THE INTERNATIONAL JOURNAL OF MEDICAL ROBOTICS AND COMPUTER ASSISTED SURGERY
Int J Med Robotics Comput Assist Surg 2008; **4**: 32–45.
Published online 14 February 2008 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/rcs.175

ORIGINAL ARTICLE

Image-enhanced surgical navigation for endoscopic sinus surgery: evaluating calibration, registration and tracking

R. Lapeer[1]*
M. S. Chen[1]
G. Gonzalez[1]
A. Linney[2]
G. Alusi[3]

[1]*School of Computing Sciences, University of East Anglia, Norwich, UK*

[2]*Ear Institute, University College London, London, UK*

[3]*St. Bartholomew's Hospital, London, UK*

*Correspondence to: R. Lapeer, School of Computing Sciences, University of East Anglia, Norwich, UK. E-mail: rjal@cmp.uea.ac.uk

Accepted: 12 December 2007

Abstract

Background Endoscopic sinus surgery (ESS) is generally applied to treat sinusitis when medication is not effective in eliminating the symptoms. Images captured by the endoscope are viewed on a monitor placed near the surgeon. Due to the separation of the handling of the endoscope from the viewing of the image, ESS requires surgeons to have well-trained hand–eye coordination. Unlike the use of the stereo surgical microscope in ENT, the endoscope does not provide the stereo cue for depth perception, hence a surgeon can only perceive depth through motion and shading, which may affect the accuracy of tool placement. Whilst the skill and experience of the surgeon are important factors to the success of ESS, the assistance of image-enhanced surgical navigation (IESN) can further reassure the surgeon's judgement and enhance surgical performance.

Methods We developed and validated an IESN system (ARView) for a rigid zero-degree endoscope, typically used for ESS. We present the interface, and calibration and registration (pre-operative and intra-operative) methods of the system. We then quantitatively assess the performance of each of the steps needed to generate the overlay of a real endoscope image with its 'virtual' counterpart, obtained from computed tomography (CT) image data of a real skull. These steps include calibration, registration, motion tracking and final overlay.

Results Calibration results using a planar calibration object displayed optimized object space errors of 0.025 ± 0.013 mm, whilst a non-planar calibration object displayed errors of 0.12 ± 0.08 mm. Target registration errors (TREs) near the region of interest (ROI), using our pre-operative registration method with the calibration object located near the mouth of the patient (skull), were 2.3 ± 0.4 mm. The proposed photo-consistency method for intra-operative registration has not yet yielded satisfactory results for ESS-based IESN. (RMS) values for tracking accuracy were found to be around 1.2 mm in a typical workspace of 400×400 mm. Object space overlay errors in a small measurement volume of $10 \times 10 \times 10$ mm were found to be around 0.4 ± 0.02 mm.

Conclusions We conclude that, in agreement with individual experiments, the current overall overlay accuracy is of the order of 2–3 mm in the x–y plane, which is in line with current conventional SN systems. The method which is most in need of improvement is registration, hence we wish to investigate the application of the proposed photo-consistency method further.

Keywords endoscopic sinus surgery; image-guided surgery; surgical navigation; augmented reality; image enhancement

Introduction

Image-enhanced surgical navigation (IESN)

Surgical navigation (SN), also referred to as image-guided surgery (IGS), is becoming increasingly more popular as a visual aid for minimally invasive surgery (MIS), since commercial systems were introduced more than a decade ago. As the term suggests, a SN system guides or navigates a surgical tool through the human body. Augmented reality (AR), also referred to as 'enhanced reality' or simply 'image enhancement' (IE), is a technology drawn from its older brother, 'virtual reality' (VR). Unlike VR, which aims to immerse a subject into a 'virtual' computer-generated world, AR aims to enhance a real-world image with virtual objects or subjects. In SN/IGS applications, where a visualization aid such as a surgical microscope or rigid/flexible endoscope is used, the observed 'real' image can be overlaid with a pre-operatively scanned image [usually by computed tomography (CT) or magnetic resonance imaging (MRI)] of the patient; the 'virtual' image, hence image-enhanced SN (IESN), also known as AR-based SN. It should be noted that some surgical microscope vendors do already sell microscopes with two-dimensional (2D) contour-like enhancements of real images. However, a three-dimensional (3D) IESN system aims to provide 3D image enhancement, i.e. not just in the plane of a single image, but across a stack of images, or in other words in 'depth' as well. Research in 3D IESN started in the early to mid-1990s. Edwards *et al*. (1,2), who worked on the Microsope-Assisted Guided Interventions (MAGI) Project, enhanced the view of a stereo surgical microscope using the concept of AR, by 'injecting' pre-operatively scanned and segmented images of the patient into the optics of the microscope. State *et al*. (3) worked on an AR-based system for ultrasound-guided needle biopsies, which originated from an earlier AR-based system for routine obstetric examinations developed by Bajura *et al*. (4). Since then, many other groups have joined the race to commercialize a fully-fledged 3D IESN system, applicable to different fields of surgery, i.e. ear, nose and throat (ENT), neuro-, maxillofacial, laparoscopic and orthopaedic surgery.

IESN for endoscopic sinus surgery (ESS)

Although some groups are already investigating IESN systems which consider deformable anatomy, typically useful in neuro- and laparoscopic surgery, where organs and soft tissues move during surgical intervention, the problem of generating accurate overlays of the virtual and real images has not yet been satisfactorily solved for rigid anatomy (i.e. bone), typically omni-present in ENT, maxillofacial and orthopaedic surgical interventions. We will discuss the overlay issue for non-deformable tissues in the next section. Our research focuses on ENT surgical interventions, and in particular endoscopic sinus surgery (ESS), which is generally applied to treat sinusitis when medication is not effective in eliminating the symptoms. The (rigid) endoscope is typically inserted through the opening of the nasal cavity. Images captured by the endoscope are viewed on a monitor placed near the surgeon. Due to the separation of the handling of the endoscope from the viewing of the image, ESS requires surgeons to have well trained hand–eye coordination. Unlike the use of the stereo surgical microscope in ENT, the endoscope does not provide the stereo cue for depth perception, hence a surgeon can only perceive depth through motion and shading, which may affect the accuracy of tool placement. Whilst the skill and experience of the surgeon are important factors to the success of ESS, the assistance of IESN can further reassure the surgeon's judgement and enhance surgical performance. Shahidi *et al*. (5,6) initially worked in the area of virtual endoscopy and conventional SN for endoscopic surgery. Their work was later extended to use an AR-enabled visualization technique (7). We will briefly discuss their work and further research in the field of IESN for ESS by Bax *et al*. (8) in the next section, after explaining the concepts of calibration and registration.

Calibration, registration and tracking

The main issue of IESN systems is the accuracy by which real and virtual images can be overlaid. This process happens in three phases and the methods/technologies used are calibration, registration and tracking. Calibration aims to align the virtual camera system with the real camera system (in our case, the latter corresponds to the endoscope optics). If we manage to do that, then a virtual object will be seen from the same camera angle and position as the corresponding real object. The calibration procedure is discussed in more depth in the Materials and methods section. Calibration relates a simple object in the real and virtual worlds. Registration relates a patient with his/her virtual counterpart (typically pre-operative CT, MRI or ultrasound images of the patient). We will see that calibration and registration can be combined in one procedure. After calibration/registration, we should arrive at an overlay of a corresponding real and virtual view of the patient (subject to an overlay error), at wherever a static optical device is currently pointing. As soon as the optical device (or possibly the patient) moves, the overlay is destroyed, hence we need to track this motion. Motion-tracking devices typically used in commercial SN systems are either optical or electro-magnetic (EM).

Shahidi *et al*. (7) presented a method for accurate calibration of a rigid zero-degree endoscope using the Tsai calibration algorithm (9). Their method requires the endoscope to be calibrated at different distances to the calibration target. Parameters obtained at several levels are then interpolated, based on the distance between the endoscope and its target. By calibrating the endoscope at different depth levels, the focal length and other

DOI: 10.1002/rcs

intrinsic parameters can be accurately estimated when the endoscope moves. Their work also includes the assessment of the endoscope tracking error distribution within the field of view (FOV) of the endoscope, using the point-based registration error theory. They showed that the tracking error increases as objects are further away from the tip of endoscope, e.g. the error at the tip of the endoscope was 0.67 mm, whilst the highest error occurring inside the FOV was 0.9 mm, which is located at the furthest point of the FOV.

Bax *et al*. (8) focused on the use of a linear technique to compensate for the non-linear radial distortion in the endoscopic image. The linear technique presented in their work involves less computation than in the non-linear radial distortion model. However, as the computer graphics hardware has become more sophisticated, such computations can be transferred to programmable shaders of the render hardware (10). Yamaguchi *et al*. (11) calibrated an oblique-viewing endoscope which could be applied to IESN.

