PII S0301-5629(98)00044-1

● *Original Contribution*

RAPID CALIBRATION FOR 3-D FREEHAND ULTRASOUND

R. W. PRAGER,[†] R. N. ROHLING,[†] A. H. GEE[†] and L. BERMAN[‡]

[†]Department of Engineering, University of Cambridge, Cambridge, UK, and [‡]Department of Radiology,
Addenbrooke's Hospital, Cambridge, UK

Abstract—**3-D freehand ultrasound is a new imaging technique that is rapidly finding clinical applications. A position-sensing device is attached to a conventional ultrasound probe so that, as B-scans are acquired, they can be labelled with their relative positions and orientations. This allows a 3-D data set to be constructed from the B-scans. A key requirement of all freehand imaging systems is calibration; that is, determining the position and orientation of the B-scan with respect to the position sensor. This is typically a lengthy and tedious process that may need repeating every time a sensor is mounted on a probe. This paper describes a new calibration technique that takes only a few minutes to perform and produces results that compare favourably (in terms of both accuracy and precision) with previously published alternatives. © 1998 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **3-D freehand imaging, Calibration, Position sensor.**

INTRODUCTION

3-D ultrasound is a new imaging modality that has already been recognised as a valuable tool for a variety of clinical applications. Conventional 2-D diagnostic imaging uses a hand-held probe that transmits ultrasound pulses into the body and receives the echoes. The magnitude and timing of the echoes are used to create a 2-D grey-level image (B-scan) of a cross-section of the body in the scan plane. 3-D ultrasound extends this concept so that volumes of intensity data are created from pulse-echo information.

High-quality, rapid 3-D imaging remains a long-term research goal. One promising approach centres around the development of a new type of phased array probe, which sends and receives echoes from a 2-D array of elements (instead of the usual 1-D array). Unfortunately, several technical challenges must be overcome before such probes receive clinical acceptance (Smith et al. 1995). Alternative approaches, which make use of conventional 2-D ultrasound technology, include the freehand and swept-volume techniques (Rankin et al. 1993; Steiner et al. 1994). Instead of taking a 3-D snapshot, these techniques construct a 3-D data set from a number of 2-D B-scans acquired in rapid succession.

The swept-volume approach uses a special mechanism inside the probe to sweep the plane of the B-scan through a volume of interest. In contrast, the freehand approach relies on the physician to guide a standard probe over the volume, while a position sensor, attached to the probe, measures the B-scans' relative positions. Each method has its advantages. Swept-volume systems are easy to use and produce standardised volumes of densely sampled data without irregular gaps. However, they require the considerable expense of a dedicated machine, and are also limited to a maximum volume dictated by hardware constraints in the probe.

Freehand systems can be used to obtain arbitrary volumes of data because the motion of the probe is unconstrained. They are also cheaper, requiring only existing conventional ultrasound systems and relatively inexpensive additional components. However, the physician needs to learn how to move the probe to acquire regular, densely sampled data sets. The cost and flexibility of freehand imaging ensures it remains a popular choice and, with recent improvements in the speed of data acquisition combined with careful scanning practice, the quality of freehand data sets is improving. For these reasons, research into freehand systems is very active, and several commercial systems have recently become available.

A 3-D freehand examination can be broken into three stages: scanning, reconstruction and visualisation

---

Address correspondence to: Richard W. Prager, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ UK. E-mail: rwp@eng.cam.ac.uk
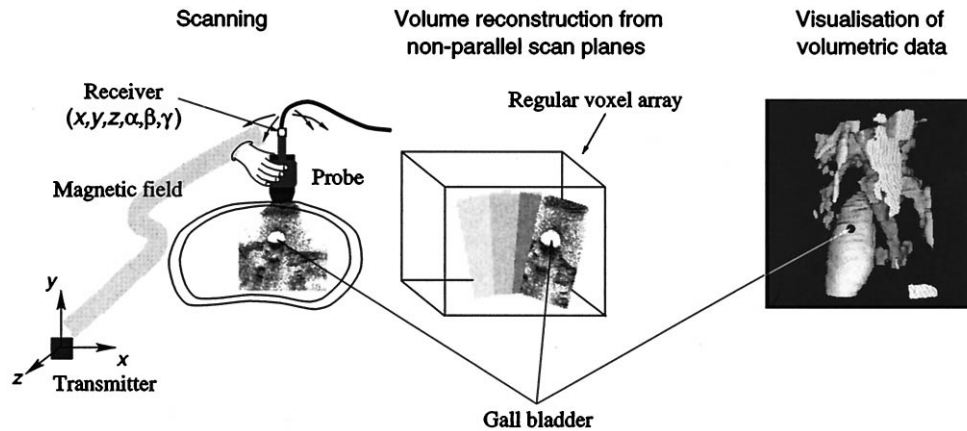
Fig. 1. 3-D freehand ultrasound imaging. Clinical freehand imaging involves scanning, reconstruction and visualisation. This example illustrates an examination of a gall bladder.

(see Fig. 1). Before scanning, some sort of position sensor is attached to the probe. This is typically the receiver of an electromagnetic position sensor, as illustrated in Fig. 1. Measurements from the position sensor are used to determine the positions and orientations of the B-scans with respect to the fixed transmitter. In the next stage, the set of acquired B-scans and their relative positions are used to fill a regular voxel array. Finally, this voxel array is visualised using, for example, any-plane slicing, volume rendering or surface rendering (after segmentation).

A key requirement of all freehand imaging systems is calibration. This involves determining the position and orientation of the B-scans with respect to the sensor mounted on the probe. The results of the calibration take the form of six constant offsets, three for position and three for orientation. These offsets must be combined with the sensor measurements to calculate the positions of the B-scans during reconstruction. Accurate calibration is essential for a consistent reconstruction that preserves true anatomical shape.

Calibration is required, regardless of the type of position sensor used in the examination. Previous work has described the use of acoustic spark gaps (King et al. 1991), mechanical arms (Ohbuchi et al. 1992), AC magnetic sensors (Barry et al. 1997; Hughes et al. 1996; Nelson and Elvins 1993), DC magnetic sensors (Detmer et al. 1994; Leotta et al. 1997), optical sensors (State et al. 1994; Trobaugh et al. 1994), and multiple sensors (Leotta et al. 1995). Only a few, however, have seriously considered the question of accurate calibration.

A rough estimate of the calibration parameters can be obtained by external measurements of the probe and position sensor. However, the origin of the sensor (for example, the centre of the wire coils in an electromag-

netic receiver) and the corner of the B-scan are not well defined with respect to the external cases of the sensor and probe. For these reasons, calibration is preferably performed by imaging a phantom, an artificial object with some known physical properties or dimensions. Measurements from an examination of the phantom, combined with its known physical properties, can be used to determine the six offsets.

Existing calibration techniques are not particularly easy to use. A typical calibration process, which often needs repeating every time a sensor is mounted on a probe, might take several hours for a skilled technician to perform. In this paper, we present a novel calibration technique that can be performed in a few minutes and produces results that compare favourably with previously published alternatives. The paper is organised as follows. First, we describe the reconstruction process and formulate the calibration problem. We then review existing calibration procedures and introduce the new technique that is the focus of this paper. Finally, we compare the accuracy, repeatability and usability of the techniques, and draw some conclusions from our results.

RECONSTRUCTION

Figure 2 shows the four coordinate systems used for reconstruction. $\mathbf{P}$ is the coordinate system of the B-scan plane, with an origin in the top left hand corner of the cropped image. The y-axis is in the beam direction, and the x-axis in the lateral direction. The z-axis is in the elevational direction, out of the plane of the B-scan. $\mathbf{R}$ is the coordinate system of the position sensor's receiver and $\mathbf{T}$ is the coordinate system of the transmitter. The reconstruction volume, created from the set of acquired B-scans, takes the form of a 3-D matrix $\mathbf{C}$ of volume
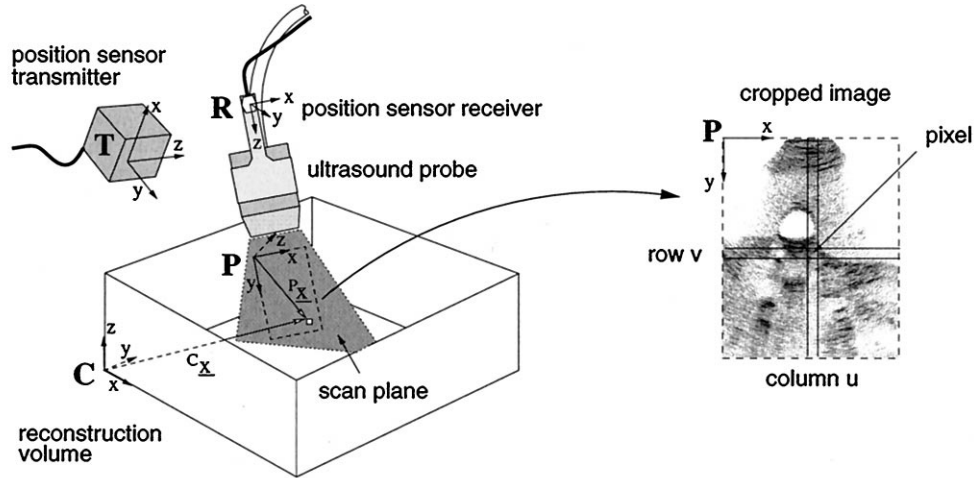
Fig. 2. Coordinate systems. The reconstruction process uses four coordinate systems. Note that the transmitter is often placed on a cantilevered support above the examination table, to give an unobstructed path to the receiver.

elements (voxels). The origin of the **C** coordinate system is at the corner of the reconstruction volume.

