated to be useful by Kikinis et al. [31, 20] in the operating room, however the manual alignment typically takes 45 min. This same group now uses a laser scanner [41] to reconstruct the patient surface in the operating room, then finds the RMS error minimization of the reconstruction's registration to the 3-D model. Colchester [28] has also repeated the registration of video to MRI, using a patterned light source for surface reconstruction from stereo video views where they use a cost-minimization approach is used for the surface matching. Other groups are also using video in the stereotactic environment [54, 2]. In the implementation presented in this thesis, points in the video image are selected, and those points need to be matched to the corresponding object points with the wand. This registration procedure is similar to the probe registration to the MRI data.

A group at Dartmouth College, headed by Roberts, began developing a system to integrate pre-recorded images with an operating microscope view [42]. In 1989 Friets et al.[16] presented this system as a frameless stereotaxic operating microscope which combines anatomical images with images from the OR. The microscope and patient position are monitored with an ultrasonic range-finder, and the image data are projected through the microscope optics by a miniature cathode ray tube and a beam splitter. Edwards et al. [10] have developed a three step procedure to feed the correct view of an anatomical model into a microscope view. They first calibrate the microscope optics, then register the patient to an image coordinate system, originating at the camera optical centre, then they track the microscope using LEDs [10].

An alternative approach to establishing live feedback to the surgeons in this environment is by ultrasound [18]. The advantage that ultrasound has over video is that internal landmarks can be used for registration and tracked whereas video can only represent the outside of the cortex or exposed structures. An ultrasound image shows acoustic impedance, or echogenic boundaries, which in a normal head are primarily the sulci and ventricles as well as blood flow from doppler signals offering the possibility of registration based on vessel positions. The transducer must be put directly against the cortex, which makes it more invasive than video, and volume image acquisition takes about 4s. Depth resolution is dependent on the frequency used to scan, but can be sub-mm, while the lateral resolution depends on the beam width and is on the order of 5mm. It may in future prove to be a cost effective competitor to interventional MRI, when used for the guidance of neurosurgical procedures.

6.2 Future Work

6.2.1 Surface Matching

If a depth or range map is generated, or the surface of the patient skin is reconstructed, it could be registered in three dimensions to the stereotactic model of the patient head. The surface can be reconstructed using laser scanning techniques [41], or by projecting textured light. To facilitate point correspondence for 3-D surface reconstruction a textured light pattern is projected onto the object while capturing the stereo video pair. Siebert and Urquhart have shown that projection of a random noise field increases the accuracy of the depth map derived from the stereoscopic views of a given 3-D surface[45].

6.2.2 Continuous Video Capture and Display

Registration is performed in IAP then the appropriate views of the tomographic objects are fed as graphics to "blend" with a live video source. Once the correct view of the 3-D model is attained, no further manipulation is necessary short of re-registration. Parameters describing the orientation of the objects in the IGNS can be used in the form that they are exported from the VW system. The VW would update the orientation of the probe relative to the anatomical objects rendered, and the relative orientation would remain valid. The renderings can then be taken as a graphics display, to be blended on the Galilleo Video board hardware by mixing analogue signals. The brightness or colour of the IAP rendered objects can be controlled or adjusted to emphasize desired contrasts.

Glossary

- **accommodation** - the automatic adjustment of the eye for seeing at different distances effected chiefly by changes in the convexity of the crystalline lens; also: the range over which such adjustment is possible.
- **Allegro** - a workstation at the MNI that accepts volumetric MRI or CT data and provides tools to segment, display and analyze 3-D structures.
- **convergence** - coordinated movement of the two eyes resulting in a turning in of the optical axes so that the image of a single point is formed on corresponding retinal areas.
- **CoP** - Centre of Projection.
- **Crystal eyes** - stereoscopic viewing system by Stereographics Corporation.
- **CT** - Computed Tomography; radiographic volume imaging modality.
- **DSA** - Digital Subtraction Angiography; radiographic 2-D imaging technique for visualizing cerebral vasculature.
- **EEG** - ElectroEncephaloGraphy; detection and record of brain waves.
- **FARO arm** - a mechanical pointing device, which is mounted on the operating table and connected to a probe.
- **fiducial marks** - points common to the two imaging modalities.