Objective

The main objective of the research presented in this paper was to quantitatively validate the accuracy (in mm) of each step required to produce an accurate overlay in an IESN system with a rigid zero-degree endoscope, typically used for ESS. These steps include calibration, registration and tracking. The experiment was run in a controlled *in vitro* environment and aimed to report accurate measurements on each individual step, in order to analyse its contribution to the overall overlay accuracy and to allow its further optimization to minimize the overlay error.

In Materials and methods we cover the essential visualization tools, as well as the calibration and registration methods used in our IESN system for ESS, called 'ARView'. The next section covers results on calibration, registration, tracking and overlay accuracy in a series of seven experiments.

Materials and methods

3D visualization, Graphical User Interface (GUI) and software architecture

A SN system needs to visualize the patient's anatomy. If we assume that the anatomy of the patient will not change during surgical intervention – a reasonable assumption for ESS – virtual images of the patient can be obtained from pre-operatively acquired CT, MRI or 3D ultrasound scans. Our IESN application called ARView was developed from an earlier C++/OpenGL in-house visualization software named '3DView' (12) (an executable and documentation of the 3DView software for personal use can be downloaded from: http://www2.cmp.uea.ac.uk/~rjal/3DView.htm). The basic rendering mode of this software is texture-based volume rendering. Today's video adapter cards have sufficient memory to store datasets of 512^3 voxels, i.e. equivalent to 128 MB for a voxel of 8-bit = 256 levels of greyscale. This is not an uncommon size, particularly for CT image data. There are two ways to exploit the texture storage and display properties of modern cards for volume rendering: using 2D texture or 3D texture. 2D texture simply adopts the original slice by slice storage structure of tomographic image data, such as CT and MRI. To visualize a 3D image of the data based on this concept, three sets of the same dataset for each viewing plane, i.e. sagittal, coronal and transverse, need to be stored. This is to avoid 'seeing the gaps' in between slices when viewing the data from a direction close to orthogonal to the image slice normal. If this is about to happen (i.e. at a 45° angle), a different dataset is visualized. 3D texture works considerably differently. A stack of planes (quads) or circular shells are rendered, upon which a 3D texture of (medical) image data is projected. When the data texture is rotated, the actual 'support' planes or shells remain static, unlike 2D texture, where each individual 2D slice is projected and 'glued' to its corresponding plane. Therefore, 3D texture does not suffer from the aforementioned problem of gaps in between slices, hence the dataset needs to be only stored once, thus saving two-thirds of memory storage. However, aliasing does occur even when using circular shells. Despite the additional cost of memory, we do prefer the use of 2D textures, which yield better visual quality. The ARView software also enables conversion from volume data to surface data, using the marching cubes algorithm (13) in conjunction with a mesh decimation algorithm (14) to reduce the polygon count. Video images, as captured from an endoscope or surgical microscope, can be displayed in conjunction (overlaid) with calibrated and registered volume- or surface-rendered image data. Figure 1 shows the ARView GUI. Figure 2 shows the workflow of the ARView IESN system in operation, covering each of the steps (calibration, registration and tracking) needed to arrive at a final overlay ready for operational use.

Calibration

Calibration of a camera system involves the derivation of camera parameters, which cannot be directly measured but need to be indirectly derived from one or more captured images used as input to a calibration 'algorithm'. The camera parameters are classified in two distinct classes: extrinsic and intrinsic. Extrinsic parameters define the position and orientation of a camera in either its own 3D coordinate system or a generic 3D 'world' coordinate system. In 3D these are represented by three position parameters, x, y and z, and three orientation parameters, i.e. Euler angles around each of the three axes. Intrinsic parameters are related to the optical system used and include effective focal length, scale factors (normally in x

DOI: 10.1002/rcs

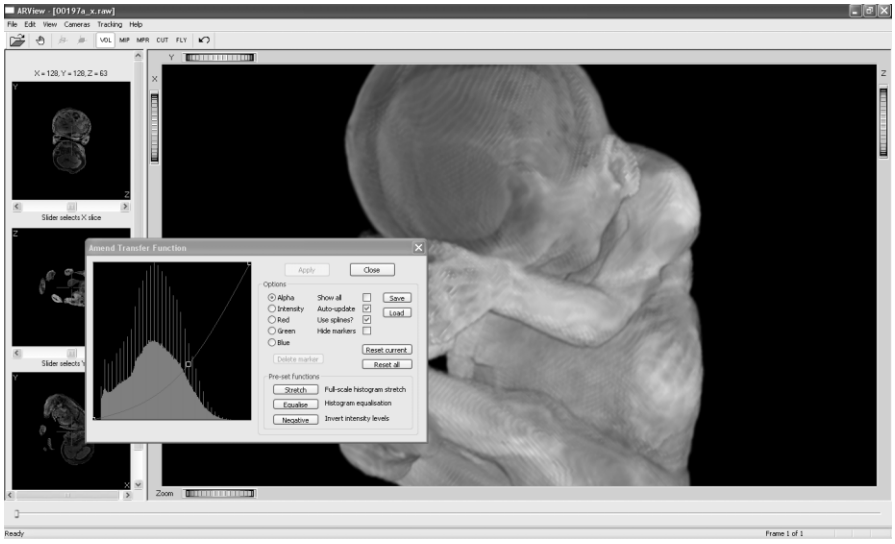

Figure 1. The ARView GUI with 3D volume-rendered fetus (MRI data)

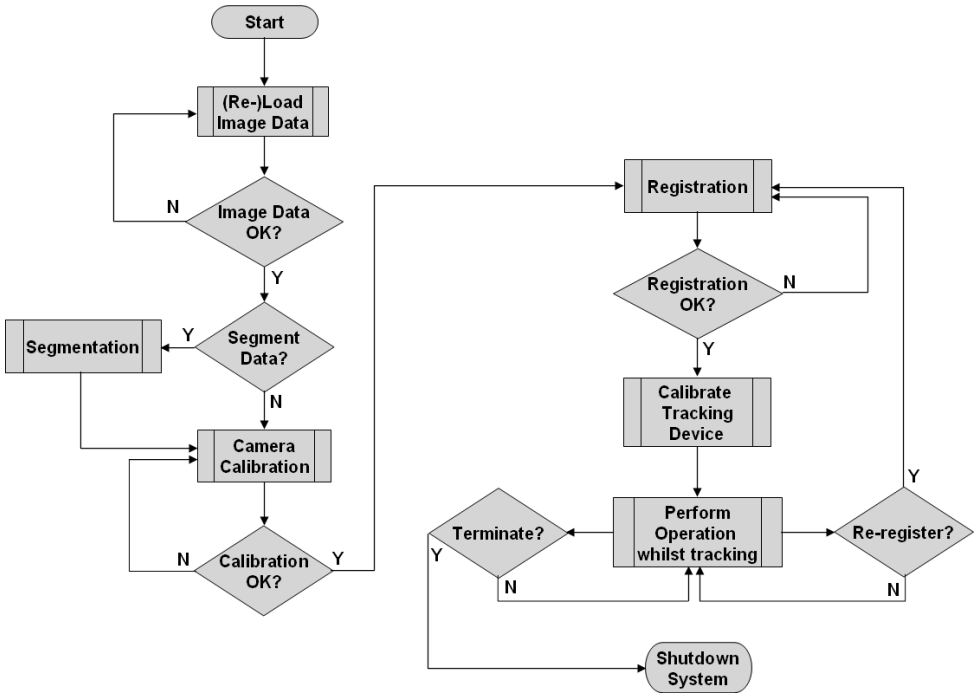

Figure 2. The basic workflow of the ARView IESN system in operation

and y, but can be reduced to x only), the optical centre of the image plane (u and v component) and radial distortion. The latter is negligible in surgical microscope systems but this is not so for endoscopes, which have been mounted with a 'fish-eye' lens to maximize the captured area in close-up viewing. If we adopt one parameter for radial distortion (see further) then 11 unknown parameters need to be recovered.

