

● *Original Contribution***RAPID, EASY AND RELIABLE CALIBRATION FOR FREEHAND 3D ULTRASOUND**

PO-WEI HSU, RICHARD W. PRAGER, ANDREW H. GEE, and GRAHAM M. TREECE

Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

(Received 21 September 2005; revised 16 February 2006; in final form 23 February 2006)

Abstract—This paper presents improvements to the plane-based technique for calibrating freehand 3D ultrasound systems. The improvements are designed to make it easier for inexperienced users to perform plane-based calibration and to know that they have got a reliable result. In particular, we enable the calibration to be performed using water at room temperature while producing a result that is valid for average soft tissue and we show how it is possible to provide feedback on the reliability of the calibration using a metric based on the curvature of the calibration criterion function. We present comprehensive results showing that these innovations improve the precision of the calibration and offer useful feedback to the user. (E-mail: pwh24@cam.ac.uk) © 2006 World Federation for Ultrasound in Medicine & Biology.

Key Words: freehand three-dimensional ultrasound, calibration.

INTRODUCTION

Freehand three-dimensional (3D) ultrasound (Fenster et al., 2001) is a technique for acquiring ultrasonic data of a 3D volume by measuring the trajectory of the ultrasound probe using a position sensor. It has applications in clinical volume measurement, analysis of complex geometry, surgery planning and radiotherapy planning (Gee et al., 2003). As the probe is swept across the volume, the sensor records the position and orientation of the probe with respect to an external world coordinate system. The recorded position, however, is the position of the sensor, rather than the scan plane, as shown in Fig. 1. We therefore need to determine the scales in the ultrasound images (B-scans) [For readability purposes, we will use the words “B-scan”, “scan” and “image” interchangeably in this paper] and the rigid-body transformation from the electrical centre of the position sensor to the corner of the ultrasound scan plane. This is done through a process called probe calibration (Mercier et al., 2005). Points in the B-scan are first transformed to the sensor’s coordinate system by the parameters found by calibration and then to the world space by the sensor’s readings. It is crucial that probe calibration be as accurate

as possible, as any errors will propagate through to subsequent stages of image analysis and measurement.

The usual technique for 3D ultrasound calibration is to perform a series of scans of an object with known dimensions (a phantom). The simplest such object is a point fiducial formed by cross wires (Detmer et al., 1994; Huang et al., 2005) or a small spherical ball (State et al., 1994; Amin et al., 2001) in a water bath. This point is scanned from different directions and segmented in each B-scan. For a given set of calibration parameters, each point in the B-scan can be mapped to the world coordinate system. Since the same point is being scanned, they should be all mapped to the same point in space. This sets up an over-determined system of constraints that can be solved using optimisation techniques.

There are several difficulties in following the point phantom approach. First of all, each point in the B-scan needs to be manually or semiautomatically segmented. This makes the calibration process long and tiresome. Secondly, due to the finite thickness of the ultrasound beam, it is very difficult to align the probe directly at the centre of the point, as the point fiducial may appear in the B-scan even when the probe is not directly pointing at it.

Muratore and Galloway (2001) used a 3D localizer, which is essentially an additional position sensor attached to the point phantom. In general, this requires an additional calibration for the localizer itself, unless it is precisely designed and manufactured. In any case, this

Address Correspondence to: Mr. Po-Wei Hsu, Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK. Tel: +44-1223-332754; Fax: +44-1223-332662; E-mail: pwh24@cam.ac.uk

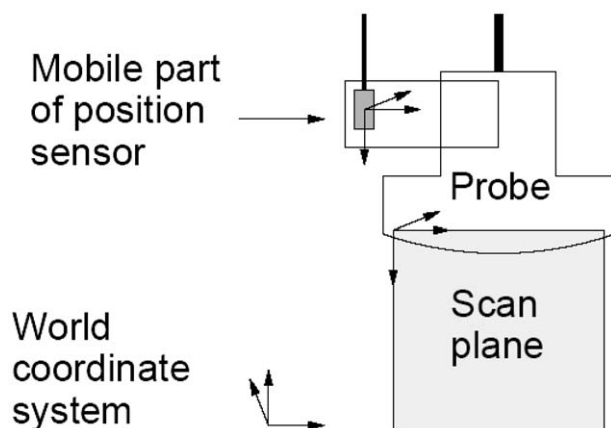


Fig. 1. The coordinates associated with the scan plane and mobile part of the position sensor.

means that the position of the phantom is known in space during calibration. This greatly simplifies the optimisation and allows for closed-form solutions.

A three-dimensional localizer is in general able to locate any point in space, and has been used to locate points on other types of phantoms. Sato et al. (1998) used a two-dimensional phantom comprising three pointed edges. These three points are located in space with a 3D localizer. The planar phantom is then aligned with the B-scan and the three points are manually segmented. Since these three points are known in space, the calibration parameters can be solved algebraically in a closed form. This approach has the advantage that only one frame is needed for calibration, but it is nevertheless extremely difficult to align the plane phantom, with the thick beam compounding the difficulty.

The alignment problem is solved by the design of Z-phantoms (Comeau et al., 1998; Pagoulatos et al., 2001; Lindseth et al., 2003), where wires are connected in a “Z” shape. The intersection of the wires with the B-scan forms a virtual plane phantom, whose location can be deduced from the imaged points. This type of calibration is fast, since only one image is required and it does not require accurate alignment with the plane. However, the thick beam is still a problem. The wires do not show up as clear dots, but as smeared ellipses. Locating the centres of such ellipses accurately is difficult, with the noisy ultrasound background compounding the problem. Nevertheless, automatic segmentation is possible with a predefined search region (Lindseth et al., 2003).

Gee et al. (2005) designed a high definition mechanical device, where the plane phantom is adjusted with micrometers and is therefore able to be aligned very accurately. The alignment is also aided by the use of wedges, which show up as bright patches in the B-scans, and so a slight misalignment of the phantom can be

easily seen from the B-scans. The wedges also serve as the fiducial points on the planar phantom, which are semiautomatically segmented. This mechanical device has been shown to be the highest definition system in the current literature.

In all the aforementioned approaches, the calibration requires a purpose-built phantom, parts of which need to be manufactured very accurately. Another calibration technique is the single wall approach (Prager et al., 1998; Varandas et al., 2004; Rousseau et al., 2005). All the hardware that is required is a flat plane immersed in a water tank. The flat plane could be as simple as the bottom of the water tank. The Cambridge phantom (Prager et al., 1998) may be used as a virtual plane to avoid the beam thickness problem and to produce strong echoes even when the plane is scanned from an oblique angle. As the plane is scanned from different angles, the images of the plane appear as clear, straight lines in the B-scans. This allows the possibility of automatic segmentation, subject to a few threshold settings prior to calibration. Prager et al. (1998) used the RANSAC line detection algorithm (Fischler and Bolles, 1981) for its robustness against outliers. Rousseau et al. (2005) used Hough transforms (Hough, 1959) to extract the lines and reject outliers from each B-scan. They went further to assume that the line varied smoothly between neighbouring B-scans in order to reject any falsely detected lines. For a given set of calibration parameters, points on these lines can be transformed to the position sensor’s coordinate system and then to the world coordinate system by using the position sensor’s readings. If the calibration is correct, these points will also lie on a plane in world space. This sets up a list of constraints involving 11 parameters. Six of these define the rigid-body transformation from the mobile part of the position sensor (attached to the probe) to the corner of the B-scan image. Three further parameters define the location of the phantom plane in 3D space. These three are only required internally for the process of calibration itself; once calibration is complete, they are discarded. Two final parameters define the horizontal and vertical scales in the B-scan image. An iterative optimisation algorithm is then utilized to find the best fit plane by adjusting the calibration parameters, see Fig. 2.