- **fuse** - to blend thoroughly by or as if by melting together fusible.
- **Ga-DTPA enhanced MRI** - Gadolinium doped MRI enhancer used to contrast vasculature with other brain tissue.
- **Gallileo Video board** - SGI video to computer interface which digitizes images, loads them into buffers, and transfers them to the computer.
- **IAP** - Image Applications Platform; object oriented, image manipulation and display environment developed by ISG.
- **IGNS** - Image Guided NeuroSurgery.
- **ISG** - I.S.G. Technologies Inc., Mississauga, Ontario, Can.
- **MIP** - Maximum Intensity Projection.
- **MRI** - Magnetic Resonance Imaging.
- **MRA** - Magnetic Resonance Angiography.
- **optical centre** - point of convergence of a video camera.
- **parallax** - the apparent displacement or the difference in apparent direction of an object as seen from two different points not on a straight line with the object; angle of instantaneous parallax- is the difference in the angles of convergence for two points at different depths from a viewer.
- **perspective fore-shortening** - a phenomenon where the projection size of an object is inversely proportional to its distance from the CoP.
- **PET** - Positron Emission Tomography.
- **principal ray** - the line perpendicular to the imaging plane which passes through the focal spot.

- **registration** - to make or adjust so as to correspond exactly or to be in correct alignment.
- **render** - to reproduce or represent by artistic or verbal means: DEPICT.
- **stereopsis** - stereoscopic vision, or binocular depth perception.
- **stere- or stereo-** [comb form NL (fr. Gk, fr. stereos solid)] 1: solid: solid body (stereotaxis) 2a: stereoscopic (stereopsis) 2b: having or dealing with three dimensions of space (stereochemistry).
- **-tax-is** [n comb form, pl -tax-es NL (fr. Gk, fr. taxis)] 1: arrangement: order (homotaxis) 2: physiological taxis (chemotaxis).
- **ste-reo-tax-ic** [NL stereotaxis, stereotaxic technique (fr. stere- + taxis)] (1919) :of, relating to, or being a technique or apparatus used in neurological research or surgery for directing the tip of a delicate instrument (as a needle or an electrode) in three planes in attempting to reach a specific locus in the brain – ste-reo-tax-i-cal-ly.
- **ste-reo-tac-tic** [NL stereotaxis, after such pairs as NL hypotaxis: E hypotactic] (1950) :STEREOTAXIC.
- **Tx3 object** - a programming object in the IAP platform which manages a set of 3-D transform parameters.
- **Solid object** - a programming object in the IAP platform which manages a group of different objects which taken together represent a 3-D object (or group) which can then be displayed.
- **VW** - Viewing Wand; an interactive, frameless, stereotaxy environment by ISG.