We adopted the widely used Tsai calibration algorithm (9) in our research for this purpose. This algorithm is well explained in Tsai's original paper (9). We will briefly outline the *radial distortion* model, as this is, as already mentioned, crucial to endoscope calibration. The mathematical model of radial distortion can be described using the following quartic polynomials:

$$x = x_d(1 + \kappa_1 r^2 + \kappa_2 r^4)$$
$$y = y_d(1 + \kappa_1 r^2 + \kappa_2 r^4) \quad (1)$$

In equation (1), (x, y) and (x_d, y_d) respectively represent correct and distorted image coordinates; $r^2 = {x_d}^2 + {y_d}^2$ is the squared Euclidean distance between a distorted image point and the image centre at $(0, 0)$. The coefficients κ_1 and κ_2 ($\kappa_1 > \kappa_2$) are respectively known as second- and fourth-order degree radial distortion factors. Higher-order coefficients (greater than four) are not necessary for practical applications, as their contribution to the distortion model is negligible. Positive coefficients

DOI: 10.1002/rcs

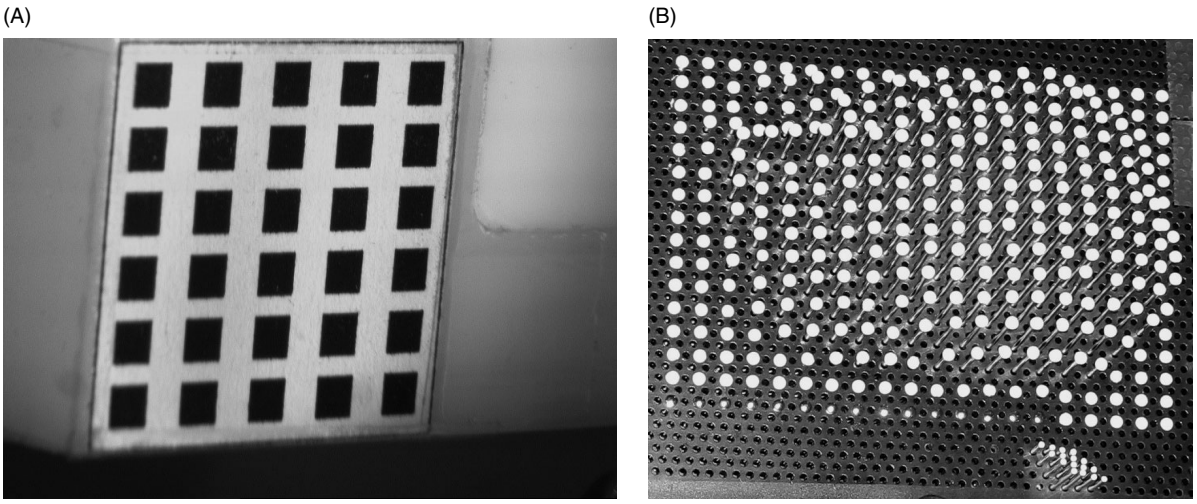

Figure 3. (A) Planar and (B) non-planar calibration patterns. The edge length of the squares on the planar object and the diameter of the dots on the non-planar object are both 2 mm

correspond to 'barrel' distortion (outwards), typically for endoscopes; whilst negative coefficients correspond to 'pincushion' distortion (inwards). In our experiments, we found that even κ_2 is of negligible magnitude, hence κ_1 is sufficient to model the radial distortion of the endoscope. The endoscope calibration procedure uses a calibration 'object' or 'pattern' which can be either planar or non-planar. This corresponds to the two Tsai calibration algorithms for coplanar and non-coplanar calibration, respectively. We have experimented with several calibration objects/patterns throughout our research and we have tested both planar (Figure 3A) and non-planar (Figure 3B) versions – see experiments 1 and 2 in the Results section.

The endoscope calibration procedure consists of the following four stages (Figure 4):

1. Acquisition of the calibration image (endoscope capture and Matrox Framegrabber).
2. Calibration image smoothing and segmentation (Canny edge detection + shape-based removal of noise).
3. Calibration marker localization.
4. Computation of camera parameters – overlay after radial un-distortion of video image.

Registration

In our method, camera calibration is a part of the registration process as the parameters found through the calibration procedure align the virtual and real calibration patterns (and correspondingly align the virtual camera system with the real camera system). However, aligning two viewpoints only partially achieves the registration for an IESN system, as it does not align the virtual patient (i.e. the image data) with the real counterpart (the endoscope image). Figure 5 illustrates the *pre-operative registration* process schematically.

To apply this procedure to the endoscope image (real subject) and image data (virtual subject), we mount the calibration object to the subject using a mouthpiece, more specifically the VBH mouthpiece (15) (Figure 6A). The whole assembly (Figure 6B), i.e. mouthpiece and calibration pattern, also contains a passive marker set-up for tracking (the headband, which can be seen at the top-centre in Figure 6B, C, was scanned to test the EM tracking device of a GE SN system. These experiments are not reported in this paper).

The subject (a laboratory skull) is pre-operatively scanned with the fixed mouthpiece and the mounted calibration object (see Figure 6B). The resulting scanned calibration object is segmented from the subject's image data (Figure 6C), using a watershed segmentation algorithm (16). This is the square object at the bottom right in Figure 6C. If we register this segmented calibration object with the pre-defined virtual calibration object, i.e. the object used to align the real and virtual cameras during calibration, then we have achieved a real-to-virtual subject registration. The iterative closest point (ICP) algorithm (17) is used for this purpose. The reasons for adopting the ICP algorithm are:

1. The algorithm works with a wide range of surface representations.
2. No prior knowledge of the correspondence between two surfaces is required.
3. The convergence rate is fast if the source surface represents a large portion of the reference surface.
4. At each iteration, an improved transformation can be obtained from a closed-form solution.

As our surfaces to be registered are inherently the same, the translational transformation is trivial as we simply align their centres. This leaves the rotation to be solved, which is typically represented by a quaternion rather than a 3×3 rotation matrix. The former has the advantage that interpolation can be done correctly between two quaternions, which is not the case for rotation matrices.

DOI: 10.1002/rcs

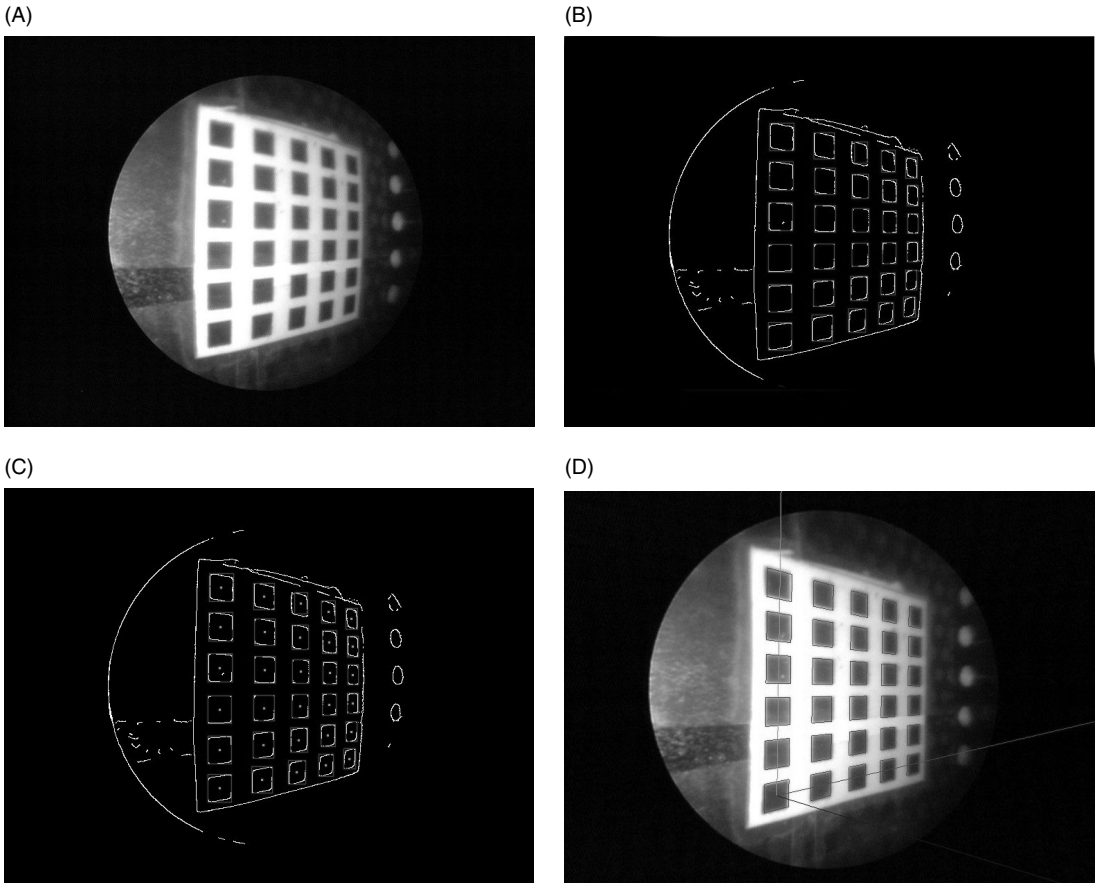

Figure 4. (A) Video image of planar calibration pattern as seen through endoscope; (B) after smoothing and segmentation; (C) localization of image coordinates (u, v); (D) overlay after un-distorting the original image