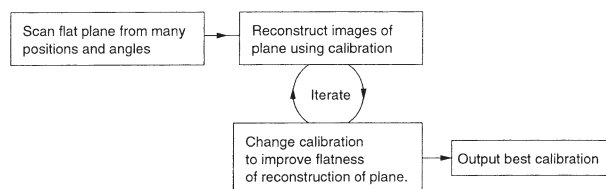


Fig. 2. Block diagram of plane-based calibration.

This paper describes some refinements to make the process of single-wall calibration easier and more reliable for inexperienced users to perform. There are two main problems:

(1) The speed of sound in water at room temperature is slower than its speed in average soft tissue. The sound speed in water at 20° C is only 1482 m/s (Bilaniuk and Wong, 1993), while sound travels at 1540 m/s in average soft tissue. For accurate calibration, the water bath should be at 48° C, which gives a sound speed of 1540 m/s. This approach was adopted by Bector *et al.* (2003) and Treece *et al.* (2003), but the water temperature is difficult to maintain and uncomfortable for the operator. Another approach is to add glycerol (Gobbi *et al.*, 2000) or ethanol (Martin and Spinks, 2001) to the water solution. This means an accurate amount of a separately acquired solution needs to be added to the readily available water, which complicates the calibration procedure. Lindseth *et al.* (2003) modified the sound speed assumed by the ultrasound machine directly to correspond to the speed in cold water. This simple and easy approach may require a special arrangement with the manufacturer. Trobaugh *et al.* (1994) and Pagoulatos *et al.* (2001) made provisions by postprocessing the B-scan images. Although the correction using this technique may be simple for linear probes, it is more complicated for curvilinear probes. Nevertheless, we have used the image processing technique in this paper, since glycerol may not be readily available in every situation. We aim to allow the user to perform calibration with as few specialised equipments as possible. This technique may be easier to use for novice users and still allows the user to use a glycerol/ethanol solution by pretending that the temperature is 48° C. If the calibration is performed in cold water without any adjustments, the sound speed difference would cause the phantom to appear shallower in the B-scans by 3.7% during calibration.

(2) There is currently no feedback to the user to indicate when the plane has been scanned from a sufficiently diverse set of positions and angles to produce a well-constrained solution to the resulting nonlinear equations. The residual error from the optimisation is not an accurate measure of calibration quality, since it is possible to perform an incorrect calibration with a low residual error. An extreme example would be to scan the plane repeatedly from the same position. In this case, the residual error would be low and the user would have no idea that the calibration was unreliable. In our experience supporting medical physicists and clinicians using our single wall calibration system (Treece *et al.*, 2003), many inexperienced users have recorded under-constrained motion sequences, leading to incorrect calibrations. In this paper, we show how it is possible to provide feedback to the user on the reliability of the calibration.

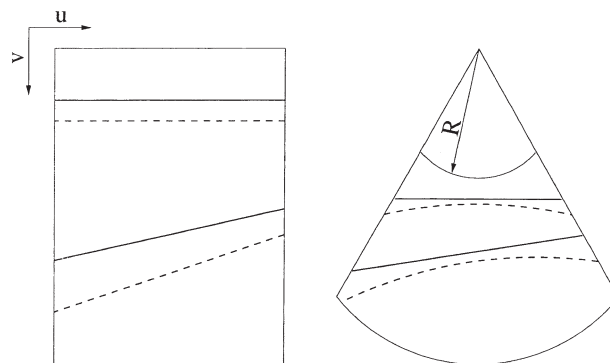


Fig. 3. The dotted lines shows the distorted the plane due to slower sound speed in cold water. The solid lines show the corrected plane. The diagram is exaggerated to emphasise the distortion.

IMPROVING THE RELIABILITY OF SINGLE WALL CALIBRATIONS

In this section, we present the modifications which we have made to current single-wall calibration systems used by Prager *et al.* (1998) and Treece *et al.* (2003). We performed and analysed a number of calibrations in order to measure the reliability of a calibration using a metric based on the curvature of the criterion function. These calibrations were used to compare with the results from Treece *et al.* (2003), in order to verify the effectiveness of the modifications which we made to the single wall calibration protocol.

Improvement to single wall calibrations

We describe the two modifications which we have made to improve current plane-based calibration systems. The first modification is to account for the difference between the speed of sound in water and in soft tissue. The second is to determine the B-scan scales explicitly.

Speed of sound

The speed of sound problem is fixed by allowing the calibration to be performed at room temperature, but requiring the user to provide a measurement of the water temperature. The ultrasound images can then be corrected using one of the well-known polynomial models for the speed of ultrasound in water as a function of temperature (Bilaniuk and Wong, 1993).

Since sound travels slower in cold water than in average soft tissue, the image of the plane (dotted lines) appears to be further away than its actual position (solid lines), as shown in Fig. 3. Corrections can therefore be made by moving the image of the plane toward the face of the probe.

We used the same edge detection algorithm as was

used by Prager et al. (1998). Vertical samples are taken within the user defined region in the ultrasound image. A likely edge point is where the derivative of the smoothed signal first exceeds some threshold. The edge can then be found by applying the RANSAC line detection algorithm to these likely edge points (Fischler and Bolles, 1981). For a linear probe, the correction is a translation of each likely edge point toward the probe face; the axial coordinates v of the points are multiplied by the temperature correction factor $t = \frac{\text{speed in cold water}}{\text{speed in average soft tissue}}$.

In the scenario of a curvilinear probe, the plane appears as a curve rather than as a line. We therefore first detect the probe shape automatically (Treece et al., 2002). The probe centre is then the intersection of the left and right edges. We further assume that the probe face is part of a circular arc with radius R centred at the probe centre. We move each point on the curve (shown as a dotted line in Fig. 3) towards the probe face, in the direction of the probe centre, by the temperature correction factor. The shifted points will lie on a straight line and can be detected using the RANSAC line detection algorithm (Fischler and Bolles, 1981). Since this correction is mathematically sophisticated and extra steps need to be followed by the user to detect the probe shape reliably, an approximate correction is considered for common abdominal curvilinear probes. Please refer to Appendix A for a detailed discussion.

Explicit scale estimation

We increase the robustness of the calibration process by reducing the number of parameters involved. We can establish the image scales at the start of the procedure by using a separate protocol and, hence, remove the horizontal and vertical image scales from the optimisation. The result is that the nonlinear optimisation only has to determine nine rather than 11 parameters and this makes the system of equations easier to constrain.

The image scales are determined by using the distance measurement tool that is available on most (if not all) clinical ultrasound machines. This tool provides the distance, in centimetres, between two points marked in the B-scan image. These points will appear in the grabbed B-scan on the computer. A simple graphical user interface has been implemented to enable the user to indicate these points on the captured B-scan and to enter their separation. Since most ultrasound machines provide square pixels to a very good approximation, we have further simplified this procedure, with little loss of accuracy, by assuming that the horizontal and vertical image scales are the same. The scale is therefore estimated by

$\frac{\text{distance (cm)}}{\sqrt{(\Delta u)^2 + (\Delta v)^2}}$, where Δu and Δv are the horizontal and vertical distances, in pixels, between the two marked points on the B-scan.

Quantification of calibration quality

In order to measure the calibration quality and to provide feedback to the user, we performed calibrations using three probes on two different ultrasound machines. The first is the Diasus [Dynamic Imaging Ltd., <http://www.dynamicimaging.co.uk>] 5–10 MHz linear-array probe. The analog radio-frequency (RF) ultrasound data, after receive focusing and time-gain compensation but before log-compression and envelope detection, was digitised by using a Gage CompuScope 14100 PCI 14-bit analog-to-digital converter [Gage Applied Technologies Inc., <http://www.gage-applied.com>], and transferred at 10 frames per second to a Pentium(R) 4 2.80 GHz PC running Microsoft Windows XP. Each frame consists of 127 vectors, each with 3827 data points sampled at 66.67 MHz synchronised with the clock of the ultrasound machine. The RF data were converted to an analytic signal using matching Hilbert filters with a unity-gain and 5–10 MHz pass-band. The amplitude of the band passed signal was suitably log-compressed so that the resulting images resemble those on the ultrasound machine. Finally, the data were linearly interpolated to produce the image displayed on the computer based on the image scale, which we have set at 0.01 cm/pixel. We also performed calibrations with the Toshiba [Toshiba Corporation, <http://www.toshiba.com>] model SSA-270A/HG 7.5 MHz linear probe and the 3.75 MHz curvilinear probe. The B-scan images were digitised using a Brooktree [Conexant Systems Inc., <http://www.conexant.com>] BT878 frame-grabber card and transferred to a 3.0 GHz PC running Linux. Both probes were tracked using an AdapTrax [Traxtal Technologies, <http://www.traxtal.com>] infrared LED target for the Polaris [Northern Digital Inc., <http://www.ndigital.com>] optical tracking system.

