Computer Assisted ENT Surgery Using Augmented Reality: Preliminary Results on the CAESAR Project

R.J. Lapeer[1], P. Chios[1], G. Alusi[2], A.D. Linney[1], M.K. Davey[2], and A.C. Tan[2]

[1] Department of Medical Physics and Bioengineering, University College London (UCL), London WC1E 6JA - UK.
[2] Institute of Laryngology and Otology (ILO), University College London (UCL), London WC1X 8EE - UK.

Abstract. The 'Computer Assisted ENT Surgery using Augmented Reality' (CAESAR) project aims to improve ENT surgical procedures through augmentation of the real operative scene during surgery: a virtual scene, which shows structures that are normally hidden to the eye of the surgeon, is superimposed onto the real scene. The main distinction of this project as opposed to previous work in the field is to create a hierarchical and stepwise implemented system which allows operations such as calibration, tracking and registration to be assessed on an individual basis. This allows us to compare different alternatives for each operation and eventually apply the best solution without interfering with the performance of other parts of the system. In this paper, we present a framework for the alignment of the objects/subject in the real and virtual operating environment before the onset of surgery, and test its performance on a phantom skull. The operations involved are thus based on a static system and include calibration of the stereo microscope and registration of the virtual patient (as reconstructed from CT data) with the real patient. The final alignment of all objects in the real and virtual operating scene is assessed by cumulating maximum errors of each individual step.

1 Introduction

Computer assisted surgery (CAS) has its roots in the field of stereotactic neurosurgery where accurate mechanical systems (stereotactic frames) were used in conjunction with CT images and powerful low cost computers [8]. The introduction of algorithms for surface and volume rendering from CT and MRI images, powerful computer graphics, the use of computer vision techniques for calibration and the commercialisation of different types of tracking devices and shape sensors have, amongst other developments, contributed to significant improvements in this field. Today, CAS has also found its way in other disciplines such as ear, nose and throat (ENT) surgery, cardiac surgery, minimal invasive surgery, maxillo-facial surgery, eye surgery and many other surgical disciplines. *Augmented reality* (AR) locates itself in between reality (the real world) and

S.L. Delp, A.M. DiGioia, and B. Jaramaz (Eds.): MICCAI 2000, LNCS 1935, pp. 849–857, 2000.

virtual reality. Imagine a line of which the two extremes comprise the real world and a totally virtual environment (see Figure 1), then augmented reality (AR) is that part of the line which lies near to the real world extremum with the predominate perception being the real world augmented by computer generated data [9].
The application of the AR concept in CAS aims to provide the surgeon with more visual information than what he can typically perceive through a 3D stereo operating microscope during surgery. Important parts of the anatomy which are not normally visible during the operation can be perceived in the augmented image. The Microscope-Assisted Guided Intervention (MAGI) system as developed at Guy's Hospital [5], which has been evaluated on six patients in a clinical experiment, augments the surgical microscope image by projecting the preoperative images back into the binocular optics of the microscope. Alternatively, the augmented image can be displayed on a separate device such as a 3D stereo display [4].
In this paper, we describe an initial setup and calibration of the participating subjects/objects in an ENT operating scene and the corresponding alignment in the virtual scene.

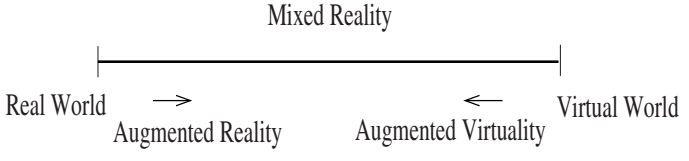

Fig. 1. The spectrum of mixed reality according to Milgram [9].

2 The Alignment of the Real and Virtual Operating Scene

The following subjects/objects are included in the operating scene and their individual coordinate systems need to be referred to one another:

1. the target object, in this case the patient's head,
2. the binocular stereo operating microscope,
3. one or more surgical tools,
4. a calibration object,
5. tracking devices.