During reconstruction, every pixel in every B-scan has to be located with respect to the reconstruction volume **C**. Each pixel's scan plane location ($^P\underline{\mathbf{x}}$) is transformed first to the coordinate system of the receiver **R**, then to the transmitter **T** and, finally, to the reconstruction volume **C**. The overall transformation can be expressed as the multiplication of homogeneous transformation matrices:

$$^C\underline{\mathbf{x}} = {}^C\mathbf{T}_T\,{}^T\mathbf{T}_R\,{}^R\mathbf{T}_P\,{}^P\underline{\mathbf{x}}, \tag{1}$$

where

$$^P\underline{\mathbf{x}} = \begin{pmatrix} s_x u \\ s_y v \\ 0 \\ 1 \end{pmatrix}.$$

J. *u* and *v* are the column and row indices of the pixel in the cropped image, and s_x and s_y are scale factors with units of mm/pixel. $^C\underline{\mathbf{x}}$ is the pixel's location in the coordinate system **C**.

A transformation between two 3-D coordinate systems has six degrees of freedom: three rotation (α, β, γ) and three translation (x, y, z). There are many ways to parameterise a rotation, including unit quaternions, rotation matrices and a variety of three-angle systems: we adopt an *x-y-z* fixed-angles scheme (Craig 1989). The rotation between two coordinate systems is effected by first rotating through γ around the *x*-axis, then through β around the *y*-axis, and finally through α around the *z*-axis (the order of rotation is significant). The fixed rotation axes are aligned with the first coordinate system. Using this convention, the homogeneous matrix describing the transformation takes the following form (Craig 1989):

$$^J\mathbf{T}_I(x, y, z, \alpha, \beta, \gamma) = \begin{pmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & x \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & y \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma & z \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2}$$

We adopt a standard notation, such that $^J\mathbf{T}_I$ is the transformation from coordinate system *I* to coordinate system

Each of the transformation matrices in eqn (1) plays a different role in reconstruction. The most straightfor-
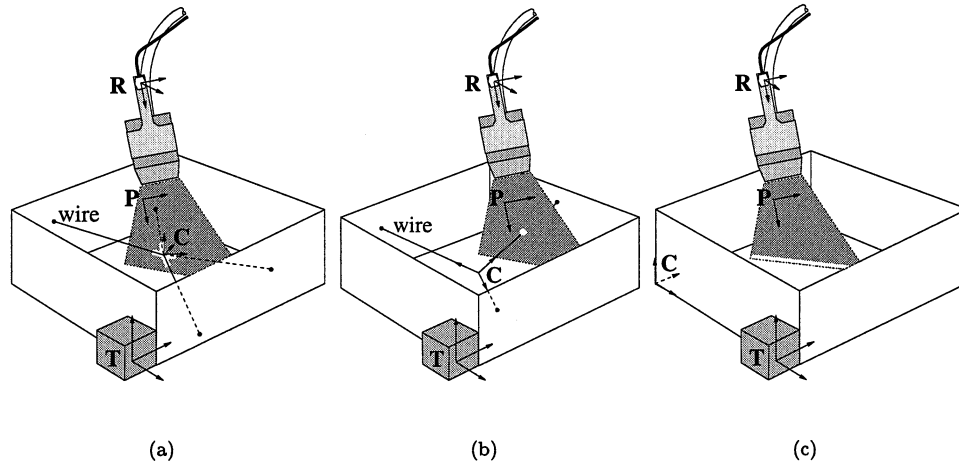
(a) (b) (c)

Fig. 3. Calibration phantoms. (a) Cross-wire phantom. A crossed wire is scanned in a water bath, producing a white cross in the B-scan. The crossed wires can be scanned from a variety of directions, but the scan plane must intersect the crossing point. The origin of **C** is placed at the point where the wires cross. (b) Three-wire phantom. Three mutually orthogonal, intersecting wires are mounted in a box. B-scans of each wire exhibit a white dot. The origin of **C** is located at the point where the wires cross, with the axes running along the three wires. (c) Single-wall phantom. The floor of the water bath is scanned, producing a white line in the B-scan. The origin of **C** is located on the floor of the water bath, with one axis orthogonal to the plane of the floor.

ward is $^{T}\mathbf{T}_{R}$, which is derived directly from the position sensor readings. In the context of reconstruction, $^{C}\mathbf{T}_{T}$ is included largely as a matter of convenience. We could omit it and align the reconstruction volume with the transmitter, but this might, for instance, place the B-scans a long way from the origin of **C**, resulting in a vast, largely empty voxel array. Or the anatomy might appear upside down in **C**, with the head toward the bottom of **C** and the feet toward the top: it all depends on the positioning of the transmitter with respect to the patient. Such effects can be corrected by allowing an arbitrary transformation $^{C}\mathbf{T}_{T}$, set as required for each examination. That leaves just $^{R}\mathbf{T}_{P}$, which needs to be determined by calibration. Although the scale factors s_{x} and s_{y} could be derived from the axis markings on the B-scan, we have decided to include them in the calibration for convenience.

After $^{C}\underline{\mathbf{x}}$ has been found for every pixel, the voxels of **C** can be set according to the intensities of the pixels they intersect. The reconstruction algorithm is described in more detail by Rohling et al. (1997). After a significant portion of the voxels are filled, the volume can be visualised by a number of methods: in this paper, we use surface rendering and any-plane slicing (Fenster and Downey 1996).

CALIBRATION PHANTOMS

Cross-wire phantom

Calibration is performed by scanning a phantom of known geometric dimensions. Equations similar to eqn

(1) can be written using knowledge of the phantom geometry and the position sensor measurements. These equations are then solved to determine the calibration parameters.

A common calibration phantom is the cross-wire phantom (Barry et al. 1997; Detmer et al. 1994). Two intersecting wires are mounted in a water bath, with the transmitter placed at some fixed location with respect to the wires (see Fig. 3a). For calibration purposes, the origin of **C** is not coincident with the corner of the reconstruction volume, but is placed, instead, at the intersection of the wires. This simplifies the calibration equations considerably.

The location where the wires cross is scanned repeatedly from different directions, with each B-scan showing a detectable cross (the reader may wish to skip ahead to Fig. 7a for a typical example). The pixel at the centre of the cross should satisfy:

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = {}^{C}\mathbf{T}_{T}{}^{T}\mathbf{T}_{R}{}^{R}\mathbf{T}_{P} \begin{pmatrix} s_{x}u \\ s_{y}v \\ 0 \\ 1 \end{pmatrix}. \tag{3}$$

The first three rows of eqn (3) give three equations involving the measurements $^{T}\mathbf{T}_{R}$, u and v, and the unknowns $^{R}\mathbf{T}_{P}$, $^{C}\mathbf{T}_{T}$, s_{x} and s_{y}. If there are m B-scans, then the equations can be stacked together to produce a system of nonlinear homogeneous equations of size $3m$:

$$\mathbf{0} = \mathbf{f}(\theta, \phi), \tag{4}$$

where θ are the measurements and ϕ are the unknowns. Given an overdetermined set of equations (so that $3m$ is greater than the number of unknowns ϕ), the system can be solved by several iterative methods: we choose the popular and robust Levenberg–Marquardt algorithm (More 1977). At iteration j, the update $\Delta\phi$ to the current estimate ϕ_j is derived from the first order Taylor expansion of eqn (4):

$$0 = \mathbf{f}(\theta, \phi) \approx \mathbf{f}(\theta, \phi_j) + \frac{\delta\mathbf{f}(\theta, \phi_j)}{\delta\phi}(\phi - \phi_j)$$

$$\Rightarrow \Delta\mathbf{f} = \mathbf{J}(\phi - \phi_j) = \mathbf{J}\Delta\phi, \quad (5)$$

where $\Delta\mathbf{f}$ is the error vector $-\mathbf{f}(\theta, \phi_j)$ and $\mathbf{J}$ is the gradient matrix $\delta\mathbf{f}(\theta, \phi_j)/\delta\phi$, also known as the Jacobian. The Levenberg–Marquardt algorithm produces the updated parameter values ϕ_{j+1}:

$$\phi_{j+1} = \phi_j + (\mathbf{J}^T\mathbf{J} + \varepsilon\mathbf{I})^{-1}\mathbf{J}^T\Delta\mathbf{f}, \quad (6)$$

where ε is a damping term chosen at each step to stabilise convergence, and $\mathbf{I}$ is the identity matrix. This method becomes the standard iterative least squares algorithm for $\varepsilon = 0$. At each step, $\Delta\mathbf{f}$ and $\mathbf{J}$ are evaluated at the current estimate ϕ_j. This process is iterated until the corrections $\Delta\phi$ are sufficiently small.