We performed four calibrations by capturing 30, 60, 90 and 120 images of the plane, using the Cambridge phantom. The calibration consisting of 60 images was formed by capturing 30 images of the plane in addition to the 30 images from the previous calibration and, similarly, for the calibrations with 90 and 120 frames. Every calibration was repeated 10 times. It is important for plane-based calibration that the plane is scanned in a nondegenerate sequence to constrain the plane in space (Prager et al., 1998; Treece et al., 2003). These calibrations all included the required motion sequences, namely vertical movements, side to side rotations, front to back rotations, translation in the x direction, translation in the y direction and rotations about the z -axis, where x and y

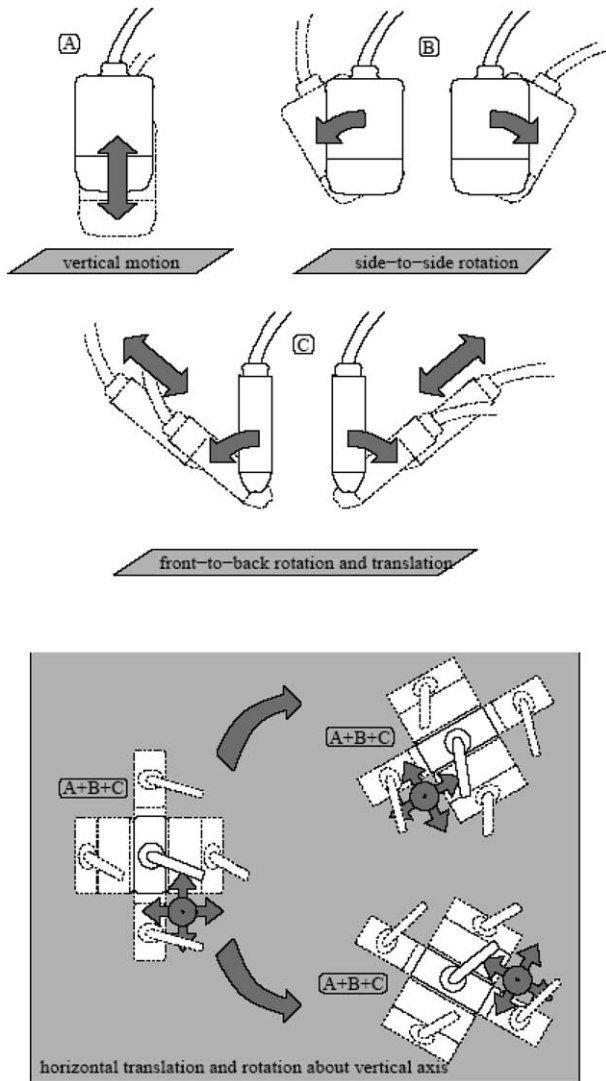


Fig. 4. The sequence of probe movement required during calibration to constrain the 11 calibration parameters (before scales estimation). The diagram is taken from Treece *et al.* (2003).

are the two axes in the plane phantom, aligned with the sides of the water tank and z is the corresponding axis perpendicularly out of the plane, as shown in Fig. 4. The probes and depth settings, as well as the cropped B-scan sizes in pixels used in the experiments, are listed in Table

1. One focus at approximately half the depth setting was used in each calibration. The scales in the Diasus calibrations were set manually to 0.01 cm/pixel, since this is the scale at which the RF data, after log-compression, envelope detection and band-pass filtration, were displayed. The distance measurement tool was utilized to estimate the scales in the Toshiba calibrations. Both the temperature and scale estimations are highly repeatable and the same readings were obtained in every measurement. Hence, there is no variation in the scales in each calibration. In all, a total of: four probes \times four numbers of plane images \times 10 repetitions = 160 calibrations were acquired.

The quality of each calibration was assessed by measuring its precision and accuracy (Treece *et al.*, 2003). Precision is measured by the consistency of the locations of the four corners and the middle of the image when located in 3D space using the calibration parameters. For each of these five points, we measured its mean location $\bar{x}_i = \frac{1}{10} \sum_{j=1}^{10} x_i^j$ in 3D space, where x_i^j is the 3D location of point i mapped using the j -th calibration. The average magnitude of its deviation from this point was then computed by: $\frac{1}{10} \sum_{j=1}^{10} |x_i^j - \bar{x}_i|$. This measure was called *mean 3D error* by Treece *et al.* (2003). These five errors were then averaged ($\frac{1}{5} \sum_{i=1}^5$ mean 3D error of point i). Figure 5 shows that the averaged error as the number of scans is increased in our calibrations. The decreasing graphs show that more consistent calibrations were achieved at a higher number of scans. The linear behaviour of the Diasus probe at 6 cm depth may appear to be peculiar at first glance, but in fact it saturates between 120 and 150 scans. This was confirmed by an additional experiment and further data analysis. This linear behaviour is also evident in other probes, most noticeably in the Diasus probe at 3 cm depth before it saturates after 90 scans.

The precision measurements indicate the variation in the image position due to spatial calibration errors. The repeatability of the calibrations is not an indication of accuracy, as the calibrations might be highly repeatable but biased.

Table 1. Probes and other settings used in the calibrations.

Machine	Probe type	Frequency	Depth	B-scan	Scales
Diasus	Linear	5–10 MHz	3 cm	360 \times 300	Preset
Diasus	Linear	5–10 MHz	6 cm	360 \times 600	Preset
Toshiba	Linear	7.5 MHz	6 cm	222 \times 416	Distance measurement
Toshiba	Curvilinear	3.75 MHz	6 cm	388 \times 380	Distance measurement

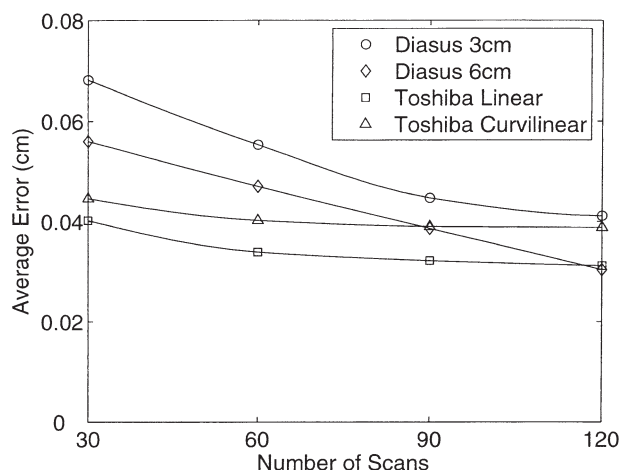


Fig. 5. The variation of the four corners and the centre of the cropped B-scan due to spatial calibration for the different types of probes.