Other assets in the operating scene include the 3D augmented display, the operating table and the operating lighting equipment and of course last but not least the surgeon, but they do not require alignment!
The virtual operating scene includes the same components as mentioned above. The aim of the alignment process is to get the objects in both the real and the virtual operating scene in the same relative position as opposed to each other,

before surgery begins. Since the moving objects in the scene, i.e. the subject's head, the 3D microscope and the surgical tools, are equipped with a tracking device, further changes in the individual coordinate systems as opposed to the reference or world coordinate system can be derived. For simplicity, the surgical tool(s) were excluded at this stage.

2.1 Stepwise Solution to the Alignment Problem

Starting from an initial (random) setup to the final alignment of the components in both the real and the virtual world, requires the following operations:

Microscope calibration The 3D surgical microscope is equipped with two CCD video cameras to capture the stereoscopic images. When combined and correctly aligned, the two images form a stereoscopic image that can be displayed by an autostereoscopic 3D monitor. This calibration step was discussed in [4]. The calibration which we discuss here involves the alignment of the microscope/camera assembly, which we will call the real camera from now on, with the virtual camera[1]. The end result of this operation is that the calibration object in the real world and its exact replica in the virtual world are viewed from the same position by the real and the virtual camera, respectively.

To allow easy calibration, a microscopic pattern of a number of points[2] is required. The subject's head would obviously be ideal to use as the calibration object but not enough landmarks or points, on a typically small area as seen with the microscope, are available. Therefore, an artificial calibration object was manufactured and used. The design is briefly discussed in section 3.1. The pattern of the calibration object is used to define the world coordinate system. This implies that the calibration object has to be mounted in the operating scene and should not move during the initial alignment until all objects in both the real and virtual world are aligned.

Referring the calibration object to the subject In the first step we aligned the real and virtual camera to be in the same relative position in relation to the real and virtual calibration object respectively. However, we aim to make the camera point at the patient rather than the calibration object, thus the position of the former to the latter has to be known. We use a tracking device to achieve this goal. The receiver of a tracking device can be fixed exactly to a frame which is part of the calibration object (see Figure 2), thus its position to the origin of the world coordinate system is known. After positioning the tracking device sensor onto the calibration object, the readout of the six degrees of freedom are set to 0. A similar frame is fixed to the (real) patient's head and the sensor is placed into this frame. New readings give us the position of the patient's head relative to the world coordinate system.

[1] Note that this is done separately for both cameras of the stereo couple.

[2] At least five for the Tsai calibration algorithm [10] but for optimal accuracy, preferably of the order of about 100.

Registration of the real head with the virtual head The 'virtual head' is usually obtained by volume or surface rendering of a set of CT or MRI images of the (real) patient's head. In case more than one dataset is used, a preliminary registration of these datasets is required. To register the real and virtual head, we first laserscan the real head with the tracking device as positioned in the previous step. A graphical interface, allows us to roughly register the laserscanned surface with the volume-rendered CT/MRI dataset. A combined distance transform/genetic algorithm allows us to automatically optimise the registration[3]. After registration, the location of the tracking device will now also be determined on the virtual head, thus the relative position of the latter will be known as opposed to the virtual calibration object.

Tracking the cameras Although we managed to relate the coordinate systems of the patient's head, the calibration object and the cameras to one another in both the real and the virtual world, we still cannot point the camera towards the patient's head without destroying the entire alignment. Therefore we simply place a tracking device on the camera of which the readings of the six degrees of freedom are set to zero. Any change of the camera's position (e.g. to point it at a particular area of the patient's head) will allow us to correspondingly align the virtual camera and point it at the same area of the virtual head.

3 Preliminary Experiments and Results

At this stage, only basic lab experiments have been performed. The patient's head is simulated by a dummy which is a skull model (see Figure 3). A first series of tests included camera calibration, referring the calibration object coordinate system to the subject coordinate system and the registration of the skull model with its virtual counterpart obtained from CT images.