Appendix

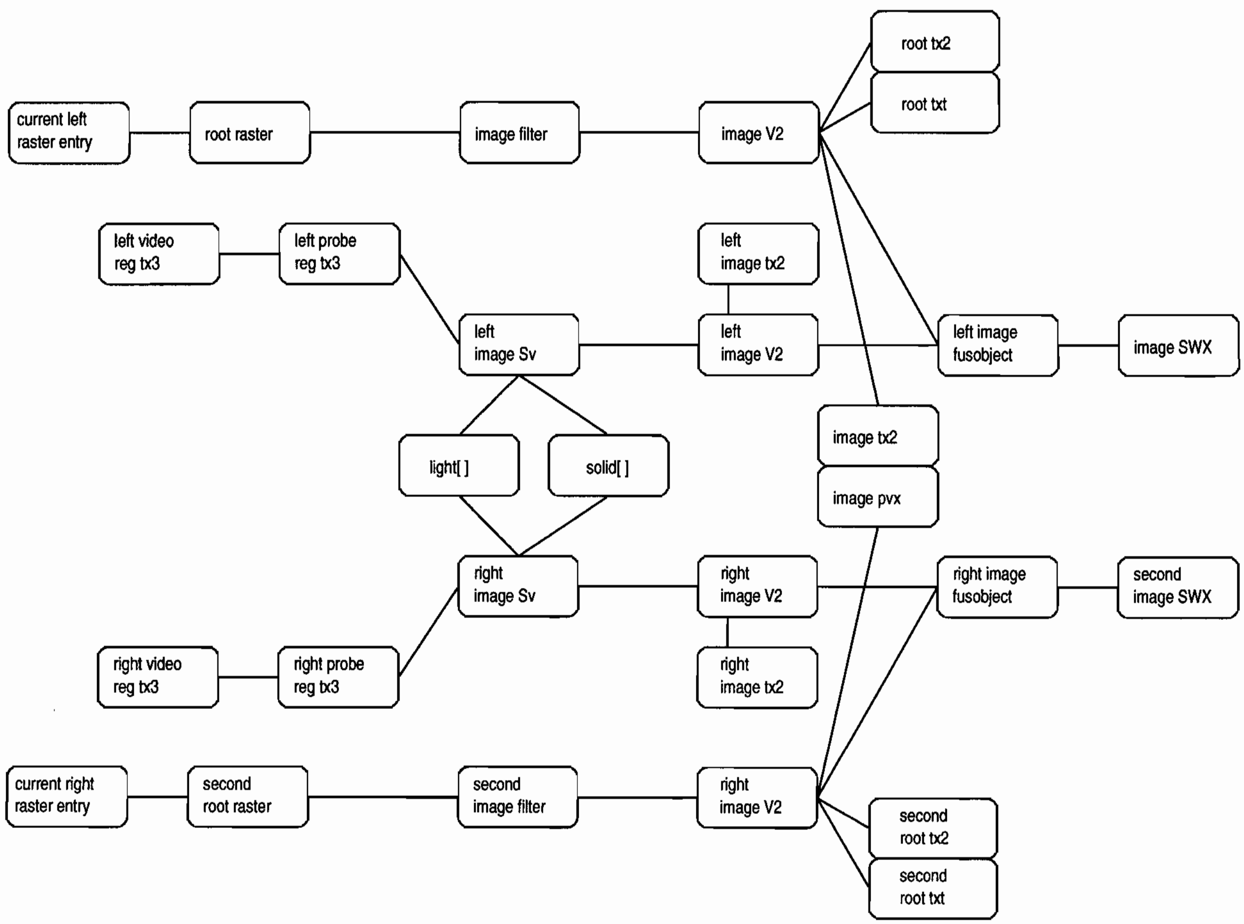


Figure 6.1: The object pipeline for the video fusion application.

Bibliography

[1] M. Bergström and T. Grietz. Stereotactic computed tomography. *American Journal of Roentgonology*, 127:167–170, 76.

[2] F. Betting, J. Feldmar, N. Ayache, and F. Devernay. A new framework for fusing stereo images with volumetric medical images. In *First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 30–39. Springer–Verlag, April 3-6 1995.

[3] P. Charland and T. M. Peters. Investigation of optimal conditions for viewing a stereopic medical image display for neurosurgery planning. *IEEE Transactions on Medical Imaging*, 1996 (in press).

[4] D. L. Collins, T. M. Peters, W. Dai, and A. C. Evans. Model based segmentation of individual brain structures from mri data. In *Proceedings of Conference on Visualization in Biomedical Computing*, pages 10–23. SPIE press, Oct. 1992.

[5] A. Cuschieri. Laparoscopic management of cancer patients. *Journal of the Royal College of Surgeons of Edinburgh*, 40(1):1–9, 1995.

[6] Bruce Davey, Patrice Munger, Roch Comeau, Laura Pisani, Andre Olivier, and Terry Peters. Applying stereoscopic image visualization to neurosurgical image guidance. In Anthony M. DiGioia, Takeo Kanade, and Russell Taylor, editors, *Proceedings of the First International Symposium on Medical Robotics and Com-*

puter Assisted Surgery, pages 264–271, Pittsburgh, Penn., September 22-24 1994. Shadyside Hospital.