We assess the accuracy of our calibrations by scanning a tissue-mimicking phantom consisting of coplanar spheres with 2 mm diameter (Kofler and Madsen, 2001). These spheres lie on a grid with known dimensions (Appendix B) and they show up as dark circles in the B-scans. We scanned the phantom with the Diasus probe at 3 cm and 6 cm depths, using the three patterns in Fig. 6. These patterns were used to assess system accuracy and were not necessary during clinical use of the free-hand 3D ultrasound system. In order to maximize the visibility of the spheres, six focal points were used in the middle region of the image. The images were recorded at approximately 7–15 frames per second. Although many foci were used during image acquisition, some of the spheres away from the focal points could not be seen clearly and segmented properly, so they were not included in the accuracy assessment. The phantom was scanned freehand, but as slowly as possible so that each sphere appeared in a number of consecutive images. A typical recording consisted of approximately 400 scans down the 10 cm length of the scanning window. The images of the spheres were then semiautomatically segmented, reconstructed in space using the parameters from our calibrations and fitted to our predefined grid of spheres (Treece et al., 2003). We measure the error as the 3D distances between the centres of the reconstructed spheres and their true location in world space.

Figure 7 shows the orthographic projection of the sphere centres' distribution. The small grey circles are the actual positions and size of the spheres forming the predefined grid. In order to make the errors more visible, the small grey circles are magnified by a factor of four and shown as larger black circles. The crosses show the centre of each reconstructed sphere, with the error mag-

nified by a factor of four. The diagrams in the left column show the reconstruction errors for the Diasus probe at 3 cm depth; the errors at 6 cm depth are shown in the right column.

Figure 8 shows the average error as a function of the number of scans. The graphs are relatively flat. This is most likely because the dominant errors are not caused by spatial calibration. Possible sources include the optical tracker, temporal misalignments, image capture and segmentation of the spheres. Nevertheless, from the low errors and the even distribution of located spheres in Fig. 7, we can deduce that our calibrations are unbiased.

In this paper, we will use precision as the measure for calibration quality. This is because precision is more sensitive to calibration errors than accuracy, since precision measures the variations caused by spatial calibration only and not a combination of calibration and other sources of errors. We have also verified that this choice is viable by the low accuracy errors in each calibration, confirming that the calibrations are unbiased and accurate.

Curvature of the criterion function at the solution

In each calibration, we obtain a set of constraints involving nine variables, since the B-scan scales have been determined explicitly. The values for these variables are solved using a nonlinear optimisation algorithm and we have ensured that the correct solution is obtained (Appendix C). We wish to measure how well the solution is constrained. As the nine parameters are similarly scaled, with distances in centimetres and angles in radians, we can get an indication of an under-constrained solution using the minimum curvature of the criterion function at the solution. If the minimum curvature is high, then the solution is well-constrained. If the minimum curvature is low, then one or more parameters will be under-constrained. We can thus spot poor solutions and advise the user to scan the plane from some more angles to introduce further constraints.

We wish to use the minimum curvature at the solution as a measure of the quality of the solution. We therefore measured the minimum curvature of our cali-

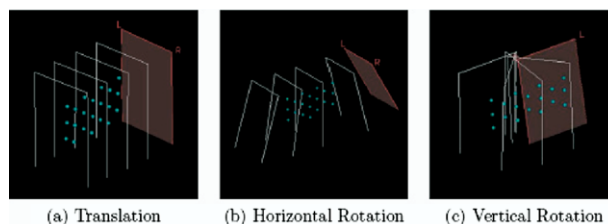


Fig. 6. Scanning patterns used when scanning the phantom. Approximately 400 scans were recorded in each case.

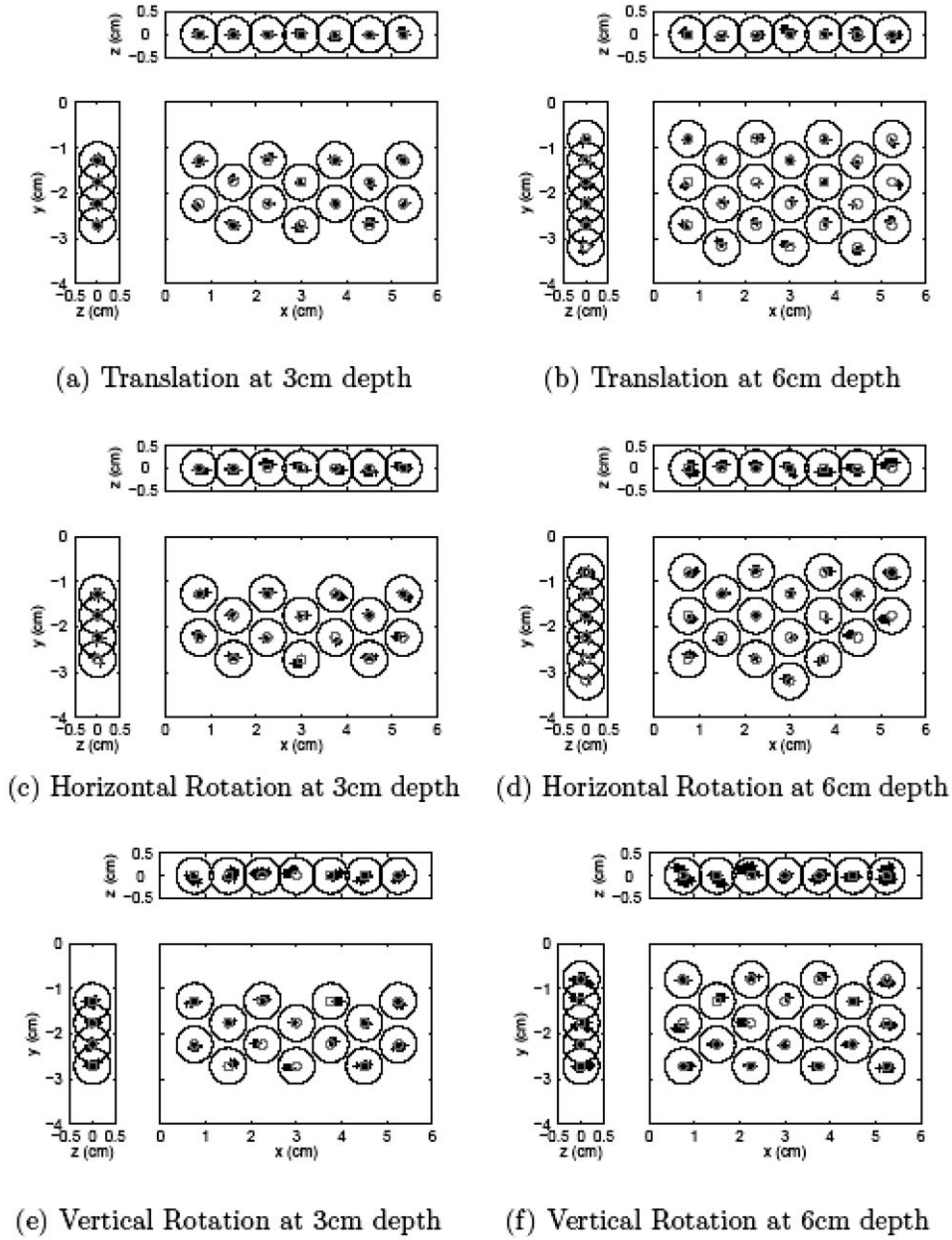


Fig. 7. Orthographic projection of the phantom and the reconstructed spheres. The small grey circles show the true size and location of the spheres in space. The larger black circles show these spheres magnified by a factor of four. The distribution of the reconstructed spheres is shown as crosses, with the error magnified by a factor of four. The consistent magnification means that the crosses and the large black circles are directly comparable.

brations by computing the Hessian, the matrix consisting of all the second derivatives of the objective function. The eigenvalues and eigenvectors of the Hessian are evaluated to give the magnitudes and directions of greatest and least curvature. In all four probes, the eigenvector corresponding to the minimum eigenvalue for each calibration is found to be in the translation direction that defines the phantom in world space. This eigenvalue is expected to be lower than other eigenvalues, because this

translation is typically an order of magnitude larger than the other calibration parameters. This means that the error in this direction is also expected to be an order of magnitude higher than other parameters. This does not bother us, as the three parameters defining the phantom in world space are subsequently discarded and are not used to judge how well a calibration is performed. We therefore use the second minimum eigenvalue, whose corresponding eigenvector is in a direction relating to a

