

A Head-Mounted Display System for Augmented Reality Image Guidance: Towards Clinical Evaluation for iMRI-guided Neurosurgery

F. Sauer, PhD¹, A. Khamene, PhD¹, B. Bascle, PhD¹, and G.J. Rubino, MD²

¹ Imaging & Visualization Dept, Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA

{sauer, khamene, bascle}@scr.siemens.com

² UCLA, Division of Neurosurgery, Box 957039, Los Angeles, California 90095-7039, USA

rubino@surgery.medsch.ucla.edu

Abstract. We developed an augmented reality system targeting image guidance for surgical procedures. The surgeon wears a video-see-through head mounted display that provides him with a stereo video view of the patient. The live video images are augmented with graphical representations of anatomical structures that are segmented from medical image data. The surgeon can see e.g. a tumor in its actual location inside the patient. This *in-situ* visualization, where the computer maps the image information onto the patient, promises the most direct, intuitive guidance for surgical procedures. In this paper, we discuss technical details of the system and describe a first pre-clinical evaluation. This first evaluation is very positive and encourages us to get our system ready for installation in UCLA's iMRI operating room to perform clinical trials.

1 Introduction

Image guidance systems help the physician to establish a mapping between a patient's medical images and the physical body. In conventional systems, a pointer or an instrument is tracked and the location visualized in the medical images. The physician observes on a screen where the pointer or the instrument is positioned with respect to the internal anatomical structures. Hence, the conventional image guidance system maps the instrument into the medical data set and displays the relationship on a screen separate from the patient.

In contrast, augmented reality (AR) image guidance maps the medical data onto the patient's body. We propose the term "*in-situ*" visualization: anatomical structures are being displayed at the location where they actually are. The physician can see beyond the surface, the patient's body becomes transparent for him. This is the most direct and intuitive way of presenting the medical image information. Our work is concerned with exploring and realizing practical benefits of this *in-situ* visualization for image-guided procedures.

AR visualization in the medical field has first been suggested and investigated at UNC for ultrasound-guided procedures [1]. Further development of UNC's ultrasound

AR system is reported in [2], another AR system for laparoscopic surgery is described in [3]. A large image guidance program is ongoing at Harvard/MIT that includes research on augmented reality. [4] reports on a system for interventional MRI where the surgeon is provided with an augmented video image, taken with a single, fixed video camera and displayed on a monitor above the patient. At Guy's and St Thomas's Hospitals in London, stereo augmented reality has been implemented on a surgical microscope [5,6]. Various other AR projects are being pursued as the interest in AR technology is growing. Faster computers and better displays are making AR more practical and affordable.

We developed an AR system around a stereo head-mounted display (HMD) of the video-see-through variety. Two miniature color video cameras are mounted on the HMD as the user's artificial eyes. The two live video streams are augmented with computer graphics and displayed on the HMD's two screens in realtime. With the HMD, the user can move around and explore the augmented scene from a variety of viewpoints. The user's spatial perception is based on stereo depth cues, and also on the kinetic depth cues that he receives with the viewpoint variations.

We gave a first general description of our system in [8]. A copy of that system was built at the University of Rochester in the course of a collaboration that put the system in a neurosurgical context [9]. In this paper, we report on a pre-clinical evaluation at UCLA that encourages us to move our system towards clinical trials. Section 2 presents technical details of our AR system. Section 3 describes our pre-clinical experience and section 4 concludes with an outlook to the work we are planning.

2 System Overview

Designing an AR system requires at least three fundamental choices. The first is the choice of the display. We chose a stereo head-mounted display over a monitor, which is externally mounted. An HMD ultimately promises the most intuitive and natural experience of the augmented world. The user can move around and observe the scene dynamically from various viewpoints. His 3D perception is based on stereo and kinetic depth cues.