For calibration, we are only interested in the scale values and ${}^R\mathbf{T}_P$, but we must also solve for ${}^C\mathbf{T}_T$, even though we will subsequently discard these values and adopt an arbitrary, convenient ${}^C\mathbf{T}_T$ for reconstruction, as explained in the previous section. At first sight, therefore, it appears that ϕ must be a 14-element vector, composed of s_x and s_y, the six parameters of ${}^R\mathbf{T}_P$ and the six parameters of ${}^C\mathbf{T}_T$. However, it is clear from inspection of the geometry that the coordinate system $\mathbf{C}$ can be at any orientation and still satisfy eqn (3). This means that the three orientation angles of ${}^C\mathbf{T}_T$ are not identifiable. These angles are therefore removed from ϕ, and for convenience set to zero in eqn (4). Table 1 lists the identifiable parameters ϕ.

Several papers (Leotta et al. 1995, 1997; State et al. 1994) have described this calibration method using other point objects (such as a pinhead or a small suspended bead) instead of the crossed wires. The calibration equations, however, remain unchanged.

Accuracy of calibration with a point object depends on how well the centre of the point can be located, as well as the degree to which the point remains fixed in space with respect to the transmitter $\mathbf{T}$ (care must be taken that both the transmitter and phantom are rigidly mounted and not susceptible to mechanical vibrations). Given the quality of ultrasound images, it is common

Table 1. Identifiable parameters.

ϕ		Cross-wire	Three-wire	Single-wall
	s_x	✓	✓	✓
	s_y	✓	✓	✓
	x	✓	✓	✓
	y	✓	✓	✓
${}^R\mathbf{T}_P$	z	✓	✓	✓
	α	✓	✓	✓
	β	✓	✓	✓
	γ	✓	✓	✓
	x	✓	✓	–
	y	✓	✓	–
${}^C\mathbf{T}_T$	z	✓	✓	✓
	α	–	✓	–
	β	–	✓	✓
	γ	–	✓	✓

Some parameters of ${}^C\mathbf{T}_T$, marked –, are not identifiable by the cross-wire and single-wall methods, so are removed from the vector of parameters ϕ.

practice to locate the point by hand in each B-scan, making the calibration process fairly time-consuming.

Three-wire phantom

Another calibration technique uses a three-wire phantom (Carr 1996), as shown in Fig. 3b. This is similar to the cross-wire method because it involves scanning wires in a water bath, but it requires that the three wires be accurately mounted in orthogonal directions. The accuracy of calibration therefore depends on the orthogonality and straightness of the wires, as well as the degree to which the wires remain fixed in space with respect to the transmitter.

We place the coordinate system $\mathbf{C}$ at the origin of the wires, and orient the x, y and z axes along the wires. Each wire is scanned, one at a time, along its length from a variety of directions. The wire appears as a detectable dot in the B-scan: see Fig. 7b. For the wire along the x-axis, the pixel at the centre of the wire should satisfy:

$$\begin{pmatrix} x \\ 0 \\ 0 \\ 1 \end{pmatrix} = {}^C\mathbf{T}_T{}^T\mathbf{T}_R{}^R\mathbf{T}_P \begin{pmatrix} s_x u \\ s_y v \\ 0 \\ 1 \end{pmatrix}. \quad (7)$$

The left hand side of the equation becomes $(0 \ y \ 0 \ 1)^T$ for the y-axis and $(0 \ 0 \ z \ 1)^T$ for the z-axis. The two zero components of each equation give two equations in the unknowns ϕ. All six parameters of ${}^C\mathbf{T}_T$ are identifiable, giving 14 unknown parameters (see Table 1). As with the cross-wire phantom, many B-scans are acquired, the wires are located by hand in each image, and the over-
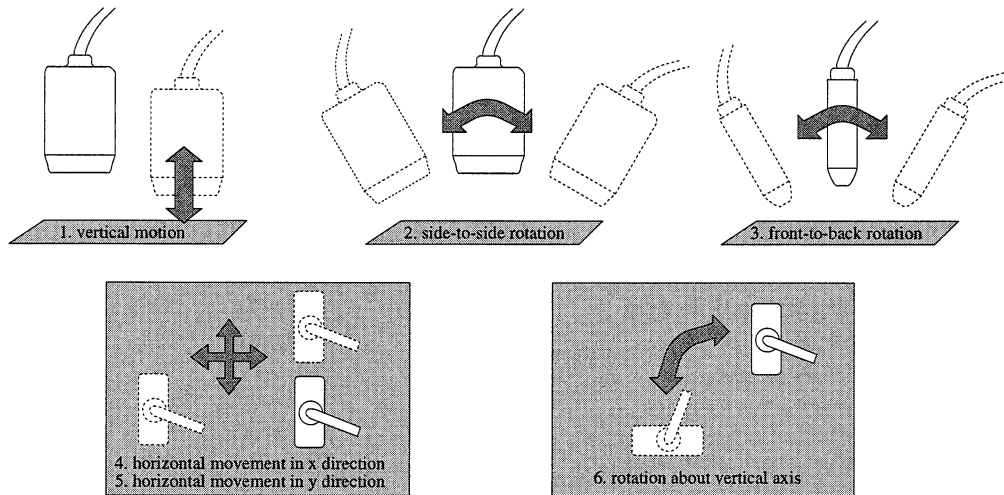
Fig. 4. Minimal sequence of motions for single-wall calibration. All six degrees of freedom must be exercised to identify the 11 parameters.

determined set of $2m$ equations is solved using the Levenberg–Marquardt algorithm.

The advantage of the three-wire method over the cross-wire method is that it is intrinsically easier to scan a length of wire than to keep the B-scan centred on a crossing point. However, each wire must be scanned separately, keeping track of which wire each B-scan intersects.

Single-wall phantom

While scanning the cross-wire and three-wire phantoms, we observed that the floor of the phantom produced clear, consistent lines in the B-scans. We therefore removed the wires and performed calibration using the floor of the water bath alone, as shown in Fig. 3c. The accuracy of the calibration will depend on the flatness of the floor, as well as the degree to which the floor remains fixed in space with respect to the transmitter.

The idea of calibrating a 3-D ultrasound system by scanning a wall of a water bath is intuitively attractive. Simple water baths with planar surfaces are readily available, so there is no need to construct a special phantom. Furthermore, the plane should show up as a strong, straight line in the image (Fig. 7c shows a typical example) and this line can be detected automatically. Points and crosses can also be detected automatically, but not as easily as lines. This is because there is much more redundant evidence for the presence of a line in an image; it is possible to locate a straight line even when portions of the line are corrupted or missing. The same cannot be said of dots and crosses. We describe a suitably robust, automatic line detection algorithm in the Appendix.

If the coordinate system $\mathbf{C}$ is defined to lie in the floor of the water bath, with the z-axis orthogonal to the floor, pixels lying on the line in the B-scan should satisfy:

$$\begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} = {}^{C}\mathbf{T}_{T}{}^{T}\mathbf{T}_{R}{}^{R}\mathbf{T}_{P} \begin{pmatrix} s_x u \\ s_y v \\ 0 \\ 1 \end{pmatrix}. \qquad (8)$$

The zero component of the equation gives one equation in the unknown parameters ϕ. However, we can write the eqn at two pixels on the line (two points uniquely define a line), giving two eqns per B-scan.

From inspection of the geometry, it is evident that several parameters of ${}^{C}\mathbf{T}_{T}$ are not identifiable. The rotation of the floor about the z-axis (α) and the translation of the floor in its own plane (x, y) do not affect the z-component of ${}^{C}\mathbf{x}$. This leaves 11 identifiable parameters (see Table 1).

The optimisation algorithm needs at least 11 independent equations to solve for these 11 unknowns. Although each B-scan can be used to write two equations, the resulting equations will not necessarily be independent unless the probe is moved appropriately while scanning the wall (see Fig. 4). As before, the overdetermined set of $2m$ equations is solved using the Levenberg–Marquardt algorithm.

Cambridge phantom

Single-wall calibration is susceptible to problems caused by the width of the ultrasound beam and the nature of specular reflection. When the beam is not
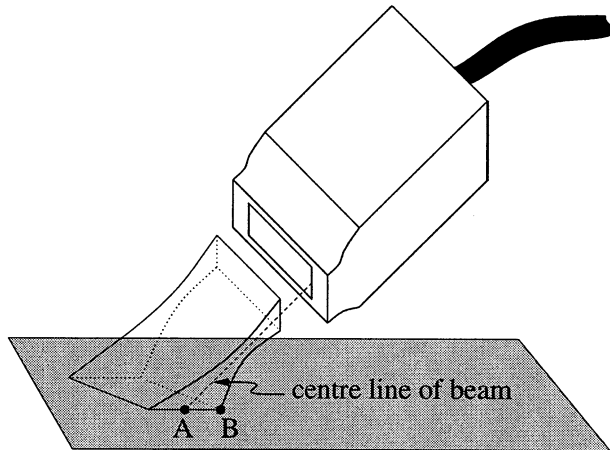
Fig. 5. Beam thickness problem in single-wall calibration. When the wall is scanned from oblique angles, the finite beam thickness results in a blurred image of the surface. This problem arises because point B is encountered by the ultrasound pulse before point A on the centre line. The echo from point B produces a response in the ultrasound image that does not reflect the true position of the wall.