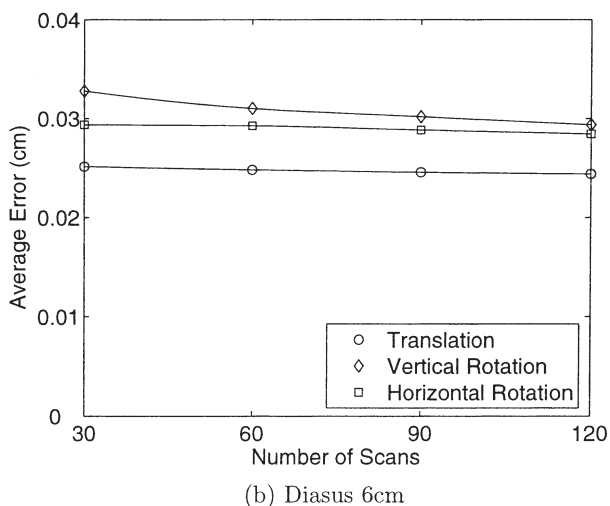
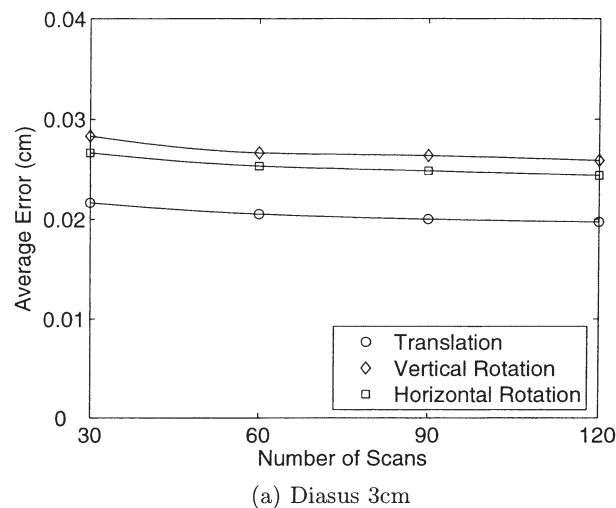


Fig. 8. The variation of the system error in terms of distance measurement as the number of scans is increased in the spatial calibrations.

combination of the six parameters that define the transformation from the mobile part of the position sensor to the corner of the B-scan image. For brevity, we will use the words “eigenvalue” or “minimum eigenvalue” in this paper while we are in fact referring to the second minimum eigenvalue (minimum eigenvalue of interest).

Figure 9 shows the graph of the second minimum eigenvalue (minimum eigenvalue of interest) of the Hessian at the calibration solution as the number of scans is increased. A clear linear relationship can be seen. This shows that the minimum curvature at the solution is related to the number of scans of the plane used in the calibration.

We have shown that the error is related to the number of scans in the calibration in Fig. 5. Hence, the number of scans is related to both error and eigenvalue. This enables us to plot the error against the eigenvalue,

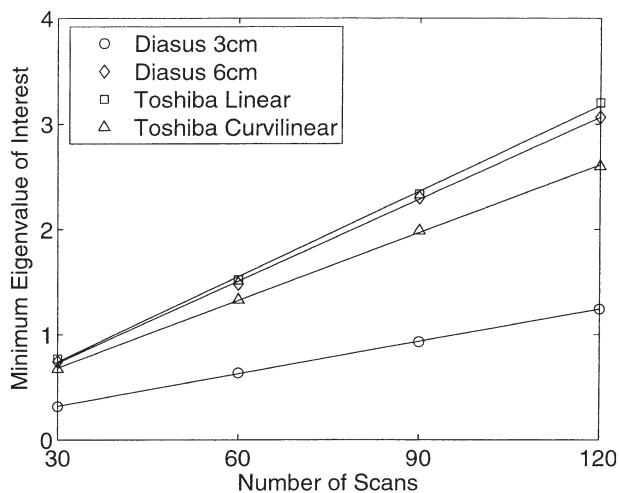


Fig. 9. The graphs shows the second minimum eigenvalue (minimum eigenvalue of interest) of the Hessian calculated at the solution of the objective function as the number of scans is increased.

as shown in Fig. 10. This shows that we can use the size of the eigenvalue as an indicator for the quality of the solution. We can therefore advise the user whether the calibration is reliable, since a high eigenvalue will ensure a well-constrained calibration.

Incomplete scanning patterns

We have shown, in the previous section, that we can use the eigenvalue to determine whether the solution is well-constrained and, hence, the reliability of the calibration. This assumes that the user has performed all the required scanning patterns in Fig. 4. Inexperienced users often neglect some of these motions inadvertently. We

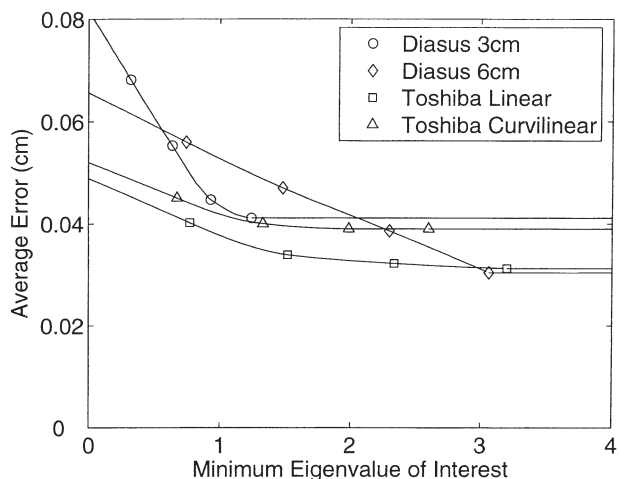


Fig. 10. The variation of the four corners and the centre of the B-scan as the calibration becomes more constrained.

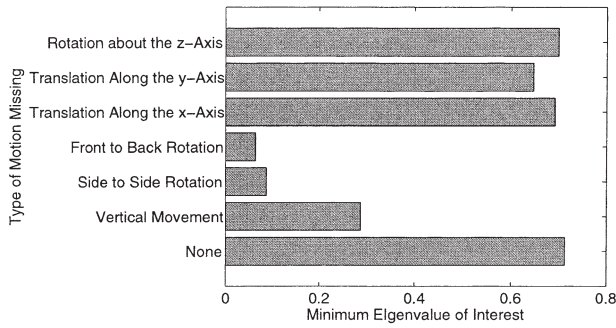


Fig. 11. The minimum eigenvalues for the calibrations where a particular motion is missing.

performed calibrations consisting of 30 images with one of the motions missing with the Diasus probe at 6 cm depth to investigate whether incomplete calibrations do result in an under-constrained solution, reflected by a low eigenvalue.

The minimum eigenvalues for these incomplete calibrations are shown in Fig. 11. In the two cases where side-to-side and front-to-back rotations are missing, the under-constrained optimisation resulting from incomplete scanning patterns is reflected by a near-zero minimum eigenvalue. For the case where vertical movements are missing, the minimum eigenvalue also dropped by more than a factor of two, indicating an insufficiently-constrained solution. In the other three cases, the eigenvalue did not drop as expected. This suggests that, although these motions are absent, the solution is well-constrained. We verify this by performing a series of calibrations without these motions and show that they are just as precise as calibrations with all motions exercised.

We performed an additional 10 calibrations with minimal rotations and translations in the aforementioned directions, but still containing the remaining three types of motions. Figure 12 shows the graph of average error against the number of scans. The graph for good calibrations where all motions are exercised using the same probe (Fig. 5) is duplicated here for comparison.