Closely linked to the display choice is the choice of whether to combine the computer graphics with an optical view of the real scene in an optical way (optical-see-through AR) or with a video view in an electronic way (video-see-through AR). Reference [10] reviews some medical AR systems of both types. Optical-see-through systems require less computing power and provide unmatched resolution of the "real part" of the augmented scene. The conventional optical-see-through systems are, however, based on semitransparent displays through which the user observes the real scene. The real view and the augmenting graphics are merged ultimately only in the user's eye, and their registration depends critically on the position of the user's eye behind the optical-see-through HMD. Correct registration is a subjective experience, which makes the system calibration imprecise and an external monitoring of the calibration impossible. Optical microscopes equipped for augmented reality visualization get around this problem by combining real and graphic views in an intermediate image plane (see e.g. [6-8]). Here the registration is fixed in an objective way, independent

of the user. We chose to work with video-see-through AR, where the combination of real and graphic views takes place in the computer. Registration is performed in an objective way. Moreover, video-see-through AR allows one the most control over the augmented image. Not only the graphics part but also the real part can now be manipulated for complete control over the final augmented visualization. This becomes essential when the surgical field contains textures and highlights that would interfere with the clear perception of the graphics information, unless one "cleans up" the video images (e.g. cuts out parts of them [2]).

The third choice concerns the tracking system. The tracking system is responsible for measuring the viewer's pose, i.e. position and orientation of the video cameras in the case of a video-see-through system. The pose information allows one to render the graphics from exactly the same vantage point and make it appear firmly anchored with respect to the real scene. Optical tracking technology achieves the highest precision. Most image guidance systems use optical tracking in the form of a stereo pair of video cameras. Instead of mounting such a commercial system next to our workspace, we employ a single tracking camera that we rigidly attach to the two other cameras on our HMD. The requirement that there be an unobstructed line-of-sight between tracker camera and optical markers is least restrictive in this head-mounted configuration. The user does not accidentally step in the way of the tracking system.

2.1 Hardware

The centerpiece of the system is a head-mounted display that provides the user with the augmented vision. Fig. 1 shows how three miniature cameras are rigidly mounted on top of the HMD. A stereo pair of color cameras captures live images of the scene. They are focused to about arm's length distance and are tilted downward so that the user can keep his head in a comfortable straight pose. A black-and-white camera, equipped with a wide-angle lens and a ring of infrared LEDs, detects retroreflective markers in the scene and is used for headtracking.

The combination of HMD and cameras weighs over 2kg, and one would not find it comfortable to wear it over extended periods of time. But with a resolution of 1024x768 (XGA) for each eye our system represents the highest resolution realtime AR system that we know of. In that respect it is well suited to investigate the potential benefits of AR visualization for image guided procedures. A fully integrated video-see-through head-mounted display weighing only 340g is described in [11]. We tested a prototype and found that its 640x480 resolution is not adequate for surgical application. The prototype makes it easy, however, to envision lightweight, comfortable, and high-resolution camera/HMD combinations for the future.

Fig. 2 shows how display and cameras are connected to two PCs. One SGI 540 processes the tracker camera images and renders the augmented view for the left eye, an SGI 320 renders the augmented view for the right eye. Both PCs communicate over an Ethernet connection to exchange information concerning camera pose, synchronization, and choice of graphics objects to be used for augmentation. Table 1 lists the particular hardware components that we are using.



Fig. 1. Video-see-through HMD with head-mounted tracker camera.

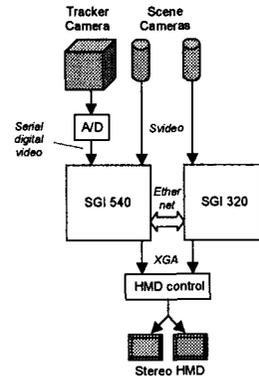


Fig. 2. System block diagram

Table 1. Hardware Components

HMD	Kaiser Proview XL35, XGA resolution, 35° diagonal FOV
Scene cameras	Panasonic GP-KS1000 with 15mm lens, 30° diagonal FOV
Tracker camera	Sony XC-77RR with 3.8mm fisheye lenses, 178° FOV
A/D converter	Miranda ASD-101i
Computers	SGI 540 and 320 with Windows 2000

2.2 System Calibration

For correct registration between graphics and patient, we need to calibrate the system. We need to determine the transformation that maps the medical data onto the patient, and we need to determine the internal parameters and relative poses of the three video cameras in order to show the mapping correctly in the augmented view.

Camera calibration and camera – patient transformation. Fig. 3 shows the calibration object we fabricated for the calibration of our camera triplet (which also doubles as a stage for AR experiments [8]). We determine the 3D coordinates of the markers with an Optotrak®, measure the 2D coordinates of the markers in the images, and calibrate the cameras based on 3D-2D point correspondences with Tsai’s algorithm [12,13]. For realtime tracking, we rigidly attach a set of markers with known 3D coordinates to the patient (resp. a head frame) defining the patient coordinate system.