3.1 Microscope Calibration

Calibration of the microscope or rather the calibration of each of the two CCD cameras mounted at the distal ends of the beam splitter typically involves the extraction of the extrinsic and intrinsic parameters. The *extrinsic parameters* are derived from the parameters of a Euclidean transform, i.e. the rotation and translation of a point in world coordinates into a point in camera-centered coordinates. The *intrinsic parameters* determine the projective behaviour of the camera, i.e. the principal point, which is defined as the intersection point of the optical axis and the image plane, and the camera constant or effective focal length. If radial lens distortion is considered, a number of distortion coefficients (two are usually sufficient) may be calculated as well.

[3] The region of the laserscanned surface, used for optimisation, should not include the tracking device for obvious reasons.

Calibration Algorithms The most basic calibration algorithm is based on a *Direct Linear Transformation* [1] and has the advantage that only linear equations have to be solved. However, it typically requires at least 100 points [7], which may pose a practical problem to fit such a number into the small area of the microscope's field of view whilst preserving sufficient accuracy. A more advanced and widely used algorithm in the field of robotics and computer vision, is the Tsai calibration algorithm [10]. This algorithm includes the treatment of lens distortion (calculation of a single first-order coefficient) and gives reasonable results for a relatively small number of points (fully optimised calibration requires at least 11 data points). Furthermore, an extra intrinsic parameter, i.e. a scale factor to account for any uncertainty in the framegrabber's resampling of the horizontal scanline of the camera, is calculated.

Calibration Objects The calibration object (see Figure 2) was manufactured on a CNC milling machine with an accuracy of 0.01mm. It consists of a parallelepiped base with six holes for mounting, a calibration prism and two pins for exact positioning of a Polhemus FASTRAK® tracking device. Two adjacent planes of the roof-like shaped calibration prism make an angle of 45° with the base. Each of them have a mixed pattern of calibration points of diameter 0.5mm and 1mm spacing in between points.

Fig. 2. Calibration object with attached Polhemus receiver.

Fig. 3. Skull used as the subject with attached Polhemus receiver.

Results A first series of laboratory experiments was performed using a Zeiss stereo operating microscope with two discrete magnification levels of 10 and 16, respectively. The focusing distance of the microscope's objective lens is 200mm. Figure 4a shows the image as captured by one of the two cameras. Eighty four points were used to derive the extrinsic and intrinsic camera parameters using the planar optimised Tsai calibration algorithm. Figure 4b shows the image as seen by the virtual camera. The first-order distortion coefficient, κ, was of the

order 10^{-6} which means radial distortion is negligible. The average (undistorted) image plane error was 1.0±0.6 pixels and the maximum image plane error, 2.7 pixels. This corresponds with an average object space error of 0.01±0.008mm and a maximum error of 0.04mm. Considering the accuracy of the points in the calibration pattern (0.01mm), a maximum calibration error of 0.05mm can be expected.

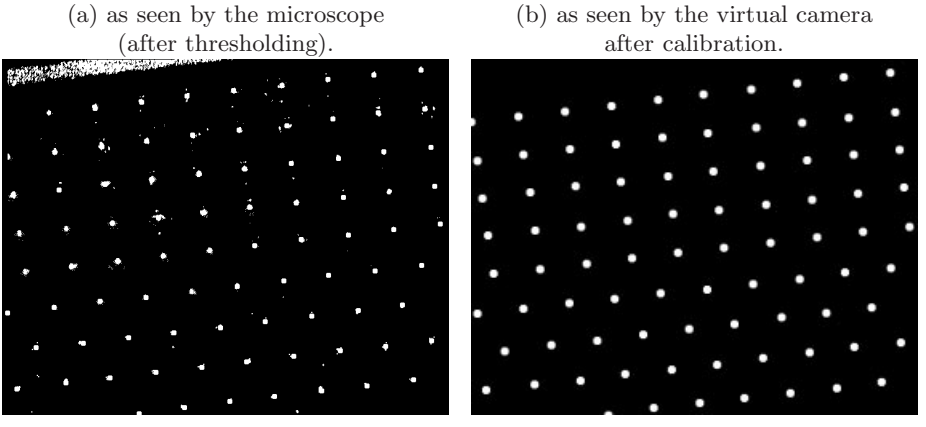

Fig. 4. Calibration pattern images: alignment of the virtual calibration object with the real calibration object - 84 points were used for calibration.