From the figure, it can be seen that the apparent paradoxical high eigenvalue for calibrations where rota-

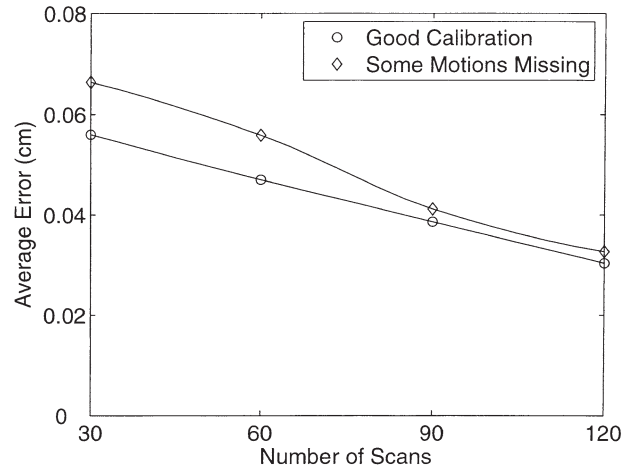


Fig. 12. Comparison between calibrations with minimal rotation about the z-axis and minimal translation in the xy -plane and calibrations where all six types of motion are exercised.

tions about the z -axis and translations in the xy -plane are missing is in fact correct. The average error is similar in magnitude from a low to a high number of scans. This means that these motions are not necessary to constrain the optimisation sufficiently for accurate calibration. This is because we have reduced the number of parameters to be optimised from 11 to nine, hence reducing the degrees of freedom which the calibration needs to constrain. This observation can be confirmed by simulating the calibration process mathematically. We have found that successful calibrations can be simulated, excluding the aforementioned motions.

Precision and accuracy

We verified the effectiveness of the improvements which we have made to the single wall calibration method by comparing our calibrations based on 120 images with the results from Treece *et al.* (2003), where the highest definition system using plane-based calibrations was reported. Figure 13 shows that the calibrations in this paper are more precise than those achieved by Treece *et al.* (2003). The error for the Diasus 3 cm probe

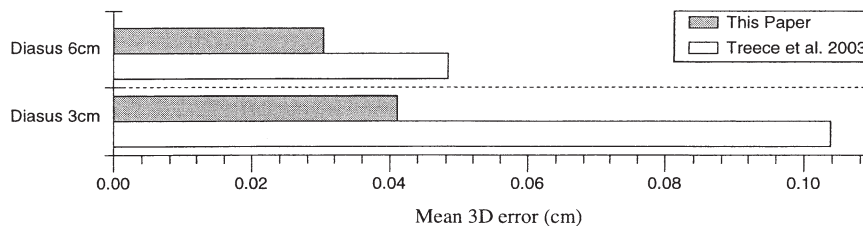


Fig. 13. Comparison of the errors due to spatial calibration alone achieved in this paper and by Treece *et al.* (2003). The errors shown are the mean 3D errors at the four corners and the centre of the B-scan.

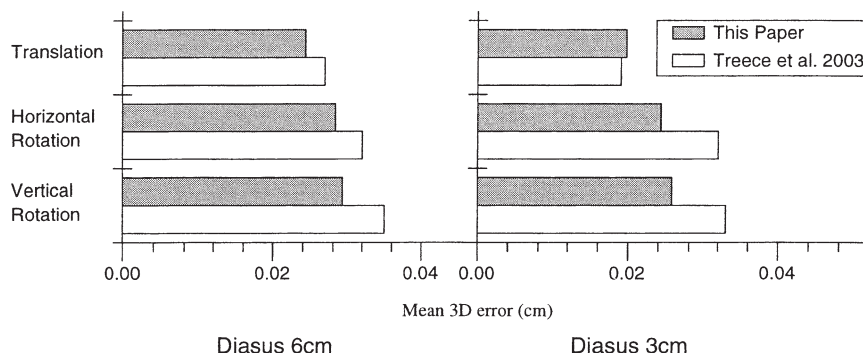


Fig. 14. Comparison of the errors from distance measurements of the whole freehand 3D ultrasound system achieved in this paper and in Treece et al. (2003). The errors shown are the mean 3D distance errors between the centres of the segmented spheres and their real locations in space.

has been reduced by more than a half and the error for the 6 cm depth setting has been reduced by more than a third.

The probe motion during calibration is restricted when the depth is set low at 3 cm. This can result in a weakly-constrained optimisation. By removing the scale factors from this optimisation, we reduce the dimensionality of the search and greatly improve the precision of the result.

The higher errors in the calibrations performed by Treece et al. (2003) are most likely caused by the variation in temperature during calibration. The hot 48° C water will cool down rapidly in a standard room environment. By correcting for the temperature difference, calibration in cold water is possible.

Figure 14 shows a comparison of our system accuracy with the results from Treece et al. (2003). It shows that our adjustments have improved the system accuracy slightly.

CONCLUSIONS

We have proposed some changes to the plane-based 3D ultrasound calibration protocol that increase its reliability and ease of use, while maintaining its state-of-the-art performance in terms of speed and accuracy. The compensation for the temperature of the water and the simple technique for explicit entry of scale values are both easy to use. In particular, fewer degrees of freedom need to be constrained by spatial calibration and so rotations about the z -axis and translations in the xy -plane are no longer needed to achieve a good calibration.

We have also produced a foolproof indicator of the calibration quality by calculating the eigenvalue of the Hessian at the solution of the criterion function. We have ensured that the global minimum is reached by the non-linear optimisation (Appendix C). Figure 10 shows the variation of the B-scan corners due to spatial calibration as a function of the minimum eigenvalue. The high

accuracy of the calibrations shown in Fig. 8 confirms that these calibrations are unbiased. Hence, we may use the eigenvalue as an indicator for the calibration quality. Figure 11 further confirms that, if a motion necessary for calibration is missing, which is a common scenario for inexperienced users, the incomplete calibration is indicated by a low eigenvalue. The eigenvalue criterion has been evaluated and shown to be effective on three different probes and at two different depth settings.

REFERENCES

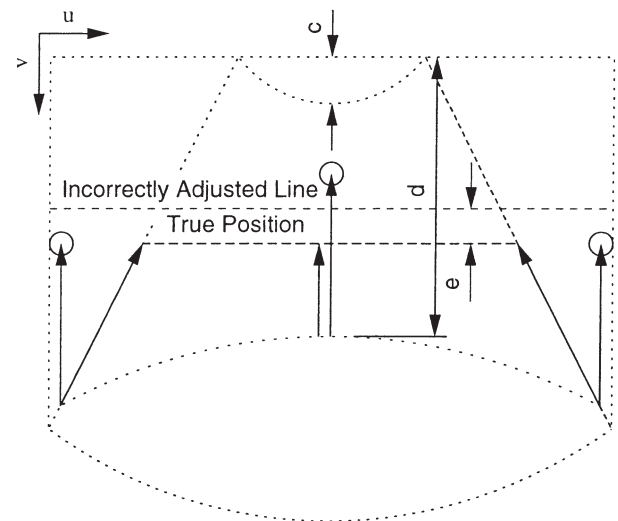
- Amin, D. V., Kanade, T., Jaramaz, B., et al. Calibration method for determining the physical location of the ultrasound image plane. In: Proceedings of the fourth International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science. Vol. 2208. Springer-Verlag, 2001, pp. 940–947.
- Bilaniuk, N., Wong, G. S. K.. Speed of sound in pure water as a function of temperature. The Journal of the Acoustical Society of America 1993 93 (3), 1609–1612.
- Boctor, E. M., Jain, A., Choti, M. A., Taylor, R. H., Fichtinger, G. A rapid calibration method for registration and 3D tracking of ultrasound images using spatial localizer. Vol. 5035. 2003. pp. 521–532.
- Comeau, R. M., Fenster, A., Peters, T. M. Integrated MR and ultrasound imaging for improved image guidance in neurosurgery. In: Proceedings of SPIE. Vol. 3338. 1998. pp. 747–754.
- Detmer, P. R., Bashein, G., Hodges, T., et al. 3D ultrasonic image feature localization based on magnetic scanhead tracking: in vitro calibration and validation. Ultrasound in Medicine & Biology 1994. 20 (9), 923–936.
- Fenster, A., Downey, D. B., Cardinal, H. N. Three-dimensional ultrasound imaging. Physics in Medicine and Biology 2001. 46, R67–R99.
- Fischler, M. A., Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 1981. 24 (6), 381–395.
- Gee, A. H., Houghton, N. E., Treece, G. M., Prager, R. W. A mechanical instrument for 3D ultrasound probe calibration. Ultrasound in Medicine & Biology 2005. 31 (4), 505–518.
- Gee, A. H., Prager, R. W., Treece, G. H., Berman, L. H. Engineering a freehand 3D ultrasound system. Pattern Recognition Letters 2003. 24, 757–777.
- Gobbi, D. G., Comeau, R. M., Lee, B. K. H., Peters, T. M. Integration of intra-operative 3D ultrasound with pre-operative MRI for neurosurgical guidance. In: Proceedings of the 22nd Annual Interna-