MR data – patient transformation. We are currently preparing our AR system for use in UCLA’s neurosurgical iMR operating room [14]. The patient’s bed can be placed in the magnet’s fringe field for the surgical procedure or swiveled into the magnet for MR scanning. The bed with the head clamp, and therefore also the patient’s head, are reproducibly positioned in the magnet with a specified accuracy of ± 1 mm. We pre-determine the transformation between the MR volume set and the head clamp with a phantom and then re-apply the same transformation when mapping the scan data to the patient’s head. Fig. 4 shows the planar phantom that consists of a set of markers visible in the MR data set and a set of optical markers visible to the tracker

camera. We track the optical markers, and – with the knowledge of the phantom's geometry – determine the 3D locations of the MR markers in the patient coordinate system. We also determine the 3D locations of the MR markers in the MR data set, and calculate the transformation between the two coordinate systems based on the 3D-3D point correspondences.

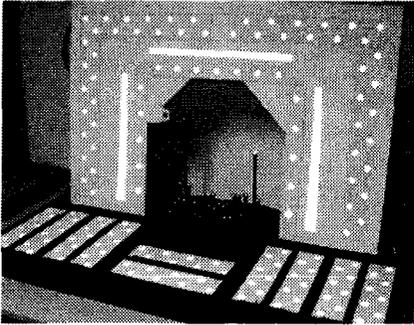


Fig. 3. Camera calibration object. The markers retroreflect the light of the camera's flash.

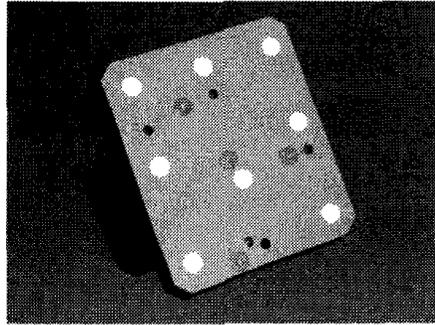


Fig. 4. Calibration phantom with MR and optical markers.

2.3 System Performance

Timing. Our AR video system is running at the full standard video rate of 30 frames per second. We synchronize video and graphics, eliminating any time lag between the real and the virtual objects. The virtual objects do not lag behind, neither does one see them swim or jitter with respect to the real scene. As the augmented view shows the graphics firmly anchored in the real scene, the user can assess the information in a comfortable way. Overall, there is a time delay of about 0.1 seconds between an actual event and its display to the user (see also section 3.2).

Registration accuracy. We initially performed an accuracy test of our AR system with the calibration setup shown in Fig. 3. The calibrated system was used to augment video images of the central marker configuration with a graphics model of the same marker configuration. The mismatches between the real markers and their virtual counterparts were recorded as errors. We found the errors to be typically below 1mm, going up to 2mm at the borders of the scene. We do not have any measurements yet with the head frame marker arrangement from Fig. 5. But the first application that we target, skin flap and craniotomy planning, does not require a high accuracy. For this application, our system's accuracy is well sufficient, even if the mapping between MR and patient space introduces additional errors in the range of a few millimeters.

3 Pre-clinical Experience

3.1 Plain Video View

The first natural concern with a video-see-through HMD is certainly this: to what extent does the video view diminish the ability to see fine details and perform delicate manual tasks? We summarize the subjective experience with our system in the following list. We gained this experience in a qualitative way through simple pick-and-place and point-and-touch experiments. One of the authors (Rubino) also tried out the use of surgical instruments on a cadaver while wearing the HMD.

Artificial eyes above head. Does the hand-eye coordination suffer when the artificial eye-point is displaced from the user's natural eye-point? When the user inserts his hand into the field-of-view, the hand does indeed appear at an unexpected position at first. But in our experience, by watching the hand's movement one is quickly able to adapt. Concentration on the limited workspace establishes natural hand-eye coordination. This situation may be similar to working under a microscope where the user can also adapt to a view that is different from what he is used to see with his naked eye. While it is possible to construct a video-see-through HMD with the camera viewpoints matching the user's eyepoints [11], it may not be necessary for surgical applications.