3.2 Referring the Calibration Object to the Subject

As described in section 2.1, this step (number 2) involves the real world only. At this stage, the Polhemus FASTRAK® was used as a tracking device. The static accuracy of this magnetic tracking device is around 0.8mm. The calibration object is foreseen of two cylindrical pins upon which the receiver fits with a maximum tolerance of 0.05mm (see Figure 2). A similar device with pins is placed onto a dummy which is the skull model as shown in Figure 3 (with Polhemus receiver mounted on it). Placing the receiver, first onto the calibration object, record the readings and then replace it to the dummy, allows us to relate the coordinate systems of the dummy and the calibration object. Consideration of all the steps involved yields a maximum error of about 1mm.

3.3 Registration of the Laserscanned Data with CT Data

The subject, in this case the skull model, was both CT scanned and laserscanned. The first operation would occur early pre-operatively, the second just before the operation starts (when the entire calibration procedure is initiated). A handheld Polhemus FastSCAN® laserscanner was used to obtain a surface of the subject.

The scan includes the tracking receiver which is fixed to the dummy. If we manage to register the laserscanned surface with the volume-rendered CT dataset, then we will know the position of the receiver on the virtual subject too, thus registering the virtual and the real subject. The algorithm to register the two datasets is based on a distance transform [3] of the voxel space which lies outside the object of interest (the virtual object). Only the points of the laserscanned surface are considered and are brought into the voxel space of the virtual object. When this is done arbitrarily, each point will fall into a region of constant distance to the virtual object surface. The ideal situation is when each point falls into the zero distance region which implies the two surfaces are registered. If we introduce the laserscanned points into the voxel space so that they are reasonably close to the object's surface, then minimising the sum of squared distances will yield an optimal fit between the surfaces. We used a genetic algorithm as the optimisation strategy. If the initial registration is sufficiently close (which can be easily achieved using a graphics interface), a global optimum will always be reached. Figure 5 shows the result before and after registration.

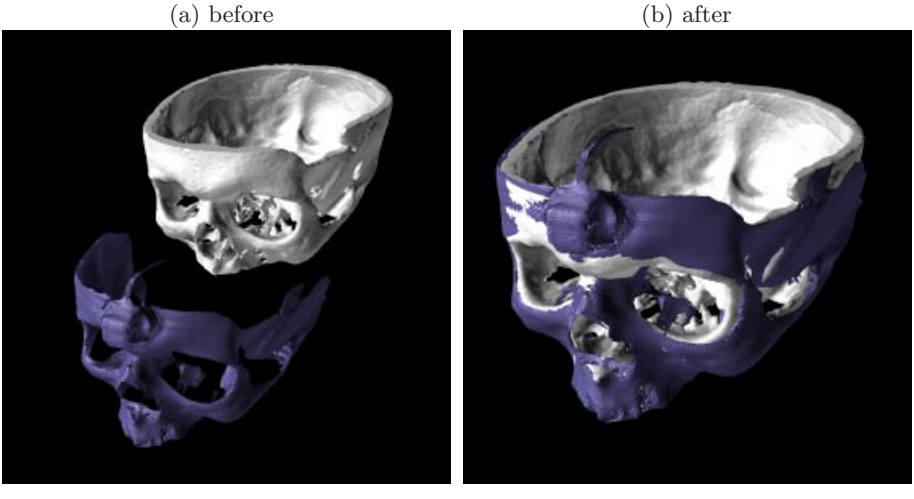

Fig. 5. Registration of the volume-rendered CT data and the laserscanned surface of the skull model.