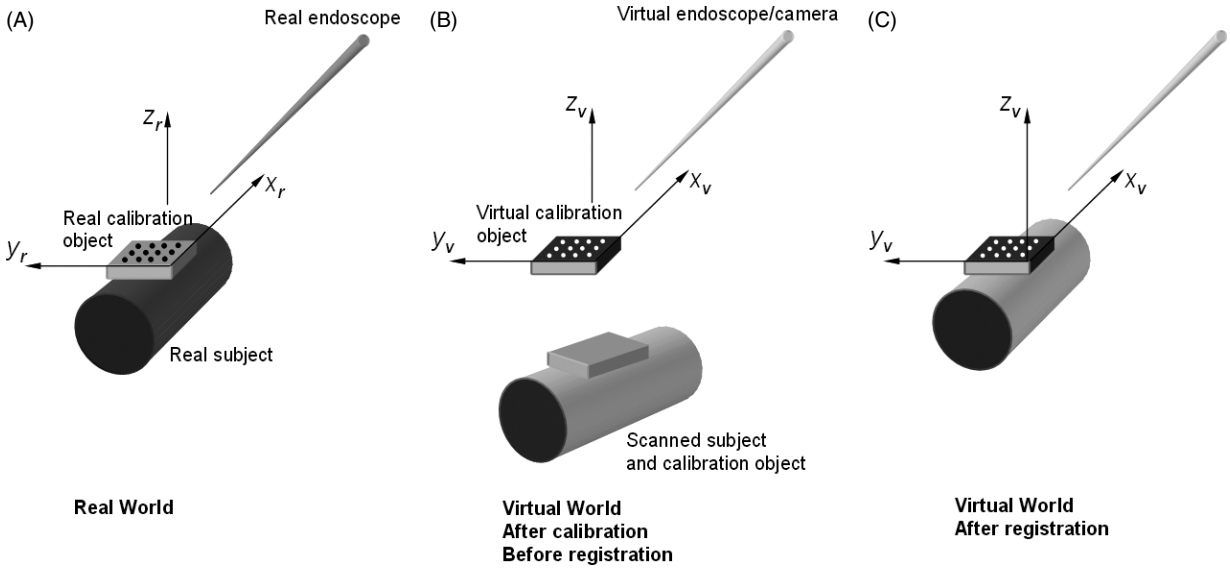

Figure 5. The real endoscope is calibrated on the real calibration object (A). This brings the virtual endoscope (this is merely a virtual camera) in line with its real counterpart relative to the pre-defined virtual calibration object, which also contains the pattern of markers (B). The subject is scanned with the calibration object. The scanned calibration object is segmented and matched with the pre-defined virtual calibration object yielding scanned to real subject registration (C)

We refer the interested reader to Van Verth and Bishop (18) for a treatise on the arithmetic of quaternions.

So far, we have only covered a pre-operative registration procedure. However, when the endoscope starts moving, the tracking device (see next section) will increase the overlay error over time, due to its limited accuracy. This will eventually require re-registration, depending on the length of the surgical procedure. We

DOI: 10.1002/rcs

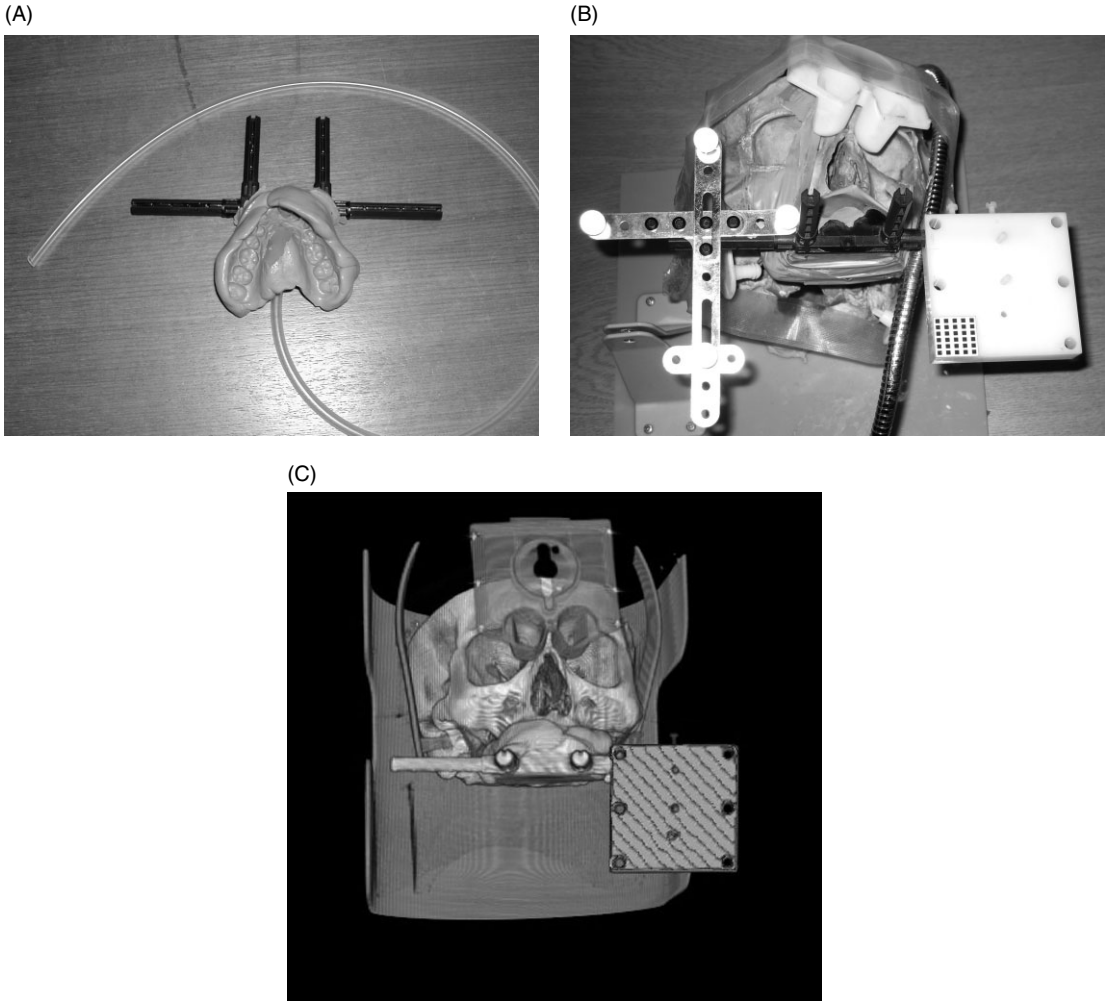

Figure 6. (A) The VBH mouthpiece. The tube is connected to a small pipe going straight through the 'gumshield'; hence, when connected to a vacuum pump, the air is sucked away between the gumshield and the subject's palate, resulting in a solid fixation of the mouthpiece to the subject. (B) The calibration block and pattern plus markers in an *ad hoc* assembly as mounted on the mouthpiece (the latter being attached to the palate of a skull). Note that despite the relatively large dimension, the assembly is lightweight (<60 g). (C) The CT scanned skull with mounted calibration block after segmentation (bottom right)

tested a method proposed by Clarkson *et al*. (19), using photo-consistency as a potential candidate for *intra-operative registration*. To use photo-consistency as the cost function for solving our registration problem, we assume that the illumination of an object is independent from the viewing direction – in other words a Lambertian model – for which the image of a visible surface point should exhibit similar intensity across different images when the lighting remains static. The registration process can be expressed using the following equation:

$$\mathbf{v}'_{n,i} = \mathbf{P}\mathbf{M}_n\mathbf{v}_i \tag{2}$$

where $\mathbf{v}'_{n,i}$, is the projection of the ith rendered pixel in the nth image; $\mathbf{P}$ is the projection matrix that is constant for all images; and $\mathbf{M}_n$ is the pose of the camera for the nth image. The corresponding pixel intensity of voxel $\mathbf{v}_i$ in each image is given by $I_1\ (\mathbf{v}'_{1,i})$ and $I_2\ (\mathbf{v}'_{2,i})$; L is the total number of visible voxel pairs in both images. Hence, the photo-consistency metric between two images ($n = 2$) (in the context of this research, the two images would be an endoscope and pre-operative scanned image, e.g. CT/MRI) is defined by:

$$C_{photo} = \frac{1}{L}\sum_{i=1}^{L} ||I_1\ (\mathbf{v}'_{1,i}) - I_2\ (\mathbf{v}'_{2,i})|| \tag{3}$$

Therefore, photo-consistency can measure the intensity difference between the images of a voxel that appears in two or more endoscopic images. The images are said to be consistent when the difference is small. If a model is correctly registered with two or more images, the photo-consistency value should be minimal and ideally zero.