Video resolution. We capture the video images with a resolution of 720x648 and display them scaled to a size of 1024x768. Of course, the upscaling of the images does not increase their actual resolution, but the appearance of the graphics benefits from the use of the XGA display. A sharper video image (HDTV in the future?) would be welcome, but – again from a subjective perspective – the present video resolution seems to be adequate to perform at least the planning for the surgical work. We find a digital zoom helpful. Even though no new details are created in the video images, the magnification gives one an easier grasp of the details that are present.

Limited depth perception cues. Our pair of cameras has a fixed convergence, and the user looks at fixed screens where he does not adjust his focus according to the distance of the objects that he is observing. The main depth cue comes from the stereo disparity of the two video images. This stereo depth cue alone provides a good depth perception for people with normal stereo vision ability. With the HMD, the user can furthermore vary his viewpoint, and as close objects appear to move faster than more distant objects, he receives additional kinetic depth cues about the structure of the scene. We found the ability to freely vary the viewpoint very helpful to understand scenes where the 3D structure was not immediately apparent.

Fixed focus lenses. The lenses on our cameras are focused to about 60cm, the arm's length distance where the user performs manual work in a comfortable way. We roughly estimate the depth of field as ten centimeters. The user has to move his head into the right distance from the object to see it sharp. Yet, for the work on the surgical field the fixed focus may not be experienced as a limitation, especially for surgeons who are used to wearing magnifying binoculars with an even smaller depth of focus. Our HMD is also not completely immersive. The user can look past the miniature displays to the right and left and in particular to the bottom and have some direct view

of the surroundings. The surgeon may take advantage of this direct view for example when he has to reach out for an instrument that is being handed to him.

Time delayed visual feedback. There is an intrinsic time delay between an actual event and its display on the HMD. The CCD chip in the video camera has to be exposed, the image data read out, converted to digital format, written into computer memory, transferred to texture memory, written into the framebuffer, and finally being displayed at the monitor's next refresh cycle. This whole process may take about 0.1 s, only a small, but noticeable delay. This time delay in the visual feedback makes precise control of fast movements difficult. In testing the use of surgical instruments guided by video vision, our surgeon collaborator (Rubino) saw the time delay as a most critical feature. He feels, though, that a surgeon can learn to adapt his technique to the slight time delay in the visual feedback.

3.2 Augmented Video View

Augmentation is the value we are adding to the video view, which alone would be inferior to the direct optical view. The usefulness of the augmented video view determines the usefulness of the system. Visualization and perception issues become important, as the goal is to provide the user with intuitive, task-oriented guidance. We list here some relevant issues.

Time lag between video and graphics. The user gets easily disturbed when, by moving his head, he sees the graphics lagging behind the real objects in the scene. Such a time lag is unavoidable in optical-see-through AR. Here the real scene changes instantaneously with the movement, but tracking the movement and rendering the graphics always takes a finite time. In our video-see-through system we delay the display of a video frame until the corresponding graphics is ready and thereby avoid the time lag completely. The graphics appears firmly anchored with respect to the real objects, what makes it easier for the user to follow the guidance.

Increased time delay of visual feedback. As we synchronize video and graphics to avoid time lag, we are adding the time required for tracking and graphics rendering to the overall delay of the visual feedback. Our tracking software takes only about 10ms to provide the pose information given the tracker camera image, which is not significant. However, when we rendered complex graphics we easily ended up delaying the final augmented image unacceptably long. Fortunately, the need to work with simple graphics matches well with the goal of providing intuitive guidance. The concentration on relevant information helps the user, whereas too much unnecessary detail may only confuse him.

Depth perception and occlusion depth cues. We have a simple experiment where the user is asked to touch the wick of a virtual candle with the tip of a (real) wooden wand. After getting used to the augmented view, most test persons are easily able to touch the correct point in space and light up a virtual fire.

The depth perception becomes more difficult, when real and virtual objects are overlapping in a way that does not reflect their correct spatial relationship. We know the viewer's vantagepoint and can make the graphics objects appear at the desired 3D locations. A correct visual interaction between real and graphics objects, however,

would require 3D information about the real objects as well. This 3D information is usually not available, the graphics objects are simply superimposed onto the 2D image of the real scene. Real objects can be hidden by virtual objects, but not vice versa. This is the well-known occlusion problem in AR. It is very critical for applications like surgical guidance where the user needs to interact with the augmented scene.