- tional Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 3, 2000. pp. 1738–1740.
- Hough, P. V. C. Machine analysis bubble chamber pictures. In: International Conference on High Energy Accelerators and Instrumentation. CERN, 1959. pp. 554–556.
- Huang, Q. H., Zheng, Y. P., Lu, M. H., Chi, Z. R. Development of a portable 3D ultrasound imaging system for musculoskeletal tissues. *Ultrasonics* 2005. 43, 153–163.
- Kofler Jr., J. M., Madsen, E. L. Improved method for determining resolution zones in ultrasound phantoms with spherical simulated lesions. *Ultrasound in Medicine & Biology* 2001. 27 (12), 1667–1676.
- Lindseth, F., Tangen, G. A., Langø, T., Bang, J. Probe calibration for freehand 3-D ultrasound. *Ultrasound in Medicine & Biology* 2003. 29 (11), 1607–1623.
- Martin, K., Spinks, D. Measurement of the speed of sound in ethanol/water mixtures. *Ultrasound in Medicine & Biology* 2001. 27 (2), 289–291.
- Mercier, L., Langø, T., Lindseth, F., Collins, D. L. A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound in Medicine & Biology* 2005. 31 (4), 449–471.
- More, J. J. The Levenberg-Marquardt algorithm: Implementation and theory. In: Numerical Analysis, Lecture Notes in Mathematics. Vol. 630. Springer-Verlag, 1977. pp. 105–116.
- Muratore, D. M., Galloway Jr., R. L. Beam calibration without a phantom for creating a 3-D freehand ultrasound system. *Ultrasound in Medicine & Biology* 2001. 27 (11), 1557–1566.
- Pagoulatos, N., Haynor, D. R., Kim, Y. A fast calibration method for 3-D tracking of ultrasound images using a spatial localizer. *Ultrasound in Medicine & Biology* 2001. 27 (9), 1219–1229.
- Prager, R. W., Rohling, R. N., Gee, A. H., Berman, L. Rapid calibration for 3-D freehand ultrasound. *Ultrasound in Medicine & Biology* 1998. 24 (6), 855–869.
- Rousseau, F., Hellier, P., Barillot, C. Confusius: A robust and fully automatic calibration method for 3D freehand ultrasound. *Medical Image Analysis* 2005. 9 (1), 25–38.
- Sato, Y., Nakamoto, M., Tamaki, Y. et al. Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization. *IEEE Transactions on Medical Imaging* 1998. 17 (5), 681–693.
- State, A., Chen, D. T., Tector, C. et al. Case study: observing a volume rendered fetus within a pregnant patient. In: Proceedings of the Conference on Visualization '94, IEEE Visualization. IEEE Computer Society Press, California, 1994. pp. 364–368.
- Treese, G. M., Gee, A. H., Prager, R. W., Cash, C. J. C., Berman, L. H. High-definition freehand 3-D ultrasound. *Ultrasound in Medicine & Biology* 2003. 29 (4), 529–546.
- Treese, G. M., Prager, R. W., Gee, A. H., Berman, L. Correction of probe pressure artifacts in freehand 3D ultrasound. *Medical Image Analysis* 2002. 6, 199–214.
- Trobaugh, J. W., Richard, W. D., Smith, K. R., Bucholz, R. D. Frameless stereotactic ultrasonography: method and applications. *Computerized Medical Imaging and Graphics* 1994. 18 (4), 235–246.
- Varandas, J., Baptista, P., Santos, J., Martins, R., Dias, J. VOLUS — a visualization system for 3D ultrasound data. *Ultrasonics* 2004. 42, 689–694.

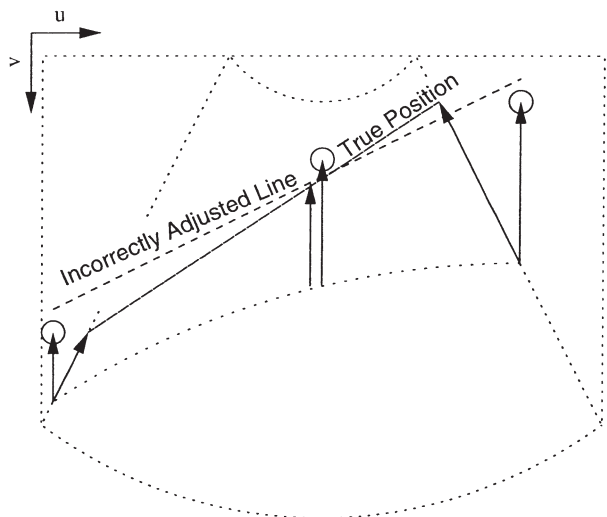
APPENDIX A: THE CURVILINEAR PROBE

It is clear that the temperature correction routine is more sophisticated for the curvilinear probe than for the linear probe. This includes the detection of the probe shape and curvature in addition to the mathematical adjustments, which is an extra step for the user during calibration. We therefore investigate what is the expected error, if we use the temperature correction routine for a linear probe to approximate the exact correction.

Figure 15 (a) shows an exaggerated diagram when the plane is scanned in such a way that the plane appears as a horizontal line, or in fact a slight curve. The exact correction is shown with arrows pointing in the direction of the shift: the result is a horizontal straight line, shown



(a) Horizontal plane



(b) Oblique plane

Fig. 15. The difference between the corrected planes when applying the correct and approximated correction routines to the displayed plane. The approximated correction is shown with arrows ending with a circle. It is assumed that a best fit line will be fitted to the nonlinear corrected points. Both diagrams are exaggerated to show the correction error more clearly.

as a dashed line in the diagram. This is the true position of the plane in space. The approximate correction is shown as arrows ending with a circle. It is clear, from the diagram, that the maximum error occurs at the middle of the image, where the point is overcorrected. If this central point appears at a depth d in the B-scan and t represents the temperature correction factor, then the point is corrected to position td . For an exact correction, the point should be corrected to $t(d - c) + c$. The overcorrection is thus the difference $(1 - t)c$. The line detection algorithm is likely to estimate the line to be around the middle of the two extreme points as the best fit line (shown as a dashed line) and so the expected error e should be about half the overcorrection error, i.e. $e \approx \frac{1}{2}(1 - t)c$. Since $t = 0.963$ for water at 20° C and c is less than 1 cm for common

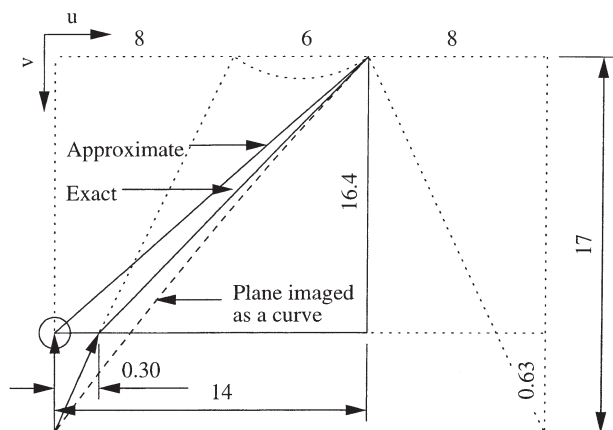


Fig. 16. An estimate of the error caused by temperature correction differences. The dimensions are shown in centimetres for our Toshiba curvilinear probe.

abdominal curvilinear probes, this error ($e \approx 0.01 - 0.02$ cm) is negligible for such probes.