Fig. 6. Cambridge phantom. Accurate calibration requires a minimal sequence of motions covering all six degrees of freedom (compare with Fig. 4).

normal to the wall, the first echo to return to the probe comes from the edge of the beam closest to the wall (see Fig. 5). This effect introduces position errors into those B-scans taken at oblique angles. The problem is compounded by the weak echo obtained when the wall is scanned obliquely, because much of the ultrasound energy is specularly reflected away from the probe and the amount of Lambertian reflection is small. Both these effects conspire to spoil the ultrasound image when the scan plane is not normal to the wall. Unfortunately, such scanning angles are required for effective calibration, as discussed previously.

A new phantom, dubbed the "Cambridge phantom" (Prager 1997), has been designed to overcome the difficulties experienced with planar calibration. The phantom consists of two parts: a clamp that fits around the probe, and a thin brass bar mounted between two circular disks (see Fig. 6). The top of the bar, which is approximately 1.5 mm thick, is squared-off and slightly roughened with emery cloth.

The idea is that the clamp constrains the thin bar to move only in the centre of the ultrasound beam. The bar is attached to the disks in such a way that the upper edge of the bar is aligned with the centre of each disk. This means that, as the disks are rolled from side to side, the upper edge of the bar stays at a constant height above the floor. In essence, the Cambridge phantom is similar to the single-wall phantom, but the wall is a virtual plane traced out by the top of the bar. This plane has two remarkable properties. First, only the slice required to reflect the centre of the ultrasound beam exists at each
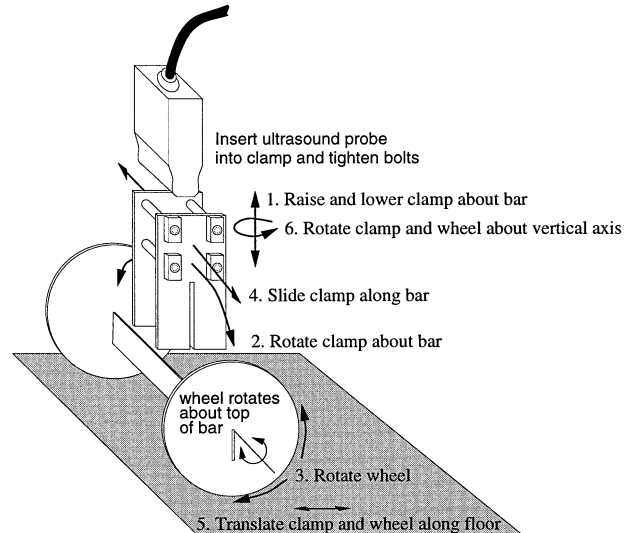
moment in time. Second, that slice is usually oriented toward the ultrasound probe, so the beam is reflected straight back to the probe, producing a strong, clear image (see Fig. 7d). The exception is when the clamp is rotated about the bar (motion 2 in Fig. 6). Even with this motion, however, we have experienced little difficulty obtaining clear images of the bar.

The calibration procedure using the Cambridge phantom may be summarised as follows:
1. Place the clamp around the ultrasound probe. Adjust it so that the slots through the two sides of the clamp
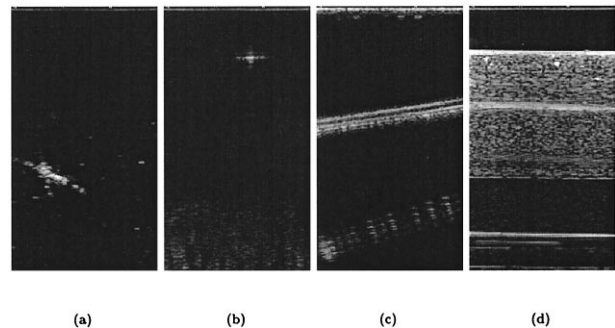

Fig. 7. Typical B-scans of calibration phantoms. (a) Cross-wire. The cross from the two wires is located near the bottom left-hand corner of the cropped image. (b) Three-wire. The dot near the top of (b) is one of the wires of the three-wire phantom. (c) Single-wall. The top line is the floor of the single-wall phantom, the second one down is a reverberation. (d) Cambridge. A typical image of the Cambridge phantom is shown; the top of the brass bar is clear and distinct.

are aligned with the scan plane of the probe. This is easily checked by ensuring that a clear image is obtained when the clamp is placed over the bar. Alternatively, the technique described in the Appendix can be used to align the bar with the scan plane more accurately.

2. Immerse the phantom in water, slot the clamp over the bar, and scan the bar with the probe from all possible angles, subject to the constraints imposed by the clamp. The shape of the phantom and the way it interacts with the clamp ensure that a clear image of the bar will always be visible in the B-scan.

3. Because of the clarity of the images and the fact that the phantom produces an edge (as opposed to a dot) in the B-scan, it is possible to detect the edge automatically in each B-scan (see the Appendix).

4. After these edges have been located, calibration proceeds in exactly the same manner as described for the single-wall phantom.

Figure 6 shows the minimum sequence of motions required for calibration. The calibration procedure is highly automated and takes about 5 minutes to complete, including scanning, edge detection and optimisation. Such speed is not possible with the wire-based techniques, because the dots need to be located by hand in each image.

COMPARISON OF CALIBRATION TECHNIQUES

Method

The cross-wire phantom was constructed by crossing two cotton wires and immersing them in a water bath, as described by Detmer et al. (1994). The three-wire phantom comprised 1 mm diameter nylon wires placed under tension between precision drilled holes in the walls of a Perspex water bath, as described by Carr (1996). The holes were positioned to make the wires orthogonal. The single-wall phantom was no more than the floor of a water bath. The weight of the water, acting on a flat table under the floor of the bath, helped to reduce any error arising from the flexibility of the floor.

The ultrasound machine used in our experiments was a Toshiba model SSA-270A/HG. Calibration experiments were performed on a 7-MHz linear array probe. A depth setting of 40 mm was used throughout. Other imaging controls, such as time-gain compensation, focus, overall gain and acoustic power, were adjusted to give optimal images of each phantom. Figure 7 shows some typical B-scans of the phantoms.

The position sensor was an AC magnetic field receiver-transmitter pair (Polhemus FASTRAK, Polhemus Incorporated, Colchester, VT). To minimise interference from the electronics inside the ultrasound probe, the
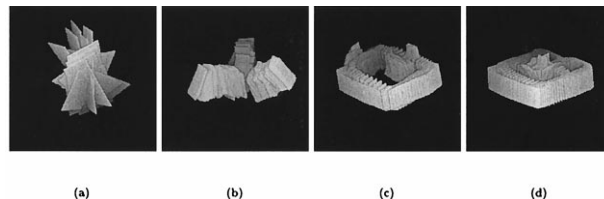

Fig. 8. Scanning protocols. (a) Cross-wire phantom. (b) Three-wire phantom. (c) Single-wall phantom. (d) Cambridge phantom. The pictures illustrate a typical scan of each phantom. Table 2 provides details of the range of angles and positions covered by each scan.

receiver was mounted about 25 cm from the probe face at the end of a rigid nylon extension bar. The mounting was not disturbed during the experiments, so successive calibration procedures were expected to produce identical results.

Images from the scanner were digitised using an 8-bit frame grabber, tagged with the position data, and stored in the memory of a Silicon Graphics Indy workstation. Although our acquisition software can record at 25 frames/s, we selected a lower frame rate for the calibration experiments, to allow significant motion of the probe between frames.

Twenty calibration experiments were performed using each of the cross-wire, single-wall and Cambridge phantoms. Practical constraints limited our evaluation of the three-wire phantom to 10 trials. The scanning protocols we adopted are described in Fig. 8a–d and Table 2. Any scans that did not reveal a clear image of the phantom were discarded. For the Cambridge phantom, the probe was completely removed from the clamp between each trial, and then repositioned using the technique described in the Appendix. For the cross-wire and three-wire phantoms, the pixel coordinates (u, v) needed

Table 2. Scanning protocols.

	Cross-wire	Three-wire	Single-wall	Cambridge
Number of B-scans	~ 53	~ 193	~ 578	~ 530
x translation (mm)	± 20	± 84	± 67	± 108
y translation (mm)	± 20	± 84	± 67	± 71
z translation (mm)	± 12	± 44	± 17	± 17
Max. inclination to z-axis	65°	50°	19°	69°
Twist of B-scan	± 87°	± 90°	± 90°	± 90°

The table gives approximate values for the ranges of angles and positions covered by the scans in Fig. 8. The translations relate to the centre of the probe face, and the z-axis is vertically upward. The angle of inclination to the z-axis defines a cone of insonation directions, and the twist angle describes rotation of the B-scan around its centre line. Relatively large numbers of B-scans were acquired for the Cambridge and single-wall phantoms because the images they produce are amenable to automatic segmentation.

for the calibration equations were extracted by hand in each B-scan. The pixel coordinates for the single-wall and Cambridge phantoms were determined automatically, using the line-detection algorithm described in the Appendix. In all experiments, we used the Levenberg–Marquardt algorithm, eqn (6), to determine ϕ, taking care to remove unidentifiable parameters and scale the remaining parameters to produce as well-conditioned a problem as possible (details can be found in the Appendix).