For correct occlusion, one needs to obtain some 3D information of the real scene. We find, however, that one can reduce the disturbing effect of wrong occlusion cues significantly with appropriate rendering of the graphics. We show segmented structures not as solids, but as wire frames, not with thick lines but with thin lines, not opaque but semitransparent, not finely structured but coarsely structured where details are not important. Overall, we show only the relevant structures in a sparse representation.

3.3 Head Phantom Augmented with MR Data

Fig. 5 shows our styrofoam head phantom as it is inserted into a mock-up head frame. The head frame is equipped with a bridge of retroreflective markers. We can detect these markers with our head-mounted tracker camera and calculate the user's viewpoint with respect to the head frame. That enables us to show an augmented view of the head phantom with the video-see-through HMD.

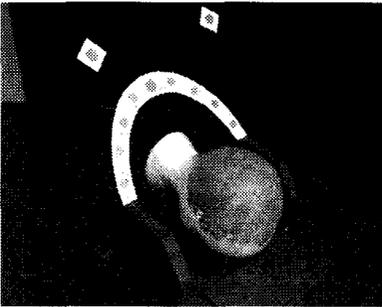


Fig. 5. Head phantom



Fig. 6. Testing the AR system.

For the augmentation, we segmented an MR data set that was taken with the Siemens OpenViva 0.2T magnet in UCLA's iMRI operating room. Figs 7 (a) and (b) are examples of two augmented views. In (a), the user sees - overlaid onto the head phantom - a set of yellow contour lines describing part of the skull, a blue wire frame model of the tumor, and one of the original MR slices. This view is helpful for studying the anatomy. In (b), the MR slice is omitted. This view is appropriate when the user wants to take action and, e.g., mark onto the head the outlines for a skin flap or craniotomy. Less graphics means less disturbance by the occlusion problem. One could even omit the skull contour lines, or at least limit their extent even further. The contour lines do not present essential anatomical information; they do support, though, the user's spatial perception of the scene. When looking at the pictures, one needs to bear in mind that the monoscopic images presented here cannot convey the user's

actual experience. This experience is strongly determined and enhanced by the stereo vision.

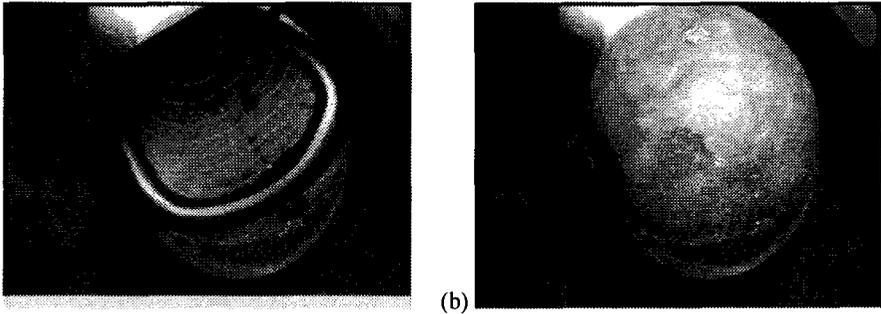


Fig. 7. Augmented views of the head phantom, showing skull contours and tumor. (a) with original MR slice, (b) without MR slice

Fig. 6 shows one of the authors examining the head phantom's augmented view. Seeing a model of the tumor in the tumor's actual location makes it very straightforward to plan appropriate access. The augmented vision supports craniotomy planning in a very intuitive and efficient way. From these initial experiences we feel strongly encouraged to move on to clinical tests. For critical cases, augmented reality guidance has the potential to become an important tool.

4 Conclusion and Future Plans

The initial evaluation of our AR system is very positive and encouraging. The head-mounted display – even though heavy and not very comfortable at present – gives the surgeon intuitive access to the structure of the anatomy. In a natural way he can explore the anatomy from various angles. Stereo and kinetic depth cues provide good 3D perception. Features that make our system particularly suited for medical applications are the high (XGA) resolution of the display, the elimination of time lag between video and graphics, and the low latency of the visual feedback of only about 0.1s.

We are currently preparing our system for use in UCLA's neurosurgical iMRI operating room [14]. Advanced augmented reality image guidance seems to be a good match with interventional MRI, where the surgical progress is monitored with a series of interoperative MR scans and the surgeon repeatedly has to map the new images onto his patient. We want to enable the surgeon to look at the surgical field and directly see e.g. the remaining parts of a tumor that still need to be resected. At first, we intend to use our system for skin flap and craniotomy planning, and then continue to explore further, more critical procedures.