[7] Bruce L. K. Davey, Roch M. Comeau, Patrice Munger, Laura J. Pisani, Daniel Lacerte, Andre Olivier, and Terence M. Peters. Multimodality interactive stereoscopic image-guided neurosurgery. In Richard A. Robb, editor, *Proceedings of Third Conference on Visualization in Biomedical Computing*, volume 2359, pages 526–536. SPIE, Oct. 4-7 1994.

[8] J. F. Donat. Long-term eeg monitoring for difficult seizure problems. *Journal of Child Neurology*, 9 Suppliment(1):S57–63, 1994.

[9] David Drascic. Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing. In *Proceedings of the Human Factors Society 35th Annual Meeting*, volume 2, pages 1367–1371, 1991.

[10] P. J. Edwards, D. L. G. Hill, D. J. Hawkes, R. Spink, A. C. F. Colchester, A. Strong, and M. Gleeson. Neurosurgical guidance using the stereo microscope. In *First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 555–564. Springer–Verlag, April 3-6 1995.

[11] L. E. Fencil, K. Doi, and K. R. Hoffman. Accurate analysis of blood vessel sizes and stenotic lesions using stereoscopic DSA system. *Investigative Radiology*, 23:33–41, 1988.

[12] R. J. Fitzgibbons Jr., R. Annibali, and B. S. Litke. Gallbladder and gallstone removal, open versus closed laparoscopy, and pneumoperitoneum. *American Journal of Surgery*, 165(4):497–504, 1993.

[13] C. M. Foley, A. Legido, D. K. Miles, and Grover W. D. Diagnostic value of pediatric outpatient video-eeg. *Pediatric Neurology*, 12(2):120–4, 1995.

[14] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics Principles and Practice.* Addison–Wesley, Reading, Mass, 1990.

[15] W. A. Friedman and R. J. Coffey. Stereotactic surgical instrumentation. In M. P. Heilbrun, editor, *Stereotactic Neurosurgery.* Williams and Wilkins, 1988.

[16] E. M. Friets, J. W. Strohbehn, J. F. Hatch, and D. W. Roberts. A frameless stereotaxic operating microscope for neurosurgery. *IEEE Transactions on Biomedical Engineering*, 36(6):608–17, Jun 1989.

[17] R. L. Galloway Jr., R. J. Maciunas, and C. A. Edwards II. Interactive image-guided neurosurgery. *IEEE Transactions on Biomedical Engineering*, 39(12):1226–1231, Dec 1992.

[18] Robert L. Galloway Jr., W. Andrew Bass, Hakan Senel, and Steven Schreiner. Enhanced interactive surgical guidance using both preoperative and intraoperative imaging. In *Proceedings of SPIE Medical Imaging*, volume 1897, pages 69–76, 1993.

[19] P. L. Gildenberg. Stereotactic surgery: Present and past. In M. P. Heilbrun, editor, *Stereotactic Neurosurgery*, pages 1–15. Williams and Wilkins, 1988.

[20] W. E. L. Grimson, G. J. Ettinger, S. J. White, P. L. Glenson, T. Lozana-Perez, W. M. Wells III, and R. Kikinis. Evaluating and validating an automated registration system for enhanced reality visualization in surgery. In *First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 3–12. Springer–Verlag, April 3-6 1995.

[21] P. Heffernan. *IAP - introduction for application writers.* ISG Technologies Inc., Missisauga, Canada, 3.2.1 edition, April 1993.

[22] P. Heffernan and D. Dekel. Imaging applications platform: Concept to implementation. *Proc. Conference on Visualization in Biomedical Computing*, 1808:13–16, October 1992.

[23] M. P. Heilbrun, P. McDonald, C. Wiker, S. Koehler, W. Peters, V. Sieminov, and C. Wiker. Implementation of a machine vision method for stereotactic localization and guidance. In R. J. Maciunas, editor, *Interactive Image-guided Neurosurgery*, pages 169–178. American Association of Neurological Surgeons, 1993.

