

● *Original Contribution***HIGH-DEFINITION FREEHAND 3-D ULTRASOUND**GRAHAM M. TREECE,\* ANDREW H. GEE,\* RICHARD W. PRAGER,\* CHARLOTTE J. C. CASH<sup>†</sup>  
and LAURENCE H. BERMAN<sup>†</sup>\*Department of Engineering, Trumpington Street, Cambridge, UK; and <sup>†</sup>Department of Radiology, Addenbrooke's Hospital, Cambridge, UK

(Received 9 August 2002; in final form 30 October 2002)

**Abstract**—This paper describes a high-definition freehand 3-D ultrasound (US) system, with accuracy surpassing that of previously documented systems. 3-D point location accuracy within a US data set can be achieved to within 0.5 mm. Such accuracy is possible through a series of novel system-design and calibration techniques. The accuracy is quantified using a purpose-built tissue-mimicking phantom, designed to create realistic clinical conditions without compromising the accuracy of the measurement procedure. The paper includes a thorough discussion of the various ways of measuring system accuracy and their relative merits; and compares, in this context, all recently documented freehand 3-D US systems. (E-mail: gmt11@eng.cam.ac.uk) © 2003 World Federation for Ultrasound in Medicine & Biology.

**Key Words:** 3-D ultrasound, Freehand scanning, Musculoskeletal ultrasound.

**INTRODUCTION**

Advances in the resolution and quality of two-dimensional (2-D) ultrasound (US) imaging are increasingly enabling detailed examination of arterial and musculoskeletal anatomy (Fornage et al. 2000; Wang et al. 1999). However, high-resolution US images (B-scans) have a limited field of view, generally sufficient for scanning the cross-section, but not the length, of the anatomy of interest. 3-D US can overcome this limitation. Not only does it provide the ability to generate extended images, it also allows the visualisation of complex structures, such as ligaments and cartilage or arterial plaque, in a much more intuitive way. A further advantage is that 3-D US, by its very nature, offers much more precise measurement of volume and the relative orientation of structures.

There are many ways to design 3-D US systems (Fenster et al. 2001). The most appropriate technique for arterial and musculoskeletal anatomy is freehand 3-D US, where the probe is moved by hand, and the resulting sequence of B-scans is located in 3-D space by either intrinsic (image-based) or extrinsic (position-sensing) means. This is the only technique that gives the clinician

complete freedom to guide the probe along the path of the anatomy.

Most recently documented freehand 3-D US systems use either an electromagnetic (Barry et al. 1997; Edwards et al. 1998; Prager et al. 1999) or an optical (Blackall et al. 2000; Bouchet et al. 2001) position sensor. B-scans are transferred from the US scanner to an external PC by digitising from the scanner's video output (Edwards et al. 1998; Prager et al. 2002; Meairs et al. 2000), from a video recording (Barry et al. 1997) or by direct digital transfer (Berg et al. 1999). All position-sensing techniques, and most image-transfer protocols, introduce additional sources of error not present in the original 2-D B-scans: typical accuracy of documented 3-D systems is on the order of  $\pm 2$  mm. This is significantly worse than the inherent resolution of high-frequency B-scan images, which is better than 0.1 mm/pixel.

The position-sensing and image-acquisition techniques are only two of many steps that affect the eventual resolution of the 3-D system. A review of the engineering challenges in such systems, together with some practical suggestions for good acquisition protocols, is given by Gee et al. (2003). This review uses our system, Stradx (Prager et al. 1999), as a worked example. Stradx is a sequential freehand 3-D US system. In this paradigm, the (arbitrarily orientated, but only gradually varying) sequence of B-scans is preserved, rather than resampled

Address correspondence to: Dr. Graham Treece, Department of Engineering, Trumpington Street, Cambridge CB2 1PZ UK. E-mail: gmt11@eng.cam.ac.uk