References

1. M. Bajura, H. Fuchs, and R. Ohbuchi. "Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery within the Patient." Proceedings of SIGGRAPH '92 (Chicago, IL, July 26-31, 1992). In *Computer Graphics* 26, #2 (July 1992): 203-210.
2. Andrei State, Mark A. Livingston, Gentaro Hirota, William F. Garrett, Mary C. Whitton, Henry Fuchs, and Etta D. Pisano, "Technologies for Augmented Reality Systems: realizing Ultrasound-Guided Needle Biopsies," *Proceed. of SIGGRAPH (New Orleans, LA, August 4-9, 1996)*, in *Computer Graphics Proceedings, Annual Conference Series 1996*, ACM SIGGRAPH, 439-446.
3. Henry Fuchs, Mark A. Livingston, Ramesh Raskar, D'nardo Colucci, Kurtis Keller, Andrei State, Jessica R. Crawford, Paul Rademacher, Samuel H. Drake, and Anthony A. Meyer, MD, "Augmented Reality Visualization for Laparoscopic Surgery," *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI '98 (Cambridge, MA, USA, October 11-13, 1998)*, 934-943.
4. W. Eric L. Grimson, Ron Kikinis, Ferenc A. Jolesz, and Peter McL. Black, "Image-Guided Surgery," *Scientific American*, June, 1999, 62-69.
5. P.J. Edwards, D.J. Hawkes, DLG Hill, D. Jewell, R. Spink, A. Strong, and M. Gleeson, "Augmentation of Reality in the Stereo Operating Microscope for Otolaryngology and Neurosurgical Guidance," *Computer Aided Surgery* 1:172-178, 1995.
6. King AP, Edwards PJ, Maurer CR, de Cunha DA, Gaston RP, Clarkson M, Hill DLG, Hawkes DJ, Fenlon MR, Strong AJ, Cox TCS, Gleeson, MJ, "Stereo augmented reality in the surgical microscope," *Presence: Teleoperators and virtual environments* 9:360-368 2000.
7. W. Birkföllner, K. Huber, F. Watzinger, M. Figl, F. Wanschitz, R. Hanel, D. Rafolt, R. Ewers, and H. Bergmann, "Development of the Variscope AR, a See-through HMD for Computer-Aided Surgery," *IEEE and ACM Int. Symp. On Augmented Reality – ISAR 2000 (Munich, Germany, October 5-6, 2000)*, pages 54-59.
8. F. Sauer, F. Wenzel, S. Vogt, Y. Tao, Y. Genc, and A. Bani-Hashemi, "Augmented Workspace: Designing an AR Testbed," *IEEE and ACM Int. Symp. On Augmented Reality – ISAR 2000 (Munich, Germany, October 5-6, 2000)*, pages 47-53.
9. C. Maurer, F. Sauer, C. Brown, B. Hu, B. Bascle, B. Geiger, F. Wenzel, R. Maciunas, R. Bakos, A. Bani-Hashemi, "Augmented Reality Visualization of Brain Structures with Stereo and Kinetic Depth Cues: System Description and Initial Evaluation with Head Phantom," *SPIE Int. Symp. on Medical Imaging 2001 (San Diego, CA, February 2001)*.
10. J.P. Rolland and H. Fuchs, "Optical versus Video See-Through Head-Mounted Displays in Medical Visualization," *Presence (Massachusetts Institute of Technology)*, Vol. 9, No. 3, June 2000, pages 287-309.
11. A. Takagai, S. Yamazaki, Y. Saito, and N. Taniguchi, "Development of a Stereo Video-See-Through HMD for AR Systems," *IEEE and ACM Int. Symp. On Augmented Reality – ISAR 2000 (Munich, Germany, October 5-6, 2000)*, pages 68-77.
12. Roger Y. Tsai, "A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, August 1987, pages 323-344.
13. <http://www.cs.cmu.edu/~cil/v-source.html>: freeware implementation of the Tsai algorithm.
14. G.J. Rubino, K. Farahani, D. McGill, B. Van de Wiele, J.P. Villablanca, A. Wang-Mathieson, "Magnetic resonance imaging-guided neurosurgery in the magnetic fringe fields: the next step in neuronavigation," *Neurosurgery* 2000; 46: 643-654.