[24] C. J. Henri, D. L. Collins, and T. M. Peters. Multi-modality image integration for stereotactic surgical planning. *Medical Physics*, 18(2):167–177, Mar/Apr 1991.

[25] C. J. Henri, D. L. Collins, G. B. Pike, A. Olivier, and T. M. Peters. Clinical experience with a stereoscopic image workstation. *Proceedings of SPIE - The International Society for Optical Engineering*, 1444, 1992.

[26] V. Horsley and R.H. Clarke. The structure and functions of the cerebellum examined by a new method. *Brain*, 31:45–124, 1908.

[27] G.N. Hounsfield, J. Ambrose, B.J. Perry, and C. Bridges. Computerized transverse axial scanning (tomography). *Brittish Journal of Radiology*, 46:1016–1051, 1973.

[28] Gongzhu Hu and George Stockman. 3-D surface solution using structured light and constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4), 1989.

[29] A. Kato, T. Yoshimine, T. Hayakawa, Y. Tomita, T. Ikeda, M. Mitomo, K. Harada, and H. Mogami. A frameless, armless navigational system for computer-assisted neurosurgery. technical note. *Journal of Neurosurgery*, 74(5):845–9, May 1991.

[30] Patrick J. Kelly. *Tumor Stereotaxis*. W. B. Saunders Company, 1991.

[31] Ron Kikinis, P. Langham Gleason, William Lorensen, William Wells, W. E. L. Grimson, Tomas Lozano-Perez, Gil Ettinger, Steven White, and Ferenc Jolesz.

Image guidance techniques for neurosurgery. In Richard A. Robb, editor, *Proceedings of Third Conference on Visualization in Biomedical Computing*, volume 2359, pages 537–540. SPIE, Oct. 4-7 1994.

[32] L. Lipton. *Foundations of The Stereoscopic Cinema.* Van Nostrand Reinhold Company Inc., New York, NY, 1982.

[33] R. J. Maciunas, editor. *Interactive image-guided neurosurgery.* American Association of Neurological Surgeons, 1993.

[34] B. D. F. Methley. *Computational models in surveying and photogrammetry.* Blackie, Glasgow, 1986.

[35] P. Milgram, D. Drascic, and J. Grodski. A virtual stereographic pointer for a real three dimensional video world. *INTERACT '90, Third IFIP Conf. on Human-Computer Interaction*, August 1990.

[36] P Milgram, S Zhai, D Drascic, and J Grodski. Applications of augmented reality for human-robot communication. In *International Conference on Intelligent Robots and Systems Proceedings*, pages 1467–1472. IEEE, IEEE Service Center, Piscataway, NJ, USA,(IEEE cat n 93CH3213-6), 1993.

[37] T. M. Peters, B. Davey, R. Comeau, C. J. Henri, P. Munger, P. Charland, A. Evans, and A. Olivier. On-line stereoscopic image-guidance for neurosurgery. In Lowell Klaisner, editor, *IEEE Medical Imaging Conference, San Francisco*, volume 3, pages 1805–1809, Nov. 4-6 1993. Paper 8M01.

[38] T. M. Peters and C. J Henri. Clinical applications of integrated 3-d stereoscopic imaging in neurosurgery, in course notes, stereoscopic visualization workshop. In *First Conf on Visualisation in Biomedical Computing*, Atlanta, May 1990.

[39] T. M. Peters, C. J. Henri, P. Munger, A. M. Takahashi, A. C. Evans, B. L. K. Davey, and A. Olivier. Integration of stereoscopic DSA and 3-D MRI for image

guided neurosurgery. *Computerized Medical Imaging and Graphics*, 18(4):289–299, Jul-Aug 1994.

[40] T. M. Peters, C. J. Henri, G. B. Pike, J. A. Clark, D. L. Collins, and A. Olivier. Integration of stereoscopic dsa with 3D image reconstruction for stereotactic planning. *Stereotactic and Functional Neurosurgery*, 55:471–476, 1990.