For all calibration experiments, the water bath was filled with water at room temperature. This means that the estimated scale factors s_x and s_y will reflect the speed of sound in water and not human tissue. Because the water temperature was stable (the temperature in the room was thermostatically controlled), any variations observed in the calibration parameters cannot be due to a change in the prevailing speed of sound. Of course, a real calibration before an *in vivo* examination should be performed in a fluid in which the speed of sound matches that observed in human tissue.

We assessed the precision of the calibration solutions by looking at their stability across the multiple trials of each phantom. To this end, we measured the change in the position (relative to the position sensor's receiver) of the bottom right hand corner (u_{max}, v_{max}) of the B-scan under two different calibrations:

$$\Delta^R\underline{\mathbf{x}} = {}^R\mathbf{T}_P^{\text{trial 1}} \, {}^P\underline{\mathbf{x}}^{\text{trial 1}} - {}^R\mathbf{T}_P^{\text{trial 2}} \, {}^P\underline{\mathbf{x}}^{\text{trial 2}}, \qquad (9)$$

where

$$^P\underline{\mathbf{x}}^{\text{trial } i} = \begin{pmatrix} s_x^{\text{trial } i} \times u_{max} \\ s_y^{\text{trial } i} \times v_{max} \\ 0 \\ 1 \end{pmatrix}.$$

This measure involves all six parameters of ${}^R\mathbf{T}_P$ and the two scale values. So $|\Delta^R\underline{\mathbf{x}}|$ indicates the precision with which points in the scan plane can be defined relative to the position sensor's receiver. It is a direct measure of calibration precision that does not depend on other sources of error associated with reconstruction, such as segmenting points in a reconstruction volume. For the Cambridge, cross-wire and single-wall phantoms, we calculated 190 values of $|\Delta^R\underline{\mathbf{x}}|$ (using all possible pairs drawn from the 20 trials). For the three-wire phantom, we calculated 45 values of $|\Delta^R\underline{\mathbf{x}}|$ (using all possible pairs drawn from the 10 trials).

We also assessed the reconstruction precision by scanning the cross-wire phantom immediately following completion of the calibration scans. The cross was observed from 49 different viewpoints, covering a range of
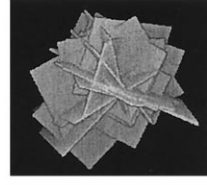
x translation	±18 mm
y translation	±20 mm
z translation	±7 mm
max. inclination to z-axis	67°
twist of B-scan	±87°

Fig. 9. Scanning pattern used for reconstruction precision experiments. With this motion, the cross was observed within the full depth range of the B-scan (40 mm). Please refer to the caption of Table 2 for an explanation of the motion parameters.

insonation angles and depths, as shown in Fig. 9. The location of the cross was identified by hand in each B-scan, and then mapped to the reconstruction space using the calibration parameters and the position sensor readings. This was done for the multiple calibration solutions obtained using each of the four calibration phantoms, giving 70 reconstructions in total.

In each reconstruction, the cross appeared as a cloud of 49 points. We assessed the reconstruction precision by looking at the tightness of this cloud. To this end, we measured the change in the position (relative to the position sensor's transmitter) of the cross observed from two different viewpoints:

$$\Delta^T\underline{\mathbf{x}} = {}^T\mathbf{T}_R^{\text{view 1}} \, {}^R\mathbf{T}_P \, {}^P\underline{\mathbf{x}}^{\text{view 1}} - {}^T\mathbf{T}_R^{\text{view 2}} \, {}^R\mathbf{T}_P \, {}^P\underline{\mathbf{x}}^{\text{view 2}}, \quad (10)$$

where

$$^P\underline{\mathbf{x}}^{\text{view } i} = \begin{pmatrix} s_x \times u^{\text{view } i} \\ s_y \times v^{\text{view } i} \\ 0 \\ 1 \end{pmatrix}.$$

So $|\Delta^T\underline{\mathbf{x}}|$ indicates thex precision with which points can be located in the reconstruction volume. We calculated 1176 values of $|\Delta^T\underline{\mathbf{x}}|$ (using all possible pairs drawn from the 49 views). We repeated the process for each of the 70 reconstruction volumes.

We also assessed the reconstruction accuracy by scanning the phantom shown in Fig. 10. The phantom, which was also scanned immediately following completion of the calibration scans, comprises 8 dressmaker's pins arranged in a plane. Each pin was observed from eight different viewpoints, covering a range of insonation angles and depths, as shown in Fig. 11. The locations of the pins were identified by hand in each B-scan, and then mapped to the reconstruction space using the calibration parameters and the position sensor readings. This was done for the multiple calibration solutions obtained using each of the four calibration phantoms, giving 70 reconstructions in total.
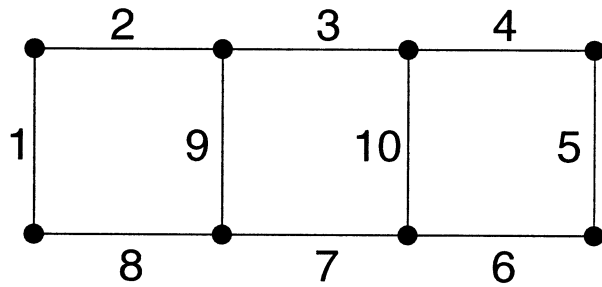
Fig. 10. Phantom used for assessing calibration accuracy. The phantom comprises eight dressmaker's pins arranged in a plane as shown. Each of the ten inter-pin distances marked on the diagram was about 100 mm.

To assess the accuracy of the reconstruction, we looked at the 10 inter-pin distances marked in Fig. 10. Each of these distances was measured directly using sophisticated metrology equipment to an accuracy in excess of 0.01 mm. In each reconstruction, the 10 inter-pin distances were estimated 64 times, using all combinations of the eight observations of each pin. The error in each of these measurements was used to assess the calibration accuracy. In total, we obtained 640 error measurements in each of the 70 reconstructions.

Results

Table 3 shows statistics indicative of the quality of the data being used to estimate the calibration parameters. For all four phantoms, a large proportion of the acquired B-scans revealed the calibration feature (wire or plane) clearly enough that the feature could be reliably segmented and used to generate some of the calibration eqns (4). In general, the Cambridge phantom produced the highest quality images: see Fig. 7 for some typical B-scans.

For all phantoms, the Levenberg–Marquardt algorithm, eqn (6), converged to a stable solution in around 15 iterations. We did not encounter any problems with local minima; the algorithm repeatedly converged to a unique solution, independent of starting position (though

Table 3. Quality of calibration data.

	Cross-wire	Three-wire	Single-wall	Cambridge
No. trials	20	10	20	20
B-scans used	51 of 53	106 of 193	480 of 578	530 of 530
RMS error (mm)	0.56	1.04	0.48	0.34

The table shows the proportion of B-scans deemed to be of acceptable quality to be included in the calibration, averaged across all trials of each phantom. It also shows the root mean square residual of the set of eqns (4) at the solution ϕ, also averaged across all trials of each phantom.

please note the discussion on "mirror" solutions in the Appendix. In an investigation to determine the conditioning of the four calibration methods, we found that the condition numbers were consistently below 100 for the Cambridge, cross-wire and three-wire phantoms, indicating good identifiability of parameters (see the Appendix). Only the single-wall phantom produced slightly ill-conditioned sets of calibration equations. This can be attributed to the limited range of scanning motions that resulted in clear images of the wall: it was not possible to observe the wall from a sufficient range of angles to produce a well-conditioned problem.

After the Levenberg–Marquardt algorithm had converged, we evaluated the RMS residual of the system of eqns (4) at the solution ϕ: these are listed in Table 3. Assuming ϕ is the globally optimal solution, the residual errors give an indication of the self-consistency of the calibration equations; perfectly consistent equations would produce an RMS error of zero.

Table 4 shows the results of the calibration precision experiments described in the previous section. These results indicate to what extent repeated trials with the same phantom produced the same calibration parameters.

Table 5 shows the results of the reconstruction precision experiments described in the previous section. These results indicate how precisely the position of a point can be defined in a 3-D reconstruction volume. They measure the size of the cloud obtained when imaging a point from different directions, but say nothing

x translation	±156 mm
y translation	±57 mm
z translation	±14 mm
max. inclination to z-axis	31°
twist of B-scan	±83°

Fig. 11. Scanning pattern used for reconstruction accuracy experiments. With this motion, the pinheads were observed within the full depth range of the B-scan (40 mm). Please refer to the caption of Table 2 for an explanation of the motion parameters.

Table 4. Calibration precision.

	Cross-wire	Three-wire	Single-wall	Cambridge
No. observations	190	45	190	190
Mean (mm)	1.47	5.37	3.27	0.92
Max (mm)	3.37	12.09	10.07	2.32

The quantity tabulated is $|\Delta^R \mathbf{x}|$ for the bottom right-hand corner of the B-scan, where $|\Delta^R \mathbf{x}|$ is defined in eqn (9). $|\Delta^R \mathbf{x}|$ was calculated for each pair of calibration solutions drawn from the multiple trials of each phantom.