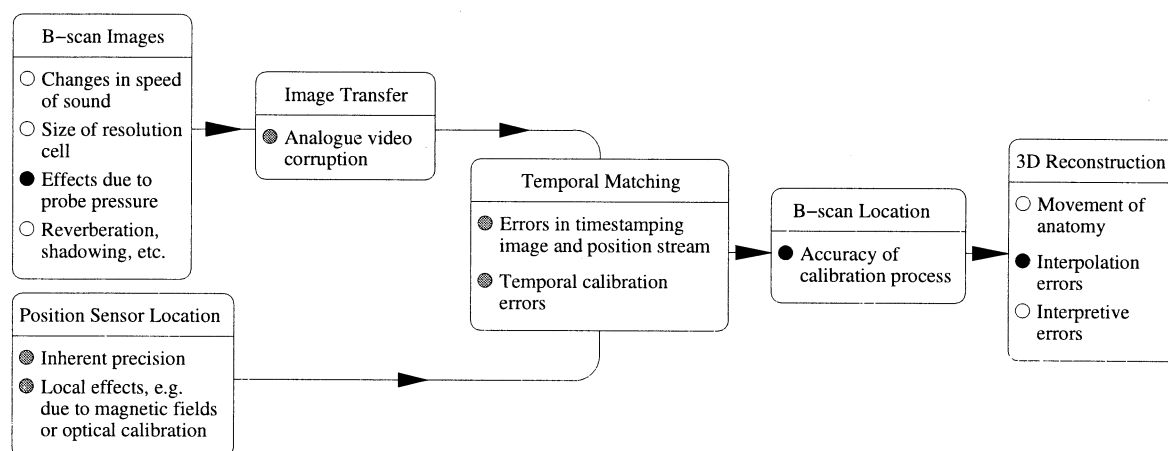


Fig. 1. Errors in freehand 3-D US systems. Some of the dominant errors are shown; those with black bullet points have been addressed in previous work, those with mid-grey bullet points are addressed in this paper, those with white bullet points are residual sources of error.

onto a regular voxel array, so that each visualisation or quantification step is calculated directly from the original data. This is an ideal starting point for improving the definition of 3-D systems because we preserve the 2-D resolution for as long as possible.

In this paper, we present recent developments to Stradx that aim to improve the overall system accuracy, so that the 3-D definition can approach that of the original, high-resolution 2-D B-scans. The system is rigorously assessed by using a purpose-built tissue-mimicking phantom, allowing us to estimate the actual errors that might be expected when making 3-D measurements in real clinical situations. We also consider the practical aspects of using the system and investigate the degradation due to remounting the position sensor on the US probe, and changing the US machine's depth setting, without carrying out a full recalibration.

#### *Analysis of errors in freehand 3-D ultrasound*

Before considering how we might improve the resolution of a freehand 3-D US system, we need a clear picture of where the errors come from. An overview of the main sources of error is given in Fig. 1. These can be grouped into errors in the B-scan images themselves, the readings from the position sensor, temporal matching of B-scans and positions, location of the B-scan relative to the position reported by the sensor and errors in the 3-D reconstruction of the B-scans.

Errors in the B-scans themselves are largely determined by the size of the resolution cell, which varies in all three dimensions. Typically, the out-of-plane resolution (or beam width) is significantly worse than the in-plane resolution, and varies across the depth of the image, dependent mainly on the out-of-plane focusing.

Variation in the speed of sound can also have a significant effect on the beam width: errors of  $\pm 5\%$  in sound speed, typical of variations in human tissue, can generate over 200% increases in beam width, as well as affecting the depth scale (Anderson et al. 2000). For high-resolution images, compression of anatomy due to probe pressure can be a large source of error, but this can be reduced by image-correlation techniques (Treece et al. 2002).