[41] Marc Rioux, G. Bechthold, D. Taylor, and M. Duggan. Design of a large depth of view three-dimensional camera for robot vision. *Optical Engineering*, 26(12):1245–1250, December 1987.

[42] D. W. Roberts, J. W. Strohbehn, J. F. Hatch, W. Murray, and H. Kettenberger. A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. *The Journal of Neurosurgery*, 65:545–549, 1986.

[43] Robert Rohling, John M. Hollerbach Patrice Munger, and Terry Peters. Comparison of relative accuracy between a mechanical and an optical position tracker for image-guided neurosurgery. *Journal of Image Guided Surgery*, 1(1):30–34, 1995.

[44] J. A. Salo. The role of videothoracoscopy in the diagnosis and treatment of chest diseases. *Annals of Medicine*, 26(6):401–4, 1994.

[45] J. P. Siebert and C. W. Urquhart. Active stereo: Texture enhanced reconstruction. *Electronics Letters*, 26(7):427–429, 29 March 1990.

[46] L. Solaini, P. Bagioni, and U. Grandi. Role of videoendoscopy in pulmonary surgery: present experience. *European Journal of Cardio-Thoracic Surgery*, 9(2):65–8, 1995.

[47] R. L. Sollenberger and P. Milgram. Stereoscopic computer graphics for neurosurgery. In G. Salvendy and M.J. Smith, editors, *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, pages 294–301. Elsevier Science Publishers B.V., Amsterdam, 1989.

[48] Randy L Sollenberger and Paul Milgram. Comparative study of rotational and stereoscopic computer graphic depth cues. In *Proceedings of the Human Factors Society*, volume 2, pages 1452–1456. Human Factors Soc Inc, Santa Monica, CA, USA, 1991.

[49] Randy L. Sollenberger and Paul Milgram. Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors*, 35(3):483–499, Sep 1993.

[50] Stereographics Corporation, P.O. Box 2309, San Rafael, CA 94912. *3Display: handbook for stereoscopic computer graphics*, 1985.

[51] Thant Tessman. Perspectives on stereo. In *Proceedings of SPIE Stereoscopic Displays and Applications*, volume 1256, pages 22–27. SPIE, 12-14 Feb 1990.

[52] W. H. Theodore, R. J. Porter, P. Albert, K. Kelley, E. Bromfield, O. Devinsky, and S. Sato. The secondarily generalized tonic-clonic seizure: a videotape analysis. *Neurology*, 44(8):1403–7, 1994.

[53] D. G. T. Thomas. *Stereotactic and image-directed surgery of brain tumours.* Churchill Livingstone, Edinburgh, 1993.

[54] M. Uenohara and T. Kanade. Vision-based object registration for real-time image overlay. In *First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 13–22. Springer–Verlag, April 3-6 1995.

[55] S. Zhai and P. Milgram. A telerobotic virtual control system. *SPIE Proceedings, Cooperative Intelligent Robotics in Space II*, 1612, 1991.

[56] Shumin Zhai and Paul Milgram. Human performance evaluation of manipulation schemes in virtual environments. In *IEEE Annual Virtual Reality International Symposium*, pages 155–161. IEEE, IEEE Service Center, Piscataway, NJ, 1993.

[57] S. J. Zinreich, S. Tebo, D. M. Long, H. Brem, D. Mattox, M. E. Loury, C. Vander Volk, W. Kotch, D. W. Kennedy, and R. N. Bryan. Frameless stereotactic integration of CT imaging data. *Radiology*, 188(3):735–742, September 1993.

Override totals:

System date/time for charges	0	Date due	0
System date/time for discharges	0	Policy table	0
Patron ID date < due date	0	Renewals limit	0

System activity totals:

Charges	0
Discharges	1
In-hand renewals	0
Not-in-hand renewals	0
Rush recalls placed	0
Recalls placed	0
Holds placed	0
Combined recalls/holds placed	0
Display items charged requests	0
Print items charged requests	0

2