Figure 15 (b) shows an exaggerated diagram when the plane is scanned from an oblique direction. The difference between the two correction routines in this case is mainly a rotational error, rather than a distance error. The incorrectly corrected line will appear not as steep as the real position of the plane, while the distance at the centre will remain roughly at the same place as before. This error is difficult to quantify, as it is dependent on the depth, curvature and width of the probe, slant of the plane and even the thresholds in the line segmentation algorithm. As an estimate of the expected error, we compute the error using the dimensions of our Toshiba curvilinear probe, as shown in Fig. 16. Assuming the depth is set at 17 cm, the maximum error occurs when the plane is scanned most obliquely. In this case, the error can be computed by simple trigonometry to be 0.6° . In practice, the error will be less than this value. Since this type of motion only occurs in approximately 20% of the calibration, its impact on the calibration error is limited.

We verify that using the simple temperature correction is a good approximation to the exact correction by performing an additional 10 calibrations with the Toshiba curvilinear 3.75 MHz probe using the simple temperature correction routine and compare this result with the previous calibrations with the same probe, but with the exact temperature correction routine applied. Since the probe shape no longer needs to be detected, it is therefore not necessary to have the sides of the probe visible within the images. This allows the B-scans to be displayed at a wide field of view, the more common setting for the curvilinear probe in clinical environments. The difference in settings, including the cropped B-scan sizes, for the two sets of calibrations is shown in Table 2.

The repeatability of the curvilinear probe shape detection is shown in Table 3, with the values in pixels. It is clear that the centre and radius of the probe are accurate to within 2 pixels (0.3 mm).

Figure 17 shows the mean 3D errors of the four corners and the centre of the B-scan for the curvilinear probe, when the two tempera-

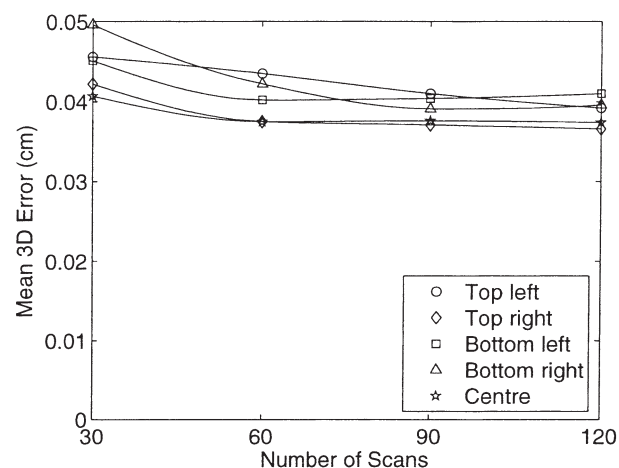
Table 2. Setting differences for the Toshiba curvilinear 3.75 MHz probe used for determining the effect of the different temperature correction routines. The depth is set at 6 cm.

Correction	Cropped B-scan	Refresh rate
Exact	388×380	36 Hz
Approximate	496×416	18 Hz

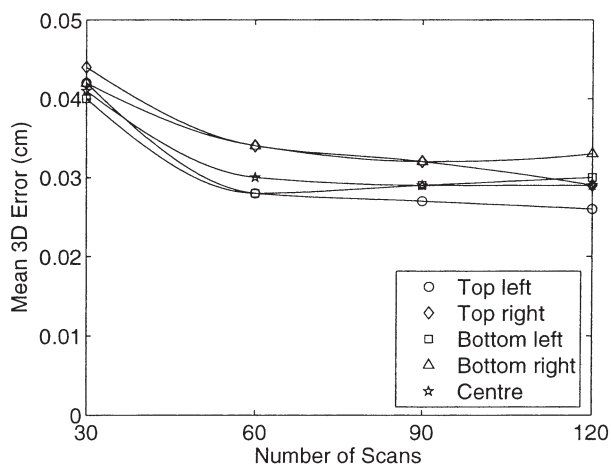
Table 3. Repeatability of the estimation for curvilinear probe centre position and radius. The unit of the values in the table is pixels. The scales in each B-scan are determined explicitly to be 0.0143 cm/pixel.

	Centre			
	x	y	2D	Radius
Mean	193.1	-489.5	(193.1; -489.5)	499.1
RMS	0.115	1.120	1.126	1.105
95%	± 0.237	± 2.314	± 2.344	± 2.283

ture correction routines are applied. The similarity of the precisions shown in the two graphs indicates that there is not a need for accurate temperature correction. Although the precision is slightly better when an approximation to the temperature correction routine is used, it is not statistically significant, given the spread of errors of these five points. The difference is also small when compared with the error achieved by



(a) Exact Correction



(b) Linear Approximation

Fig. 17. The effect of temperature correction routines on the consistencies of the four corners and the centre of the B-scan.

Table 4. The measured and original manufactured distances between two or three spheres, depending on the number of spheres that could be seen in each B-scan. All distances are in centimeters, displayed in the form mean \pm standard deviation

Probe	Measured distances		Original distances	
	Horizontal	Vertical	Horizontal	Vertical
Diasus 3 cm	3.021 ± 0.0087	0.947 ± 0.012	3 ± 0.01	1 ± 0.01
Diasus 6 cm	3.005 ± 0.0181	1.914 ± 0.013	3 ± 0.01	2 ± 0.01

Treece *et al.* (2003) and in this paper, as shown in Fig. 5. An exact temperature correction routine for common abdominal curvilinear probes is therefore not necessary to achieve high precision.

APPENDIX B: VERIFICATION OF THE SPHERE DISTANCES

Although the location of each sphere was precisely manufactured, the phantom has shrunk over the years by losing water. This has caused the spheres to be closer than the manufactured dimensions, typically near the scanning window at the top away from the rigid walls. We first measured the distances between the spheres. This was done by capturing multiple ultrasound images in the direction of the coplanar spheres with the Diasus 5–10 MHz probe with multiple foci down its depth. The horizontal and vertical distances between the

spheres were measured using the distance measuring tool provided by the ultrasound machine in a single image. The measurements are shown in Table 4. The vertical distances between spheres have shrunk by approximately 5%, while the horizontal distances remained roughly unchanged. We modelled the locations of the spheres to lie on a rectangular grid, with the measured vertical distance and the manufactured horizontal distance. We have kept the manufactured horizontal distance, as we do not anticipate any shrinking or expanding in the horizontal direction because the phantom is surrounded by rigid walls on its sides. The very slight difference in the measured distances is within measurement error.

APPENDIX C: SOLUTION OF THE CRITERION FUNCTION

We have utilized the nonlinear Levenberg-Marquardt optimisation algorithm (More, 1977) to solve for the solution to the system of nonlinear equations involving nine parameters. The calibration parameters are solved in three stages: the angles are first solved, then the distances and then a final global optimisation to fine-tune the solution (Prager *et al.*, 1998). As is the case for all optimisation algorithms, a starting point is necessary. This is chosen at random, but within a reasonable estimate of what the parameters can be. This means that the solution given by the optimisation algorithm is dependent on the starting point. Hence, we need to ensure that the optimisation converges to the global minimum, rather than to any other local minimum. Based on trial runs of the optimisation, the convergence to a nonglobal minimum occurs less than 50% of the time. We therefore repeat the optimisation 50 times at different starting points and accept the solution with the minimum error. This process takes less than 1 s on our 3.0 GHz PC. The probability that the optimisation still converges to the incorrect minimum is 1 in 10^{15} .