Table 5. Reconstruction precision.

	Cross-wire	Three-wire	Single-wall	Cambridge
No. observations	23520	11760	23520	23520
Mean (mm)	1.65	2.67	3.43	2.17
Max (mm)	4.78	10.31	11.67	7.13

The quantity tabulated is $|\Delta^T\underline{x}|$ for a point target (the cross in the cross-wire phantom), where $|\Delta^T\underline{x}|$ is defined in eqn (10). $|\Delta^T\underline{x}|$ was calculated for each pair of views drawn from the 49 views of the cross, giving a total of 1176 values per reconstruction.

about reconstruction accuracy (whether the cloud is in the right place or not). Table 6 shows the results of the accuracy experiments with the pinhead phantom described in the previous section.

DISCUSSION

There are many ways to assess the performance of a calibration procedure. We might start by asking, "If I do another calibration tomorrow, will I get the same result?" The figures in Table 4 help us answer this question. The calibration parameters define the position and orientation of a plane, the scan plane, with respect to the position sensor's receiver. We have measured how a supposedly fixed point on that plane (the bottom right-hand corner of the B-scan) moves around under repeated calibrations with the same phantom. The Cambridge phantom appears to provide the most stable calibration parameters. Of course, we have no idea whether these calibration parameters are correct or not: they're just stable. However, it is clear that stability is a prerequisite for a "perfect" calibration technique.

To decide if the calibrations are correct, we might again ask a question: "Given a calibration, if I look at a point from different angles, will it always map to precisely the same point in the reconstruction volume?" The figures in Table 5 relate to this question. The Cambridge and cross-wire calibrations appear to produce the most precise reconstructions; if we look at a point from two significantly different directions, we can expect a distance of about 2 mm between the two reconstructed

positions of the point. With reference to the results in Table 4, this suggests that the Cambridge calibrations are somewhat biased. They are tightly bunched together, but not evenly spread around the optimal solution. The cross-wire calibrations are more variable, but more equally disposed around the optimal solution.

We have yet to establish full confidence in the calibrations. We know that points are reconstructed as, more or less, points, but we do not yet know whether these points are in the right place or not. So we ask one, final question: "If I scan something that is 100 mm long in the real world, will it come out as 100 mm long in the reconstruction?" The figures in Table 6 help us answer this question. The cross-wire calibration is remarkably accurate in this respect. There is hardly any bias in the measured lengths, and the root mean squared error is just over 1 mm, about 1% of the length being measured. The Cambridge phantom is also very good; it has a slight tendency to produce overestimates of length (by about 0.25 mm on average), but can still measure lengths to within about 1.5% of their true values.

The calibrations produced by the single-wall and three-wire phantoms are not as good. There is significant uncertainty in the calibration parameters (Table 4) that leads to relatively poor point-reconstruction precision (Table 5). The length accuracy figures (Table 6) do not reflect these errors to the same extent, because the measurements are not sensitive to point-reconstruction errors out of the plane of the pinheads.

In motivating the design of the Cambridge phantom, we have already identified what are probably the two main sources of error in single-wall calibrations. The first concerns identifiability of the calibration parameters. Because only a limited amount of energy is reflected back toward the probe, calibration scans can be acquired from only a limited range of viewing angles. This leads to relatively poor conditioning of the calibration problem. This factor could be diminished by roughening the floor of the water bath, producing better images of the plane from a wider range of angles. The other probable source of calibration error is the effect of the finite beam width.

Table 6. Reconstruction accuracy.

	Cross-wire	Three-wire	Single-wall	Cambridge
No. observations	12,800	6400	12,800	12,800
Mean (mm)	0.04	−0.15	0.14	0.23
Std. dev. (σ_{n-1}) (mm)	1.12	2.18	1.63	1.33
Min (mm)	−3.98	−8.10	−6.53	−5.62
Max (mm)	4.31	7.83	5.72	4.46

The quantity tabulated is the error in each of the 10 inter-pin distances shown in Fig. 10. The true inter-pin distances were all approximately 100 mm. As explained in the text, 640 error measurements were obtained in each reconstruction.
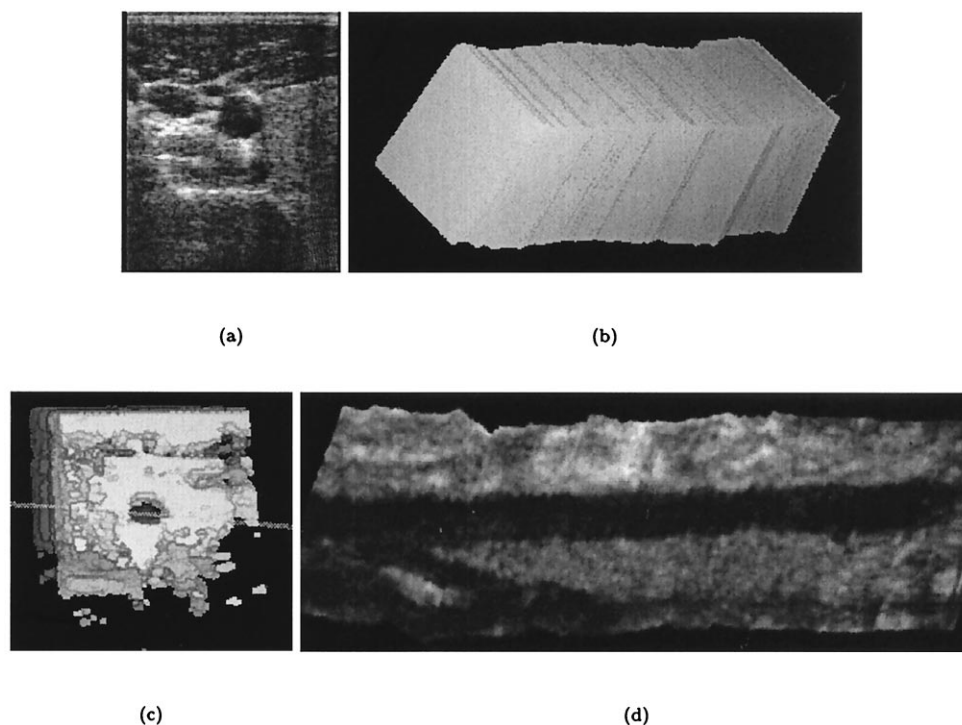
Fig. 12. Carotid artery. (a) B-scan of the carotid artery; the artery is the dark circle just above the centre of the image. (b) Outline of B-scans; a surface rendering of the outlines of the B-scans is reconstructed using a calibration obtained with the Cambridge phantom. The reconstruction volume has dimensions of $277 \times 242 \times 299$ voxels, with a voxel dimension of 0.3 mm. (c) Axial view of artery; a surface rendering is produced with a narrow grey-level threshold, looking down the artery. (d) Reslice of voxel array; a grey-level image is obtained by slicing the voxel array at the location shown in (c). A 90 mm length of the carotid artery is visible: such a view cannot be obtained using standard 2-D ultrasound.

Roughening the floor of the water bath will do nothing to reduce beam width effects.

The performance of the three-wire phantom is disappointing. In Table 3, we see that the equations solved for the calibration parameters are not as self-consistent as with the other techniques. This suggests inaccuracies in the phantom construction; we are trying to solve equations for mutually orthogonal wires, and yet the wires are not, in reality, mutually orthogonal. The three-wire phantom was by far the most difficult to construct, requiring great precision in the positioning of the wires.

Other researchers (Detmer et al. 1994; King et al. 1991; Leotta et al. 1995, 1997) have used similar measures of reconstruction accuracy and precision; these should be compared with extreme caution to those reported in this paper. Measured reconstruction performance will depend critically on the range of scanning angles and positions, as well as the type of position sensor and the quality of the ultrasound equipment. The results presented in Tables 5 and 6 are meaningful only in the context of examinations using similar ultrasound equipment and position-sensing devices, and covering a similar range of angles and positions to those described

in Figs. 9 and 11. In range-of-motion studies, we have observed that many common clinical examinations[1] fall within the envelope of motions described in Figs. 9 and 11. Thus, Tables 5 and 6 provide an upper bound on the reconstruction accuracy and precision one can expect in such clinical examinations. We say "upper bound" because *in vivo* reconstructions are also susceptible to a whole range of other errors, including motion and refraction artifacts.

A reconstruction using B-scans obtained from a wider range of scanning angles (as might be required for a bilateral carotid artery examination, for example) will probably be less accurate. In such circumstances, accuracy might be improved by calibrating against a phantom that allows approach from a wider range of directions, such as a sealed version of the cross-wire phantom (Detmer et al. 1994).

Another important factor by which a calibration technique should be assessed is ease of use. It is rather

---

[1] We monitored routine clinical examinations of the left kidney, right kidney, liver, Achilles tendon, rotator cuff, aorta and thyroid gland.

difficult to quantify ergonomics, but there can be little doubt that calibrating with the Cambridge phantom is significantly easier and quicker than with the cross-wire phantom. Although the Cambridge technique cannot be claimed to be fully automatic, many of the more time-consuming processes inherent in the use of the other phantoms are circumvented. For instance, there is the automatic segmentation of the calibration feature, and also the alignment of the scan plane with the feature; this has to be done once for the Cambridge phantom (when inserting the probe into the clamp), but once for every observation of the cross for the cross-wire phantom (when positioning the probe to obtain an image of the cross from the centre of the ultrasound beam).