Of all the position-sensing techniques, optical position sensors are the most accurate, although they require a line of sight between the probe and the camera. Such systems can achieve a spatial accuracy (for the position sensor alone) of up to  $\pm 0.2$  mm (Blackall et al. 2000). Electromagnetic position sensors can achieve an accuracy of up to  $\pm 0.5$  mm in location and  $\pm 0.7^\circ$  in orientation (again, for the position sensor alone) when optimised for very specific situations and small spatial ranges (Barratt et al. 2001). In general use, however, they are subject to distortions that impair their accuracy considerably (Birkfellner et al. 1998).

Errors due to image transfer and temporal calibration (the matching of images to positions) are less widely discussed in the literature; most researchers opt for the practical solution of digitising the analogue video output of the US machine at between 10 and 25 frames per second. The images are corrupted by conversion to and from analogue video formats, and the temporal resolution is limited by the low frame rates. A notable exception is the system described by Berg et al. (1999), where images are transferred digitally at 150 frames per second; thus, giving a temporal resolution of 7 ms. Temporal calibration at 25 frames per second can be achieved to within 40 ms by looking for sudden changes in the image

and position streams (Meairs *et al.* 2000; Prager *et al.* 1999). Inaccurate temporal calibration will result in spatial errors with a magnitude dependent on the speed with which the probe is moved. Such errors can also reduce the spatial calibration accuracy.

Spatial calibration, the estimation of the rigid body transformation between the position sensor's reference frame and the B-scan plane, is one of the most dominant sources of error in freehand systems. The various calibration techniques are compared by Prager *et al.* (1998). The calibration process involves scanning a known object from a variety of orientations; this can be a single point (Legget *et al.* 1998), a set of points (Berg *et al.* 1999), a crosswire (Barry *et al.* 1997; Meairs *et al.* 2000), a "z-shape" (Bouchet *et al.* 2001), a real or virtual plane (Prager *et al.* 1998) or, in fact, any known shape (Blackall *et al.* 2000). By constraining the 3-D reconstruction to match the known geometry of the scanned object, it is possible to derive a system of equations for the eight spatial calibration parameters (six defining the location and orientation of the B-scan relative to the position sensor, and two defining the  $x$  and  $y$  scales of the B-scan in mm/pixel). The system of equations can either be inverted directly or, more usually, optimised iteratively.

Even after the location of each B-scan has been correctly determined, there are still further sources of error. An unspoken assumption in the subsequent 3-D reconstruction is that the subject has not moved during the acquisition; any such movements result in a distortion of the 3-D data. External movement can be ameliorated by attaching a coordinate reference to the patient (Chuang *et al.* 2001) and repetitive internal movement (*i.e.*, due to cardiac activity) by the use of an electrocardiogram (ECG) to gate the acquisition of B-scans (Belohlavek *et al.* 1994; Palombo *et al.* 1998). If possible, it is best to acquire data within a single breath-hold and to review it immediately for motion artefacts, so that the scan can be repeated if necessary (Gee *et al.* 2003). Where this is not possible, for instance, in acquiring dense ECG-gated data or where the patient has difficulty holding their breath, ECG or respiratory gating procedures must be used; however, these will inevitably reduce the accuracy of the 3-D freehand data.

Visualising the semistructured 3-D freehand data involves interpolating the data onto some sort of regular pixel or voxel array. Significant interpolation errors can arise as a consequence of the scanning pattern (Cardinal *et al.* 2000) combined with simplistic interpolation schemes, optimised for speed rather than quality (Rohling *et al.* 1999). Such errors can be limited by not resampling onto a regular voxel array, as described earlier. This approach also suppresses some interpretive errors; for instance, when delineating structures in artefact-ridden out-of-plane reconstructions (Bailey *et al.*

2001). Other interpretive errors arise from poor cursor placement when making measurements (Goldstein 2000).

#### *Assessment of system accuracy*

There are many ways of assessing the performance of a freehand 3-D US system and, unfortunately, there is no agreed standard. More confusingly, results are generally quoted simply as system "accuracy," despite differences in what was tested, where it was tested and how the results were analysed, which can lead to as much as a factor of 3 variation in the quoted result. Sometimes, insufficient information is provided to be able to interpret the quoted "accuracy" at all.