Results on the accuracy of the presented registration methods are covered in Results.

Tracking

We used the Northern Digital (NDI) Hybrid Polaris optical tracking system with passive markers. When using passive

DOI: 10.1002/rcs

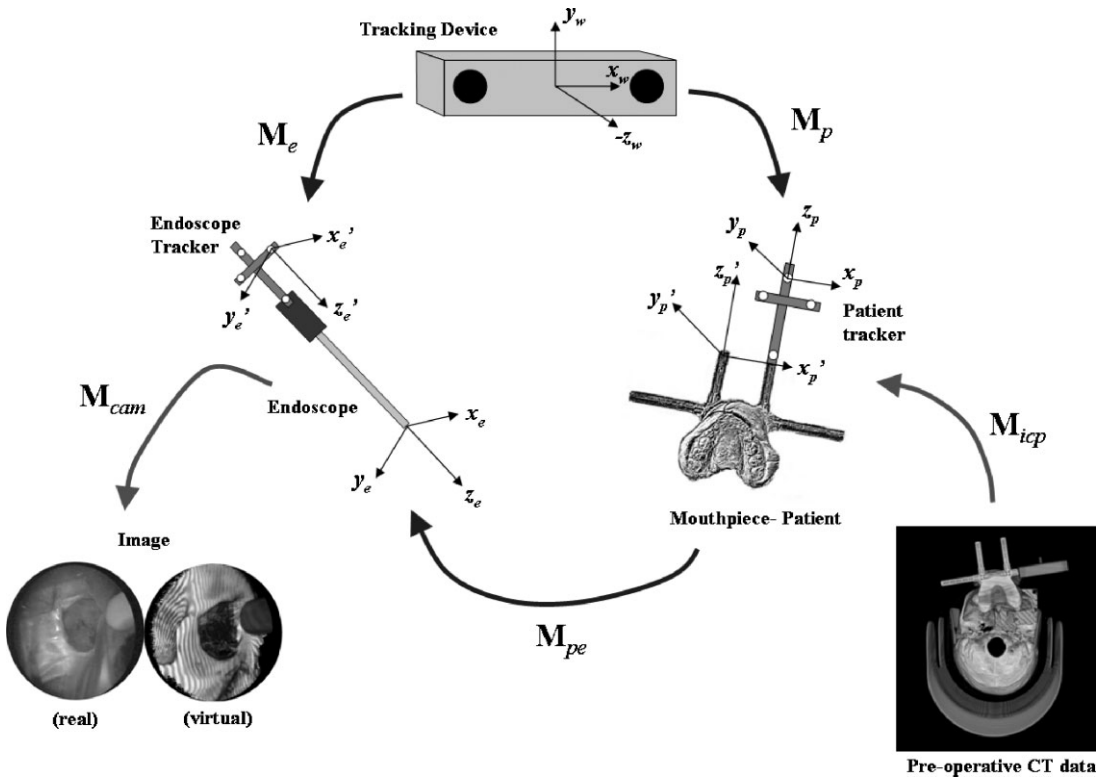

Figure 7. Coordinate systems of tracking device camera (world), endoscope and patient. The primed coordinate systems have the same orientation as the non-primed systems (related to tracking assemblies) and differ in a known translation (which is omitted in this figure, and the equations in the text, for simplicity). Transformations from one system to the other are explained in the text

markers, the Polaris identifies different tracked objects through the use of different marker layouts. At least three markers are needed; however, an additional fourth marker is often used to increase reliability, for example when another marker has been occluded. The layout is subjected to the following requirements:

1. The minimum segment length is 50 mm.
2. Each segment length has to differ from other segments by at least 5 mm.
3. Each angle between two segments must differ from other pairs by at least 0.5°.

An example of such a passive marker combo is shown on the bottom left in Figure 5B.

Figure 7 shows an overview of the coordinate systems (a) and corresponding transformations (b) needed to arrive from endoscope captured image to overlaid image, including the tracking device transformations. This process can be described in the following three transformation steps:

1. *Patient to endoscope – initial*:

$$\mathbf{M}_{pe,0} = \mathbf{M}_{e,0}\mathbf{M}_{p,0}{}^{-1} \tag{4}$$

2. *Patient to endoscope – in motion* (the pre-multiplication with the inverse of the initial patient-to-endoscope matrix is to set the system to coordinates (0, 0, 0) before the onset of motion):

$$\mathbf{M}_{pe,i} = \mathbf{M}_{pe,0}{}^{-1}(\mathbf{M}_{e,i}\mathbf{M}_{p,i}{}^{-1}) \tag{5}$$

3. *Image overlay*:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{M}_{cam}\mathbf{M}_{pe,0}{}^{-1}\mathbf{M}_{e,i}\mathbf{M}_{p,i}{}^{-1}\mathbf{M}_{icp}\begin{bmatrix} x_d \\ y_d \\ z_d \end{bmatrix} \tag{6}$$

$$\text{where } \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \mathbf{M}_{icp}\begin{bmatrix} x_d \\ y_d \\ z_d \end{bmatrix}$$

where the index p refers to the patient coordinate system and the index d to the image data; $\mathbf{M}_{icp}$ is the transformation matrix to transform from image data to world coordinates (registration); and $\mathbf{M}_{cam}$ is the projective transformation matrix to transform from camera to image (calibration).

Results

This section will cover experimental set-ups and results of a number of experiments run on each of our calibration, registration and tracking methods, as presented in the previous section. Finally we present an overall accuracy on the 'end-product', i.e. the overlay of the real (endoscope) and virtual (CT data) image. The purpose of the experiments presented is to pinpoint a number of important issues which can greatly affect the accuracy of each of these procedures and, ultimately, the final result.

DOI: 10.1002/rcs

The rigid endoscope used in our experiments is a STORZ TriCam with a 110 mm shaft.

Calibration

Experiment 1

The purpose of the first calibration experiment was to examine the influence of the angular position of the planar calibration target/pattern (see Figure 3A) to the accuracy of the camera parameters estimation, and in particular the influence on correlated parameters, such as the effective focal length f and the depth position of the camera t_z.

Indeed, scaling these two parameters with the same factor will result in an identical image when the image plane is perpendicular to the optical axis. However, if t_z (which represents the distance between the camera and the calibration target) is scaled, hence deviating from its correct value, then the overlay accuracy will be poor when the endoscope starts moving after initial calibration. Figure 8 shows the experimental set-up schematically. Table 1 shows the absolute error on the depth increment d_z after calibration when the endoscope is moved 5 mm upwards and downwards, respectively, from its initial position.

It should be noted that angles >45° are impractical for calibration with our current endoscope, as in this position not all calibration markers can be brought into focus. This is illustrated in Table 1, where even for a 45° angle only a positive shift of 2 mm could be achieved without loss of focus, as compared to the 5 mm shift in both directions for smaller angles.

Table 1. The absolute error in computed depth displacement d_z: the physical d_z indicates the physical displacement measured by the height gauge; the computed d_z is the displacement value estimated by the calibration routine; the absolute error shows the difference between the physical displacement and the estimated displacement

Angle (°)	Physical *dz* (mm)	Computed *dz* (mm)	Absolute error (mm)
0	5.00	35.13	30.13
	−5.00	−34.96	29.96
25	5.00	6.47	1.47
	−5.00	−6.50	1.50
30	5.00	5.72	0.72
	−5.00	5.58	0.58
40	5.00	5.35	0.35
	−5.00	−4.64	0.36
45	5.00	5.14	0.14
	7.00	7.27	0.27

Experiment 2

In this experiment, we compared the planar (Figure 3A) and non-planar patterns (Figure 3B). From the previous experiment, we can already conclude that the accuracy of the parameters after calibration increases with increasing positioning angle. We further discuss this in the appropriate section, but will adopt nonetheless the most favourable angular positions for the planar calibration object (40° and 45°).

The non-planar pattern is a modified pin-art set-up, where pins of a diameter of 1 mm (head is 2 mm) can be positioned at various depths by pressing a shape underneath. This allows us to make arbitrary calibration patterns. For this purpose, a half skull model of a fetus was used as the shape model because it is mostly smoothly curved but shows some irregular curvature near the zygomatic bone.

Table 2 shows calibration errors for the non-coplanar pattern over different inclinations, one calibration for each of 0°, 15°, 30°, 40° and 45°. We had already derived from experiment 1 that the planar pattern performed best at the highest possible inclination of 40–45°. Table 3 shows best and worst results over five trials for each of these inclinations.