To illustrate the convenience of the Cambridge phantom, we include in Fig. 12a–d an *in vivo* 3-D reconstruction of a carotid artery, available within 15 minutes of attaching the position sensor to the ultrasound probe. The calibration was performed using a Cambridge phantom submerged in suitably warm water (so that the speed of sound matches that observed in human tissue). The *in vivo* data acquisition was not gated to an ECG, so the pulsatile motion of the artery results in small ripples in the reconstruction. Other reconstruction errors are caused by variation in the contact pressure between the probe and the skin (this is particularly evident in the irregular skin line at the top of Fig. 12d), refraction of the ultrasound beam and changes in the speed of sound as the pulse travels through different kinds of tissue (Wells 1993). These errors cannot be addressed by calibration. Despite these limitations, fairly clear and consistent 3-D reconstructions are obtained.

CONCLUSIONS

In this paper, we have addressed the issue of freehand 3-D ultrasound calibration. We have reviewed existing techniques and assessed their performance under representative (though not extreme) scanning conditions. Although the cross-wire technique produced impressively accurate and repeatable calibrations, the calibration process took a long time to perform, typically several hours. This motivated us to examine calibration using a flat plane, because the line it produces in the B-scan can be detected reliably and automatically, greatly speeding up the calibration process. Problems with beam-width effects led to the design of the Cambridge phantom, which suppresses these effects and guarantees clear images of the virtual plane. The performance of the Cambridge phantom compared favourably with the other techniques, and calibration was comfortably performed in a matter of minutes.

REFERENCES

Barry CD, Allott CP, John NW, et al. Three-dimensional freehand ultrasound: image reconstruction and volume analysis. Ultrasound Med Biol 1997;23:1209–1224.

Carr J. Surface reconstruction in 3D medical imaging. Ph.D. Thesis, University of Canterbury, Christchurch, New Zealand 1996.

Clarke JC, Carlsson S, Zisserman A. Detecting and tracking linear features efficiently. Proc British Machine Vision Conference, 1996: 415–424.

Craig JJ. Introduction to robotics: mechanics and control. Reading, MA: Addison–Wesley, 1989.

Detmer PR, Bashein G, Hodges T, et al. 3D ultrasonic image feature localization based on magnetic scanhead tracking: *In vitro* calibration and validation. Ultrasound Med Biol 1994;20:923–936.

Fenster A, Downey DB. 3-D ultrasound imaging—a review. IEEE Eng Med Biol Mag 1996;15:41–51.

Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 1981;24:381–395.

Hollerbach JM, Wampler CW. The calibration index and taxonomy for robot kinematic calibration methods. Intl J Robot Res 1996;15: 573–591.

Hughes SW, D'Arcy TJ, Maxwell DJ, et al. Volume estimation from multiplanar 2D ultrasound images using a remote electromagnetic position and orientation sensor. Ultrasound Med Biol 1996;22:561–572.

King DL, King DL Jr, Shao MY. Evaluation of *in vitro* measurement accuracy of a three-dimensional ultrasound scanner. J Ultrasound Med 1991;10:77–82.

Lawson CL, Hanson RJ. Solving least squares problems. Englewood Cliffs, NJ: Prentice–Hall, 1974.

Leotta DF, Detmer PR, Gilja OH, et al. Three-dimensional ultrasound imaging using multiple magnetic tracking systems and miniature magnetic sensors. Proc IEEE Ultrason Symp 1995;2:1415–1418.

Leotta DF, Detmer PR, Martin RW. Performance of a miniature magnetic position sensor for three-dimensional ultrasound imaging. Ultrasound Med Biol 1997;24:597–609.

More JJ. The Levenberg–Marquardt algorithm: implementation and theory. In: Watson GA, ed. Numerical analysis. Lecture Notes in Mathematics 630. Berlin: Springer-Verlag, 1977:105–116.

Nelson TR, Elvins TT. Visualization of 3D ultrasound data. IEEE Comput Graphics Applic 1993;13:50–57.

Ohbuchi R, Chen D, Fuchs H. Incremental volume reconstruction and rendering for 3D ultrasound imaging. Proc Visualization Biomed Comput 1992;SPIE1808:312–323.

Prager RW. Ultrasound machine calibration. UK patent application no. 9716994.0. Filed: 11 August 1997.

Rankin RN, Fenster A, Downey DB, et al. Three-dimensional sonographic reconstruction: techniques and diagnostic applications. Am J Roentgenol 1993;161:695–702.

Rohling R, Gee A, Berman L. 3-D spatial compounding of ultrasound images. Med Image Anal 1997;1:177–193.

Schröer K. Theory of kinematic modelling and numerical procedures for robot calibration. In: Bernhardt R, Albright SL, eds. Robot calibration. London, UK: Chapman & Hall, 1993:157–196.

Smith SW, Davidsen RE, Emery CD, Goldberg RL, Light ED. Update on 2-D array transducers for medical ultrasound. Proc IEEE Ultraso Symp 1995:1273–1278.

State A, Chen DT, Tector C, et al. Case study: Observing a volume rendered fetus within a pregnant patient. Proc IEEE Visualization 1994:364–368.

Steiner H, Staudach A, Spitzer D, Schaffer H. Three-dimensional

ultrasound in obstetrics and gynaecology: technique, possibilities and limitations. Hum Reprod 1994;9:1773–1778.

Trobaugh JW, Trobaugh DJ, Richard WD. Three-dimensional imaging with stereotactic ultrasonography. Comput Med Imaging Graph 1994;18:315–323.

Wells PNT. Advances in ultrasound techniques and instrumentation. New York: Churchill Livingstone Inc., 1993.

APPENDIX

Mounting the probe in the Cambridge phantom

    A key step in the calibration procedure with the Cambridge phantom is mounting the ultrasound probe in the clamp (see Fig. 6). This has to be done so that the plane defined by the slot in each side of the clamp is aligned with the centre line of the ultrasound beam. For routine clinical work and experimentation, we have often performed this operation by eye; the probe is adjusted in the clamp until a clear ultrasound image of the bar is obtained.

    To increase the accuracy of this process, it is possible to clip four identical wedges over the bar, as shown in Fig. 13. A pair of wedges are mounted to the left side of the scan, and another pair are mounted on the right. When the left side of the scan beam is centred correctly, the two wedges on the left will be imaged as two vertical bars (toward the left of the B-scan) with the same height. Should the probe be moved off-centre, then one wedge will produce a longer vertical bar than the other. The symmetry needs to be checked for both left and right wedges, and for the top and bottom of the scan beam (by moving the probe toward and away from the bar). Typically, this process needs to be iterated a couple of times until full symmetry is achieved. The clamps are then tightened to prevent further slipping. The procedure takes about 2 minutes to perform.

Automatic detection of lines in ultrasound images

    For the Cambridge and single-wall phantoms, the plane (virtual or physical) appears as a straight line in the B-scans. Measuring its position and orientation comes down to a fairly standard edge detection procedure. However, given the low signal-to–noise ratio of the images (especially with the single-wall phantom), it is necessary to choose a robust edge-fitting algorithm that incorporates some outlier rejection.

    The approach we adopt is essentially a simplified version of the scheme described by Clarke et al. (1996). n vertical lines are sampled at regular intervals from within a region-of–interest in the ultrasound image. By sampling the image sparsely, instead of looking at every column, we improve the efficiency of edge detection without significantly degrading the performance. The intensities $I(y)$ along each line are smoothed by convolution with a 1-D Gaussian function:

$$s(y) = g_\sigma(y) * I(y), \tag{11}$$
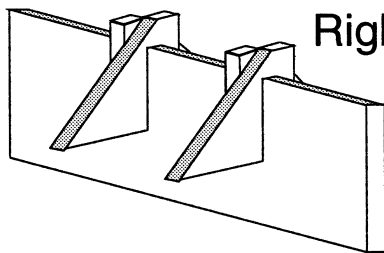

Left wedges

Right wedges

Fig. 13. Wedges to aid Cambridge phantom alignment. Four wedges are slipped over the Cambridge phantom's bar. The probe is correctly mounted when symmetrical images of the wedges are obtained at all scanning depths.
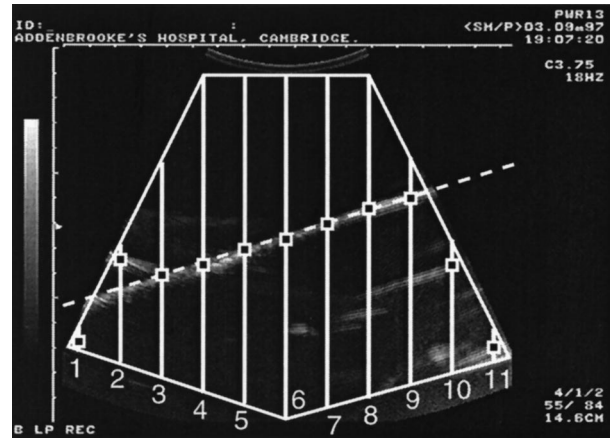