It is, therefore, necessary to clarify the differences between some of these measures before attempting to compare those systems that are described in the literature and place our system among them. This can, helpfully, be done by asking three questions: "what part of the system is included in the measurement," "what is it a measurement of," and "how are the measurements analysed?"

First, "what part of the system is included in the measurement?" With reference to Fig. 1, when designing a system, it might be helpful to know the accuracy of a specific part, for instance, the position sensor alone (Barratt *et al.* 2001), or the spatial calibration alone (Prager *et al.* 1998). Ultimately, however, it is the accuracy of the entire system, in the context in which it will be used, that is relevant to the clinician. *In vivo* accuracy is very difficult to assess, so *in vitro* accuracy is usually reported instead, by scanning a specially designed phantom in a water bath. This excludes some of the B-scan image errors, such as the speed of sound variation in human tissue and tissue deformation due to probe pressure. It also excludes some of the 3-D reconstruction errors, specifically those due to movement of anatomy and, to some extent, those due to interpretation of data (because phantom images are often significantly less complex than *in vivo* images). Clearly, the accuracy of the entire system can only be worse than that of its component parts.

Second, "what is it a measurement of?" There are a variety of possibilities here, in terms of the quantity measured, where it is measured and what it is compared with. Generally, the quantity is either the location of a fixed point (Barry *et al.* 1997; Blackall *et al.* 2000; Bouchet *et al.* 2001; Meairs *et al.* 2000; Prager *et al.* 1998), the distance between points (Blackall *et al.* 2000; Legget *et al.* 1998; Prager *et al.* 1998) or the volume of a defined object (Barry *et al.* 1997; Berg *et al.* 1999). The location of a point can, perhaps, be regarded as a more fundamental measure because volume and, to a lesser extent, distance are not affected by certain distortions of the 3-D data. Where the quantity is measured is partic-

ularly important if the spatial calibration has been optimised from the same data used to assess the system accuracy (which is, unfortunately, common practice in the literature). If this is the case, then what is being measured is only the calibration residual error (Barry et al. 1997), but how well this reflects the actual system accuracy is highly dependent on how well-conditioned the calibration optimisation is, and how well the calibration scanning pattern represents actual scanning practice. Even a repeated scan of the point on which the calibration was based (Prager et al. 1998) can be misleading; it is better to assess accuracy based on a completely different set of measurements. Finally, measurements can either be compared to “true” values (known from some other independent source), in which case they reveal system accuracy, or to themselves, in which case they reveal only the precision of the system.

Third, “how are the measurements analysed?” As an example, consider a set of measurements of point location, with independent errors in each of the  $x$ ,  $y$  and  $z$  dimensions that are normally distributed with zero mean and 1 mm SD. The 95% confidence limits in each dimension are approximately twice the SD (*i.e.*,  $\pm 2$  mm). However, the absolute 2-D location error (for instance, in the  $x$ - $y$  plane) is not normally distributed: it follows a Rayleigh distribution. The absolute 3-D location error has an even more complex distribution. The mean absolute 2-D error is approximately 1.25 mm, and 95% of the points lie within 1.85 times this (*i.e.*, within 2.3 mm of the true location). For the 3-D case, the mean error is 1.6 mm and the 95% limit is  $<2.8$  mm (approximately 1.75 times the mean). Note that these confidence limits are lower than the pessimistic estimate from simply summing the variances in each dimension, as in Legget et al. (1998). The SD of the 2-D and 3-D errors is sometimes also quoted; this is a misleading quantity because these errors are not normally distributed and can lead to optimistic assessments of system accuracy. For the example above, the SD is approximately 0.7 mm, for both the 2-D and 3-D cases.