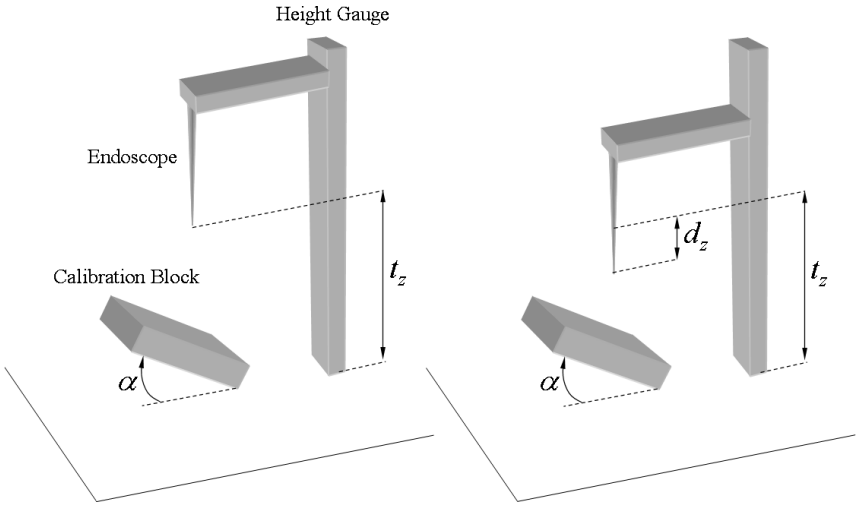

Figure 8. Schematic representation of a laboratory experiment to test the influence of the placement angle α of a planar calibration pattern on the accuracy of the camera parameters t_z, d_z and f

DOI: 10.1002/rcs

Table 2. Calibration average errors (SD in parentheses) for a non-coplanar target under different inclinations

Inclination (°)	Distorted image plane error (pixels)	Undistorted image plane error (pixels)	Object space error (mm)
0	1.2 (0.8)	1.3 (0.9)	0.12 (0.08)
15	1.2 (0.9)	1.2 (1.0)	0.12 (0.09)
30	1.3 (0.8)	1.4 (0.8)	0.14 (0.08)
40	2.5 (1.6)	2.7 (1.7)	0.25 (0.16)
45	2.8 (1.9)	3.0 (2.0)	0.27 (0.18)

Table 3. Worst and best calibration average errors (SD in parentheses) for five trials for a coplanar target under 40° and 45° inclinations

Inclination (°)	Distorted image plane error (pixels)	Undistorted image plane error (pixels)	Object space error (mm)
40 (worst)	0.38 (0.19)	0.40 (0.20)	0.025 (0.013)
40 (best)	0.31 (0.17)	0.32 (0.18)	0.020 (0.012)
45 (worst)	0.53 (0.40)	0.57 (0.43)	0.031 (0.024)
45 (best)	0.38 (0.14)	0.40 (0.14)	0.022 (0.009)

Registration

For these experiments, we used a real skull and its CT-scanned version. The skull was scanned with the VBH mouthpiece upon which the calibration block and pattern were mounted – see Figure 6B and C for the real and virtual skull, respectively.

Experiment 3

As already mentioned, we used the calibration blocks, i.e. the predefined and scanned virtual versions, for registration. We could easily assess the accuracy of the registration by checking the decreasing/converging sum squared error (SSE) on the block matching. However, our essential interest was the SSE in the target area or the region of interest (ROI), where the actual surgical intervention happens. This error is known as the target registration error (TRE), as suggested by Fitzpatrick *et al.* (20). To do this, we first used the ARView software to segment and triangulate a ROI for ESS, i.e. the frontal and maxillofacial area of the skull. Then we did a dry run of registrations, picked the best version with the smallest SSE of the block, and further minimized the error by manual adjustment until it was below 0.1 mm. Then we applied the resulting transformation on the ROI/target surface and considered this as the target surface 'ground truth'. Then we performed another five registrations from scratch and recorded the calibration block SSE and the TRE. Table 4 shows the result of these trials in columns 1 and 2, respectively. As we expected, the TRE for the ROI was higher than the SSE for the calibration block. As we realized that the difference between these errors depends on the distance of the calibration block from the ROI, we repeated the experiment on an area closer to the block (ideally, we should have placed the calibration block closer to the ROI; however, this would have required redesigning the mouthpiece and scanning the skull all over again, which is a non-trivial operation. It was therefore decided to simply pick an area on the skull closer to the block, which is trivial to do and proves the same point). That result is shown in column 3 of Table 4 and shows again what we expected; that the TRE for this area is lower than for the actual ROI. This implies that the calibration block should be as close as possible to the ROI to minimize the TRE. We elaborate further on this in the Discussion.

Table 4. Sum squared errors and target registration errors for experiment 1 over five trials (SD in parentheses)

SSE for block (mm)	TRE for ROI (mm)	TRE for closer area (mm)
0.48 (0.04)	2.31 (0.4)	1.30 (0.1)

Experiment 4

The intra-operative photo-consistency registration technique was successfully tested by Chen on simulated images of a complete skull (21). Here we present the same technique as applied to endoscopic images where the endoscope is in close-up, targeted at the orbit of the skull. The real and virtual skull models were first aligned using the pre-operative registration method. Unlike the experiment as reported in (21), we did not manage to improve the registration using the photo-consistency technique. Therefore we investigated the photo-consistency metric in each dimension. Figure 9 shows two example plots for one DOF, i.e. rotation around the x axis, $\mathbf{R}x$, for the experiment presented in (21) and the current one. It is obvious that the first plot shows a clear global minimum in the metric, whilst the second one is cluttered with several local minima and no obvious global minimum. The same pattern emerges for the other five DOF's. This implies that the photo-consistency metric is not really suitable for the close-up endoscopic images, which show little variation in intensity as compared to the synthetic complete skull images from (21).

Tracking

Experiment 5

This experiment tested the accuracy of the translational components reported by the tracking device (the fluctuation of the tracking signal, for each DOF under static conditions, was smoothed out using a Kalman/FIR filter; the delay introduced in response to motion is acceptable, as a surgeon does not move tools when inspecting a ROI). We simply attached the passive tracking tool to a support and moved it around on a measurement bench of 400×400 mm, which allows accurate positioning to 0.05 mm accuracy. We set the three translational DOFs to zero when initializing and recording the first position. Then we moved the tracked object around the grid and recorded the positional coordinates and the coordinates reported by the tracking

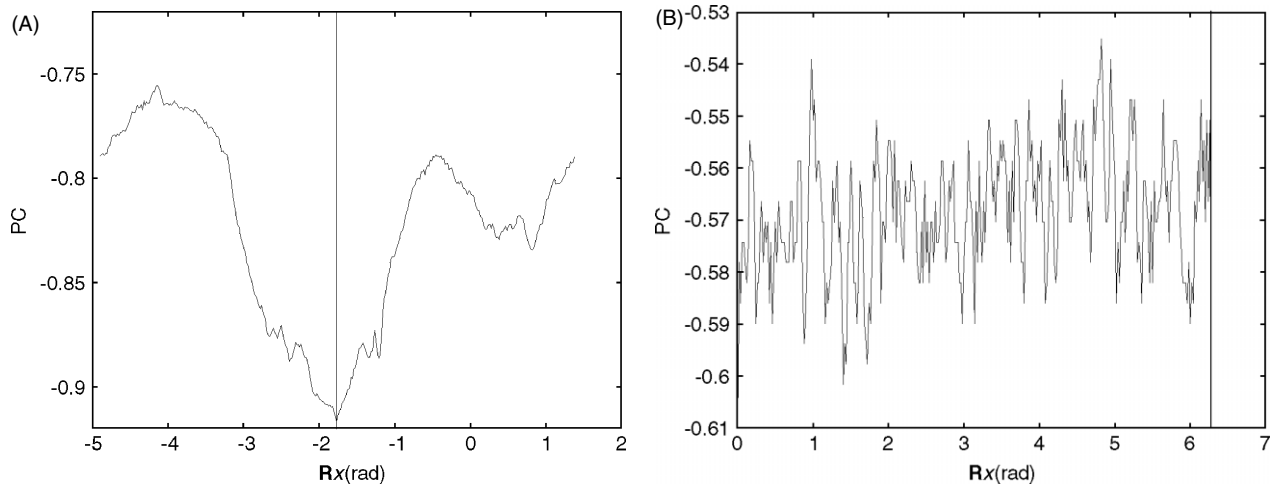

Figure 9. Plots show photo-consistency (PC) vs. rotation position around the *x* axis (R*x*) between two images, each acquired from two objects to be registered. (A) Result for simulated images of an entire skull; (B) Result of close-up endoscopic images of the orbit. The vertical line shows the global minimum

device. Of course, one cannot directly compare the tracked coordinates with the grid positions, as it is practically impossible to position the cameras so that the tracked z coordinate is zero across the grid. Therefore, we fitted a plane through the data points reported by the tracking device, i.e. five points for each of five positions on the grid, which makes 25 points in total (note that we used eigenvalue decomposition, where the eigenvector with the smallest eigenvalue represents the normal to the plane which goes through the centroid of the data points). Then we calculated the average distance to the fitted plane across all data points as a measure of tracking accuracy. We found that, for the optimized position of the camera to capture the entire tracking volume (195 cm from the centre of the grid plate), this error was 1.2 mm.