4 Discussion

The planar and non-planar optimised Tsai calibration algorithms give stable results even for the rotational degrees of freedom. However, the solution is very sensitive to small perturbations in the input data and particularly to the accuracy by which the u,v coordinates in the image plane are determined. The accuracy improves with increasing number of calibration points. The current

calibration error could be improved to sub-pixel level by better estimates of the u,v image coordinates, by using automated centre detection and using the collinearity feature to optimise centre positions.
The Polhemus FASTRAK® is a relatively inexpensive tracking device and thus suitable for initial experimentation. It proves however to be one of the main sources of error for the total accuracy of the system. Moreover, the sensitivity of the device to the introduction of any metal object in the working space may cause errors of several millimeters when used in a clinical environment [8].
The registration algorithm, based on distance transforms at voxel level, yields results with an accuracy of around 1mm. Knowing the resolution of the Polhemus FastSCAN® is also about the same order of magnitude, maximum errors of around 2mm were found.
Since the policy of the CAESAR project is to avoid any invasive solution (e.g. fiducial bone implants) for the sole purpose of tracking or registration, the placement of tracking devices on the head of the patient will be accomplished using a dental stent and/or mouthpiece [5, 6] or strapping a plastic headband around the patient's forehead or use a purpose-built head-holder [2].
From the previous experiments, a maximum alignment error of about 4mm can be expected. The main sources of error are the registration and the tracking device, which require further optimisation. For example, replacement of the Polhemus FASTRAK® with an active optical system such as the Northern Digital OPTOTRAK® with typical static accuracy of 0.1-0.15mm for distances up to 2.25m may improve the overall accuracy.

5 Conclusion

We presented a general framework to align the participating objects and subject in an ENT surgical operating environment with those in the corresponding virtual environment, the latter having the sole purpose of augmenting the real operating scene as perceived by the surgeon through the stereo operating microscope. The framework provides a solid base to experiment with different devices and techniques for tracking, calibration and registration to improve the overall accuracy of the system as each step in the framework can be optimised separately without influencing other steps. We outlined a series of initial experiments on microscope calibration and registration of the subject with its virtual counterpart. Maximum alignment errors of up to 4mm can be expected. Further experimentation should allow us to compare overlay errors with the estimated maximum error as a summation of individual steps. Although, from a statistical point of view, overlay errors would nearly always be less than the maximum possible error, the latter remains the worst case error and can only decrease by improving the methodology of the system (i.e. by minimising the number of critical steps) or the technology of the individual devices and imaging techniques! Further work will also involve the assessment of the system in a dynamic, simulated environment and, after further optimisation, in a clinical environment.

Acknowledgements

We wish to thank the staff of the CT/MRI unit at the Royal Free Hospital, London, for the acquisition of the CT data. The first and second authors would like to thank the EPSRC and the Defeating Deafness organisation for funding this project.

References

[1] Y.I. Abdel-Aziz and H.M. Karara. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Proceedings ASP Symposium on Close Range Photogrammetry*, pages 1–18, Urbana,Illinois,USA, 1971.

[2] R.J. Bale, M. Vogele, and W. Freysinger. Minimally invasive head holder to improve the performance of frameless stereotactic surgery. *Laryngoscope*, 107:373–377, 1997.

[3] G. Borgefors. Distance transformations in arbitrary dimensions. *CVGIP*, 27:321–345, 1984.

[4] P. Chios, A.C. Tan, G.H. Alusi, A. Wright, G.J. Woodgate, and D. Ezra. The potential use of an autostereoscopic 3D display in microsurgery. In C. Taylor and A. Colchester, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 998–1009. Springer, September 1999.

[5] P.J. Edwards et al. Design and evaluation of a system for microscope-assisted guided interventions (MAGI). In C. Taylor and A. Colchester, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 842–851. Springer, September 1999.

[6] A.R. Gunkel, M. Vogele, A. Martin, R.J. Bale, W.F. Thumfart, and W. Freysinger. Computer-aided surgery in the petrous bone. *The Laryngoscope*, 109:1793–1799, 1999.

[7] R. Klette, K. Schlüns, and A. Koschan. *Computer Vision - Three-Dimensional Data from Images*. Springer, Singapore, 1998.

[8] S. Lavallée, P. Cinquin, and J. Troccaz. *Computer integrated surgery and therapy: state of the art*, volume 30 of *Technology and Informatics*, chapter 10. IOS Press, 1997.

[9] P. Milgram et al. Augmented reality: A class of displays on the reality-virtuality continuum. In H. Das, editor, *SPIE Proceedings: Telemanipulator and Telepresence Technologies*, volume 2351, pages 282–292, 1994.

[10] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.