A further complication arises from the use of paired analysis (Blackall et al. 2000; Prager et al. 1998), where a set of point measurements is analysed by considering the distribution of the absolute distance between all possible pairs of measurements. If we continue the example above, this analysis would give a mean 3-D error of 2.3 mm and 95% confidence limit of  $<4$  mm, whereas we already know that 95% of the values will lie within 2.8 mm of the correct location. In effect, the paired analysis measures relative distance accuracy, which has twice the variance of the point-location distribution because it is a measure of difference.

Although it is not, in general, possible to compare accuracy results that differ with regard to the first two of these questions, it is possible to use the example above to

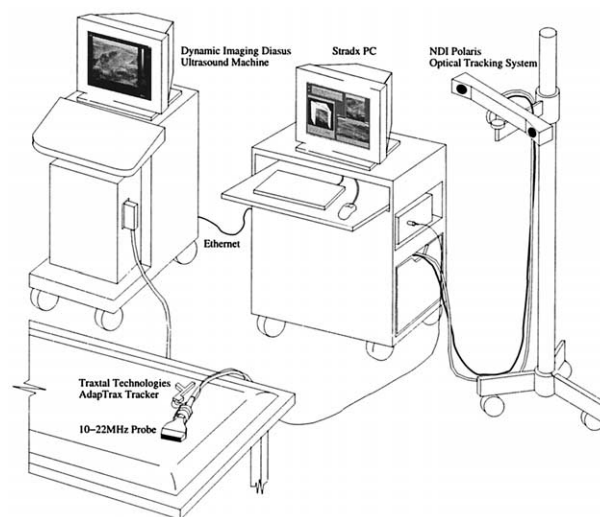


Fig. 2. Physical layout of the high definition freehand 3-D US system, indicating the approximate size and the required connections between the components, to help give a feel for the system's usability.

convert between results that are analysed differently, provided sufficient information has been given to determine the nature of the result that has been presented. In the comparison below, results are converted to the “3-D confidence limit”; this is the distance away from the mean location (for precision) or known location (for accuracy) within which 95% of the measured points will lie. This conversion assumes an unbiased distribution of errors; significant bias in the results will introduce errors in this process.

The performance of a freehand 3-D US system is clearly dependent on the type of position sensor and the frequency of the US. Prager et al. (1998) used an electromagnetic position sensor and a 7 MHz probe at a 4 cm depth setting; Blackall et al. (2000) used an optical position sensor and a 10 MHz probe, again at a 4 cm depth setting; Meairs et al. (2000) used an electromagnetic position sensor and a 5–12 MHz probe; Legget et al. (1998) used an electromagnetic position sensor and a 3 MHz probe; and Bouchet et al. (2001) used an optical position sensor and a 3.5 MHz probe. Systems using optical position sensors and higher-frequency probes will tend to be more accurate. Lower depth settings can also result in greater accuracy, depending on whether or not the B-scan is zoomed in the video display and how this relates to the actual resolution of the B-scan.

Since spatial calibration is such an important step in the design of an accurate system, several authors quote the point precision due to the spatial calibration alone. The 3-D confidence limits achieved for this value are  $<1.2$  mm (Prager et al. 1998) and  $<2.3$  mm (Blackall et al. 2000) (both derived from mean paired absolute error). Another frequently quoted value is the point precision of

the entire system (as measured by scanning a phantom; note the earlier comments about *in vitro* measurements). 3-D confidence limits for this value are <2.7 mm (Prager *et al.* 1998), <1.4 mm (Blackall *et al.* 2000) and <2.6 mm (Meairs *et al.* 2000) (all derived from mean paired absolute error), <3.4 mm (Legget *et al.* 1998) (derived from the sum of the variances in each dimension) and <2.2 mm (Bouchet *et al.* 2001) (derived from the mean absolute error in each dimension). Finally, several authors quote the errors in distances between several points; this leads to a measure of accuracy within a particular data set that is sensitive to any distortion introduced into the data, but not to systematic errors in point location. 3-D confidence limits for point location accuracy based on this measure are <1.9 mm (Prager *et al.* 1998), <1.0 mm (Blackall *et al.* 2000) and <1.1 mm (Legget *et al.* 1998), all derived from the SD of the paired signed distance errors.