Overlay

In the final series of experiments we tested the end result, i.e. the overlay of virtual and real, of which the accuracy is of course dependent on the accuracy of each of the previously tested methods.

Experiment 6

The purpose of this experiment was to assess the in-plane overlay error at different depths, hence giving us an idea of the overlay error in 3D (unlike commonly reported in-plane errors whilst ignoring depth). In this experiment we used the same pins set-up used for the non-coplanar calibration pattern. The base plate supporting the pins comprised the x–y plane, whilst the extent by which the pins stuck out of the plate was the z component. The x–y grid was accurately defined to ~0.5 mm, whilst the z component was measured with a height gauge with accuracy 0.01 mm. This allowed us to reconstruct a virtual equivalent of the pins configuration. We then performed the calibration (planar pattern at 40°) and registration procedures as mentioned and tested before, whilst moving the endoscope up and down ±10 mm in each of x, y and z. Table 5 shows the average overlay errors for each x, y position at different depths (note that $z = 0$ is the initial calibration/registration position when no tracking has yet occurred). Figure 10A, B shows the real and virtual configurations after moving the endoscope to an extreme point with coordinates 10, 10, 10 (mm).

Table 5. Average errors (SD in parentheses) of overlay across a volume of 10 × 10 × 10 mm for different tracked positions in *z* (*z* = 0 is initial position)

z Position	Image space error (pixels)	Object space error (mm)
−10	5.61 (0.32)	0.405 (0.013)
−5	4.20 (0.15)	0.392 (0.017)
0	2.49 (0.09)	0.357 (0.007)
5	3.55 (0.24)	0.401 (0.012)
10	4.27 (0.20)	0.409 (0.010)

Experiment 7

Finally, we assessed the overlay on the ROI of a real human skull. The virtual skull was obtained from a CT scan at a resolution of 256 × 256 × 209 with an in-planar voxel size of 1.2 mm and slice thickness of 1 mm. It is non-trivial to exactly quantify this error, but simply looking at pixel discrepancies between corresponding landmarks gave us some idea of what the expected overlay errors were. Figure 11 shows the real and virtual version of the orbit of the laboratory skull. Worst-case errors in x–y near the orbit were in the range 2–3 mm; however, this reduced significantly when measuring closer to the calibration block.

Discussion

The calibration results showed us two interesting findings. We started using the coplanar pattern, which was also

DOI: 10.1002/rcs

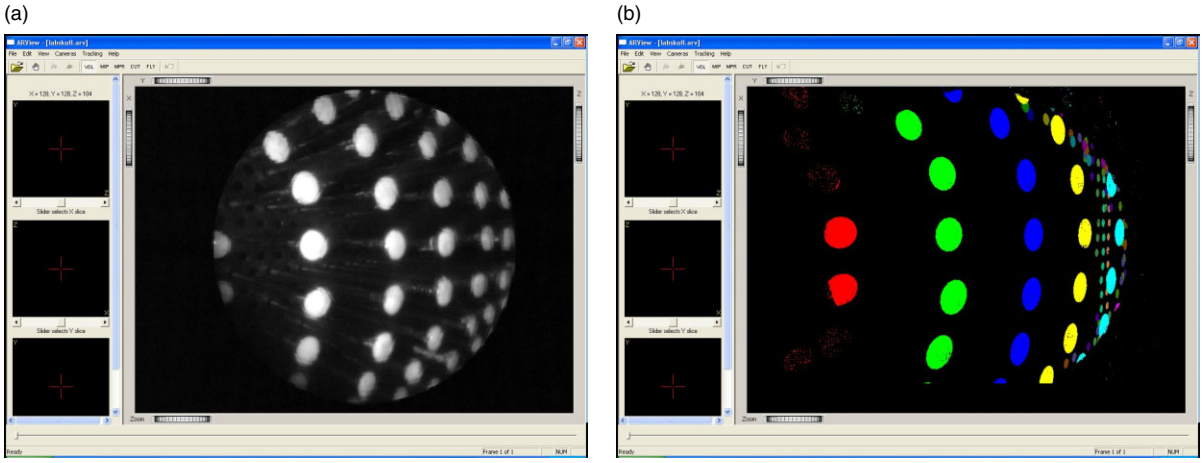

Figure 10. Real (A) and virtual (B) endoscope images, respectively, of pins grid after calibration, registration and tracking a moving endoscope

used in (6,7) for the obvious reason that, because it is flat, it is easy to integrate into any support attached to the patient – in our case the VBH mouthpiece (15). The first experiment proved that the coplanar pattern has to be inclined at an angle of about 40° to yield accurate results. Tsai (9) had already reported that a coplanar pattern needed to be inclined for optimal results. Here we showed how much this inclination should be for the purpose of our IESN system. Table 3 shows little distinction between best errors (not statistically significant) for inclinations of 40° and 45°, respectively. However, the worst-case error was significantly higher for the latter. This is because at 45° it becomes difficult to identify the u, v coordinates accurately, as the markers on the pattern became distorted due to perspective projection, further aggravated by radial distortion. Table 2 shows that the non-coplanar pattern should not be inclined, as the best performance is found at 0° of inclination. The same reason as before is valid, because the non-coplanar pattern already has different levels of depth for the markers. Thus, placing it under an angle influenced the accuracy by which u, v coordinates could be extracted, due to distortion (perspective and radial). Currently, the non-coplanar pattern does not perform as well as the planar pattern. It should be noted, however, that the non-coplanar pattern had inferior accuracy (ca. 0.5 mm) as compared to the planar pattern (<0.05 mm). A future objective is to manufacture a non-coplanar pattern to the same standard of accuracy as the planar version and redo the experiments.

Pre-operative registration results showed clearly how important it is to place the calibration block as closely as possible to the ROI. In the present study we did not have the resources left to change that configuration, but did prove the concept by simply selecting a region closer to the calibration block to compare TREs.

Intra-operative registration results showed potential for the photo-consistency method, introduced in (19) and applied by us for registration in IESN systems. At this stage, we could only improve pre-registration results on images with a sufficient range in intensity. When using close-up images, as typically seen through an endoscope, the method failed to converge, due to too many local minima in the objective function. Future work will aim to address this by using the calibration block for this purpose as well: if the block is closer to the ROI and is partially shaped and manufactured to create a wide range of intensities under endoscope lighting, then this method may well prove successful for intra-operative registration in IESN systems.

Translational tracking errors were found to be slightly >1 mm across an area of 40×40 cm. We did not assess rotational tracking accuracy, due to the absence of accurate measuring equipment.

The overlay experiments gave us an overall indication of what accuracy our IESN system for ESS currently can guarantee. The first experiment on the pin-based non-coplanar pattern showed us errors not exceeding 0.5 mm in x, y. However, this was only measured in a $2 \times 2 \times 2$ cm volume, which is significantly smaller than the volume which would typically be roamed (approximately $10 \times 10 \times 10$ cm). The reason why we could not test larger volumes is down to the equipment set-up, i.e. the height gauge in combination with the mounted endoscope. An interesting finding in experiment 6 was that the error slightly increased with increasing z. This gave us an idea of depth (z) accuracy in the overlay. The final overlays on the laboratory skull after calibration, registration and tracking showed what we expected – worst-case errors in the ROI of 2–3 mm in x, y. Recalling the results of experiment 3, where registration errors were ca. 2 mm in the ROI, together with knowing that calibration errors were as low as 0.025 mm when the inclined coplanar pattern was used, it is easy to derive that registration is still the bottleneck of improving accuracy in our IESN system. Add to this a series of tracked motions (ca. 1 mm accuracy), and it is obvious that worst-case errors run into several millimetres.