Fig. 14. Edge detection and RANSAC line fitting. Vertical lines are sampled within a user-specified region-of–interest in the ultrasound image. For each vertical line, a likely edge point is marked at the location nearest the top of the image where the derivative of the smoothed signal exceeds some threshold. Many, but not all, of these points will lie on the echo from the plane. The line through points 3 and 9 has a consensus set of size 7 because it is consistent with points 3–9. The maximum size of this line's consensus set is 8 because points 1, 10 and 11 could not possibly lie on the line. The line would be accepted if $p < 7/8$.

where

$$g_\sigma(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right).$$

The smoothed intensities $s(y)$ are then differentiated. The derivative theorem of convolution allows us to calculate $s'(y)$ from $I(y)$ using just one convolution:

$$s'(y) = g'_\sigma(y) * I(y). \tag{12}$$

Peaks in $s'(y)$ correspond to prominent dark-to–light transitions in the intensities. Because we are looking for the top of the first bright bar in the image, we search for the first maximum of $s'(y)$ that exceeds some threshold T. If no maximum of $s'(y)$ exceeds T, we settle for the biggest local maximum of $s'(y)$. Repeating this process for each of the sampled vertical lines, we obtain a list of points (x_i, y_i), $i \in \{1 \dots n\}$, corresponding to the first strong dark-to–light edge on each line (see Fig. 14).

    In the absence of noise, these points would lie on a straight line corresponding to the strong echo from the plane. However, given the quality of ultrasound images, it is likely that some of the points will be located some distance from the plane's actual position. The presence of outliers means that a least-squares fit of a straight line to the points (x_i, y_i) is inappropriate. Instead, RANSAC (random sample consensus) (Fischler and Bolles 1981) is used to fit the line to the data.

    RANSAC works by fitting a straight line through every pair of points (x_j, y_j) and (x_k, y_k) in the set (x_i, y_i), $i \in \{1 \dots n\}$. This generates $^nC_2$ lines. Each line is scored according to the size of its consensus set: a point (x_i, y_i) contributes to the consensus set if it is no more than a distance d above or below the line. The line with the largest consensus set is retained as the most likely position of the plane in the ultrasound image. Note how one or two outliers in the set (x_i, y_i) will have little effect on the position of the fitted line.

    Finally, a check is made that the size of the consensus set is

plausible before the line is entered into the calibration data. If the edge cannot be clearly detected, then it is better to discard that frame than to use unreliable data for calibration. The line is considered accurate if the size of the consensus set is at least a fraction p of the maximum consensus m. Note that m does not necessarily equal n: if the region-of–interest (ROI) is not rectangular, then some of the vertical scans may not intersect the fitted line[2] (see Fig. 14). As a final step, the equation of the line is refined using a least-squares fit to all points in the consensus set.

The algorithm is not particularly sensitive to the parameters n, σ, T, d and p. These are set by the user through a simple graphical user interface. The user can visibly check the performance of the algorithm on a few randomly selected frames (and adjust the parameters if necessary) before running the algorithm on the entire data set. The graphical interface also allows the user to draw the ROI over that portion of the image that contains ultrasound data. The entire process of setting the ROI, checking the parameters and extracting the lines typically takes less than 1 minute for a data set containing several hundred frames.

Numerical issues

The accuracy of calibration depends, to some extent, on the numerical methods employed. In the literature of robot kinematic calibration (a problem very similar to freehand ultrasound calibration), there are guidelines for establishing a well-conditioned problem (Hollerbach and Wampler 1996). The two key issues are rank determination and scaling.

To ensure that all the parameters of ϕ are identifiable, the Jacobian matrix $\mathbf{J}$ is evaluated (near the solution) to check that it is of full rank. To do this, a singular value decomposition (SVD) is performed:

$$\mathbf{J} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \qquad (13)$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Sigma} = \mathrm{diag}(\mu_1, \ldots, \mu_q)$ is the matrix of ordered singular values. Here, $q = 11$ for the cross-wire, single-wall and Cambridge phantoms, and $q = 14$ for the three-wire phantom.

The smallest singular value μ_q will be near zero if there is an unidentifiable parameter. If all μ are nonzero, then the Jacobian is said to have full rank. A convenient measure of identifiability is the condition number κ, which is simply the ratio of the largest to smallest singular values. A large condition number indicates that the problem is ill-conditioned, and at least one parameter cannot be accurately identified.

The condition number may be large for one of three main reasons. The first is that a parameter may not be intrinsically identifiable using the particular calibration technique. For example, when using the single-wall or Cambridge phantoms, three parameters of $^C\mathbf{T}_T$ are not identifiable because they do not affect the height or orientation of the plane. Removing these unidentifiable parameters from ϕ reduces κ. The second reason is that the motion of the probe may not have covered all the necessary degrees of freedom to identify the parameters in ϕ. Extending the range of scanning directions will help to reduce κ. The third reason concerns degeneracy in the fixed-axis representation of rotation we have adopted. If $\beta = \pm 90°$, then there exist multiple triples (α, β, γ) that produce the same transformation $^R\mathbf{T}_P$ (Craig 1989). In effect, one of the angles becomes unidentifiable. This situation can be avoided by mounting the position sensor at a skew angle to the ultrasound probe, such that β is not close to $90°$.[3]

The second important issue is scaling. Task-variable scaling is the process of scaling the errors produced by eqn (4) so they contribute equally to the optimisation. Because each of the equations has the same

Table 7. Condition numbers for calibration with the Cambridge phantom.

	κ
14 parameters	6.97×10^{18} ($\rightarrow \infty$)
11 identifiable parameters	4867
11 scaled parameters	87.5

The condition number approaches infinity when ϕ includes the full set of 14 parameters. By removing the three unidentifiable parameters x, y, and α (see Table 1), the condition number is considerably reduced. Column scaling of the remaining 11 parameters reduces the condition number to below 100.

units and describes a deviation from zero, they are directly comparable. However, this is not true for the parameters to be identified. The different parameters have different units (mm, radians, mm/pixel) and different effects on the residual of eqn (4). This issue can be addressed by parameter scaling, which involves modifying eqn (5) as follows:

$$\Delta\mathbf{f} = \mathbf{J}\Delta\phi = \mathbf{J}(\mathbf{H}\mathbf{H}^{-1})\Delta\phi = (\mathbf{J}\mathbf{H})(\mathbf{H}^{-1}\Delta\phi). \qquad (14)$$

$\mathbf{H}$ is a scaling matrix chosen such that $\mathbf{J}\mathbf{H}$ is better conditioned than $\mathbf{J}$ alone. Equation (6) is then evaluated using $\mathbf{J}\mathbf{H}$ and $\mathbf{H}^{-1}\phi$ in place of $\mathbf{J}$ and ϕ. After convergence, ϕ is determined by unscaling the solution $\mathbf{H}^{-1}\phi$.

The simplest way of selecting $\mathbf{H}$ is by column scaling (Lawson and Hanson 1974), whereby $\mathbf{H} = \mathrm{diag}(h_1, \ldots, h_q)$ and:

$$h_i = \begin{cases} \|j_i\|^{-1} & \text{if } \|j_i\| \neq 0 \\ 1 & \text{if } \|j_i\| = 0 \end{cases}. \qquad (15)$$

j_i is the ith column of $\mathbf{J}$, evaluated at a reasonable guess of the solution. This scaling ensures that each parameter has the same effect on $\Delta\mathbf{f}$.

The combination of removing unidentifiable parameters and scaling the remaining parameters should reduce κ to an acceptable level. For robot kinematic calibration, a common rule of thumb is that the condition number should be less than 100 for accurate parameter identification (Schröer 1993). Table 7 shows how to establish a well-conditioned calibration problem.

There are, however, several different values of ϕ that can achieve the same, well-conditioned global minimum. We call these "mirror" solutions. This is not the same as the singularity problem discussed earlier, where an infinite number of solutions exist and the problem is ill-conditioned. A mirror solution can be constructed, for example, by adding a multiple of 2π to any of the angles. Even though mirror solutions represent the same calibration, they need to be expressed in a canonical form to allow straightforward comparison of results.[4] We require that:
1. s_x and s_y are positive;
2. α and γ are in the range $\pm \pi$; and
3. β is in the range $\pm \pi/2$.
These constraints can be enforced using the following procedure:
1. Limit all the angles to $\pm \pi$ by adding or subtracting 2π;
2. If β is outside $\pm \pi/2$, add (or subtract) π to correct it, and add π to both α and γ;
3. If $s_y < 0$ then $\gamma \leftarrow \gamma + \pi$ and $s_y \leftarrow -s_y$;
4. If $s_x < 0$ then $\alpha \leftarrow \alpha + \pi$, $\beta \leftarrow -\beta$, $\gamma \leftarrow \pi - \gamma$ and $s_x \leftarrow -s_x$;
5. Check that α and γ are still within $\pm \pi$: if not, repeat Step 1.
Note the order of the corrections for negative scales. If both scales require correction, the composite transformation for γ must be $\gamma \leftarrow -\gamma$.