In summary, we have performed a number of quantitative experiments on the accuracy of calibration, registration, tracking and final overlay for an IESN system

DOI: 10.1002/rcs

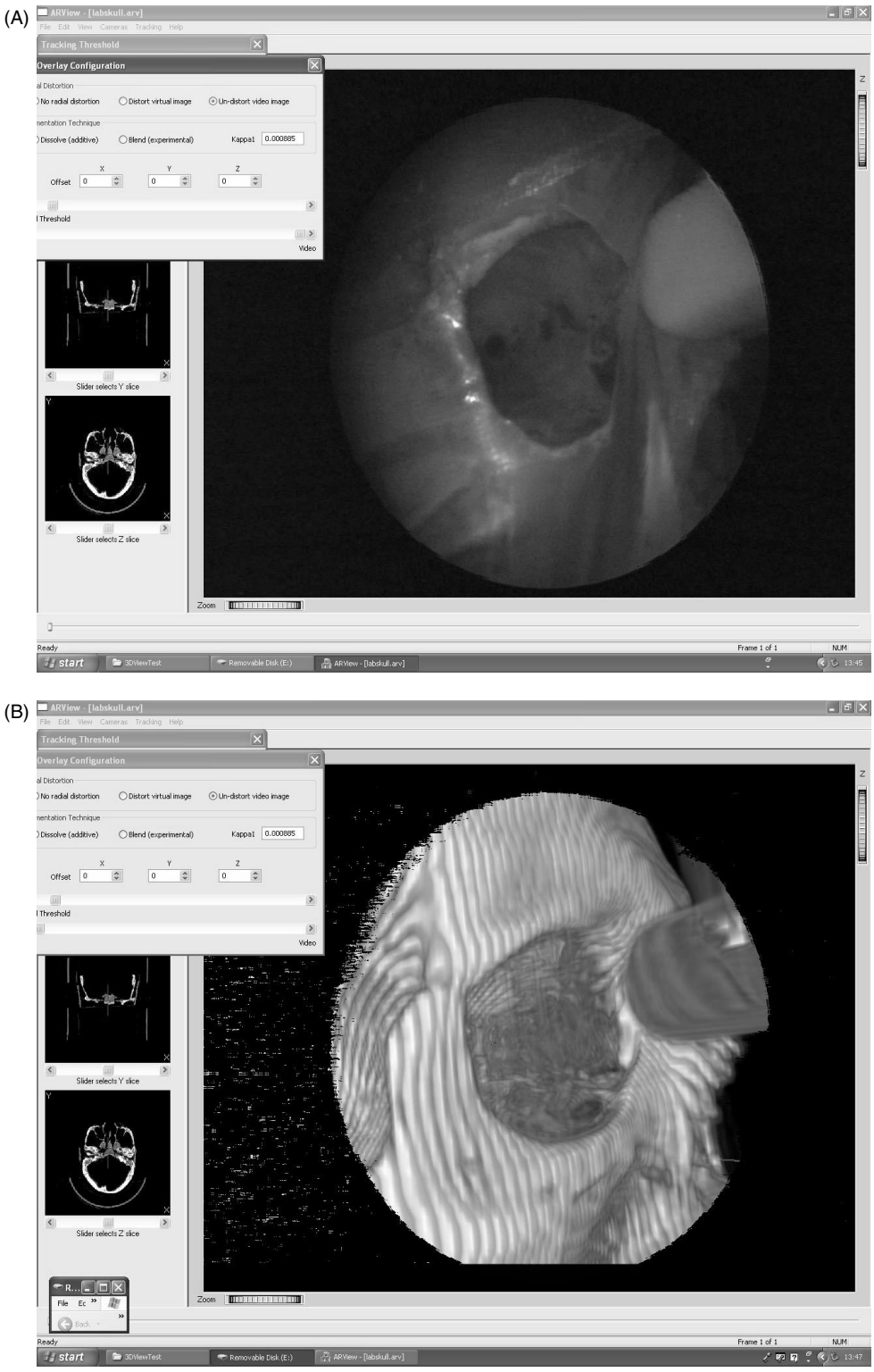

Figure 11. Real (A) and virtual (B) images of the skull, showing the orbit, after calibration, registration and tracking a moving endoscope. Note that a slider on the pop-up window (top left) allows to fade from real to virtual in ARView

for ESS. We found results which were in line with the expected accuracy of current (IE)SN systems. In the future we wish to continue this trend of quantitatively assessing an IESN system before introducing it into the operating theatre – which is the next stage of our research. To improve quantitative assessment, we are aiming to construct an accurate measurement bench which also allows us to measure rotational DOFs, something we omitted at this stage for the registration and overlay accuracy measurements, due to current unavailability of an accurate angle measurement system. The method which is most in need of improvement is registration, hence we wish to investigate further the application of the proposed photo-consistency method. However, due

DOI: 10.1002/rcs

to limited accuracy of current tracking systems and the accuracy of standard CT and MRI scans (0.5–1 mm), the much craved for sub-millimetric accuracy may remain elusive at present.

Acknowledgements

We wish to thank the EPSRC (GR/R86027/01) for sponsoring this project and Mr Ben Pattinson from Storz, UK, for lending us the endoscope equipment.

References

1. Edwards P, Hawkes D, Hill D, *et al*. Augmentation of reality in the stereo operating microscope for otolaryngology and neurosurgical guidance. *J Comput Assist Surg* 1995; **1**(3): 172–178.
2. Edwards PJ, King AP, Maurer CR, *et al*. Design and evaluation of a system for microscope-assisted guided interventions (MAGI). *IEEE Trans Med Imag* 2000; **19**(11): 1082–1093.
3. State A, Livingston MA, Hirota G, *et al*. Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. *Proc SIGGRAPH* 1996; 439–446.
4. Bajura M, Fuchs H, Ohbuchi R. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *Comput Graphics* 1992; **26**(2): 203–209.
5. Shahidi R, Argiro V, Napel S, *et al*. Assessment of several virtual endoscopy techniques using computed tomography and perspective volume rendering. *Visualization in Biomedical Computing (VBC'96) LNCS 1131*, 1996; 521–528.
6. Shahidi R, Wang B, Epitaux M, *et al*. Volumetric image navigation through stereotactic endoscopes. Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI'98, LNCS, 1496), Conference, 1998; 241–252.
7. Shahidi R, Bax MR, Maurer CR. Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. *IEEE Trans Med Imag* 2002; **21**(12): 1524–1535.
8. Bax MR, Khadem R, Johnson JA, *et al*. Image-enhanced endoscopy calibration and image registration accuracy testing. *Proc Med Imag Visualiz Imag Guided Proc Display* 2002; **4681**(1): 119–126.
9. Tsai RY. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Trans Robotics Automat* 1987; **RA-3**(4): 323–344.
10. Chen MS, Lapeer RJ, Rowland RS. Real-time rendering of radially distorted virtual scenes for endoscopic image augmentation. *Stud Health Technol Informat* 2005; **111**: 87–89.
11. Yamaguchi T, Nakamoto M, Sato Y, *et al*. Camera model and calibration procedure for oblique-viewing endoscope. Proceedings of Medical Image Computing and Computer Assisted Intervention (MediVi2/IV'04), London , 2004; 62–72.
12. Lapeer RJ, Rowland R, Chen MS. PC-based volume rendering for medical visualization and augmented reality based surgical navigation. Proceedings of (MICCAI'02), LNCS 2488, 2002; 596–603.
13. Lorensen WE. Marching cubes: a high resolution 3D surface construction algorithm. *ACM Comput Graphics* 1987; **21**(4): 163–169.
14. Garland M, Heckbert P. Surface simplification using quadric error metrics. In *Proc SIGGRAPH* 1997; 209–216.
15. Martin A, Bale RJ, Vogele M, *et al*. Vogele–Bale–Hohner mouthpiece: registration device for frameless stereotactic surgery. *Radiology* 1998; **208**: 261–265.
16. Lapeer RJ, Tan A, Aldridge R. Active watersheds: Combining 3D watershed segmentation and active contours to extract abdominal organs from MR images. Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI) Conference, 2002; 596–603.
17. Besl PJ, Mckay ND. A method for registration of 3D shapes. *IEEE Trans Pattern Anal Machine Intell* 1992; **14**(2): 239–256.
18. Van Verth J, Bishop L. *Essential Mathematics for Games and Interactive Applications*. Elsevier: Amsterdam, 2004.
19. Clarkson MJ, Rueckert D, Hill DL, Hawkes DJ. Using photoconsistency to register 2D optical images of the human face to a 3D surface model. *IEEE Trans Pattern Anal Machine Intell* 2001; **23**(11): 1266–1280.
20. Fitzpatrick J, West J, Maurer C. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imag* 1998; **17**(5): 694–702.
21. Chen MS. Calibration and Registration of an Image Enhanced Surgical Navigation System for Endoscopic Sinus Surgery. PhD Thesis, School of Computing Sciences, University of East Anglia, UK, April 2007.