## HIGH-DEFINITION SYSTEM

### Physical layout

Figure 2 shows the physical layout of our freehand 3-D US system. B-scans are acquired with a Diasus US machine<sup>1</sup>, using 5–10 MHz and 10–22 MHz linear-array probes, on 2 cm, 3 cm and 6 cm depth settings. Eight-bit digital log-compressed data are transferred *via* Ethernet at 25 B-scans per second to an 800-MHz PC running Linux. The probe position is measured by a Polaris<sup>2</sup> optical system tracking an AdapTrax<sup>3</sup> target attached to the probe. The camera is mounted on a stable, but highly manoeuvrable stand<sup>4</sup> designed for studio video cameras. Calibration, acquisition, processing and display of the data are performed on the PC by Stradx (Prager *et al.* 1999)<sup>5</sup> that can also be used in conjunction with a number of other video sources and position sensors.

To avoid having to recalibrate the system each time a different probe is used, the probe mount shown in Fig. 3 was designed. The mount can be attached to most US probes by simply winding a length of Velcro<sup>TM</sup> tightly around the mount and the probe; the rubber holds the mount firmly in place and the locating hole ensures that the position sensor can be fixed to the mount at a known, repeatable position. The mount is small, and its presence on the probe does not inhibit normal use. In practice, we have sufficient mounts for all the probes in use and these mounts remain on the probes between 3-D scanning

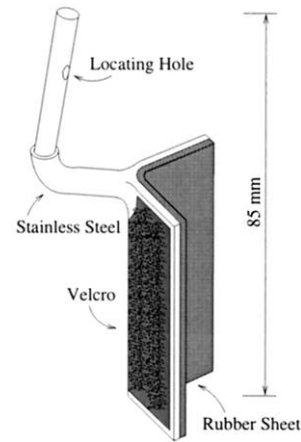


Fig. 3. Generic probe mount for the optical tracker. The mount can be quickly attached to almost any type of US probe, and is sufficiently small and cheap to remain attached between 3-D scanning sessions. The optical tracker can be fixed at the same position on the mount to a high degree of precision, so that spatial calibration is only necessary if the mount itself is removed from the probe.

sessions.

### System design

Figure 4 shows the basic design of the part of the system involved with delivering B-scans, tagged with their correct spatial position and orientation, to Stradx's front end visualisation and quantification algorithms. The sequential reslice, manifold and panoramic displays are described in Prager *et al.* (1999), volume rendering extensions in Gee *et al.* (2002) and volume measurement and surface reconstruction in Treece *et al.* (1999). We are concerned in this paper with the quality of the raw data on which these algorithms are based.

Key components of this system, discussed in the following sections, are the transfer of US images to the PC, the synchronisation of the PC and US machine timers, the matching of images to positions and the gating of acquired images. Temporal and spatial calibration of the system will be discussed later.

### Ultrasound image transfer

In almost all other freehand 3-D US systems, B-scans are transferred between the US machine and the PC in an analogue video format. This is an appealing transfer mechanism because most US machines have a suitable analogue video output, and video framegrabbers for PCs are cheap and commonplace. However, the unnecessary digital-to-analogue conversion in the US machine, followed by an analogue-to-digital conversion in the framegrabber, introduces unwanted noise into the B-scans. For our high-definition system, we would like

<sup>1</sup> Dynamic Imaging Ltd., Livingstone, UK, <http://www.dynamicimaging.co.uk/>

<sup>2</sup> Northern Digital Inc., Waterloo, Ontario, Canada, <http://www.ndigital.com/>

<sup>3</sup> Traxtal Technologies, Toronto, Ontario, Canada, <http://www.traxtal.com/>

<sup>4</sup> Unicol, Engineering, Oxford, UK, <http://www.unicol.co.uk/>

<sup>5</sup> <http://svr-www.eng.cam.ac.uk/~rwp/stradx/>

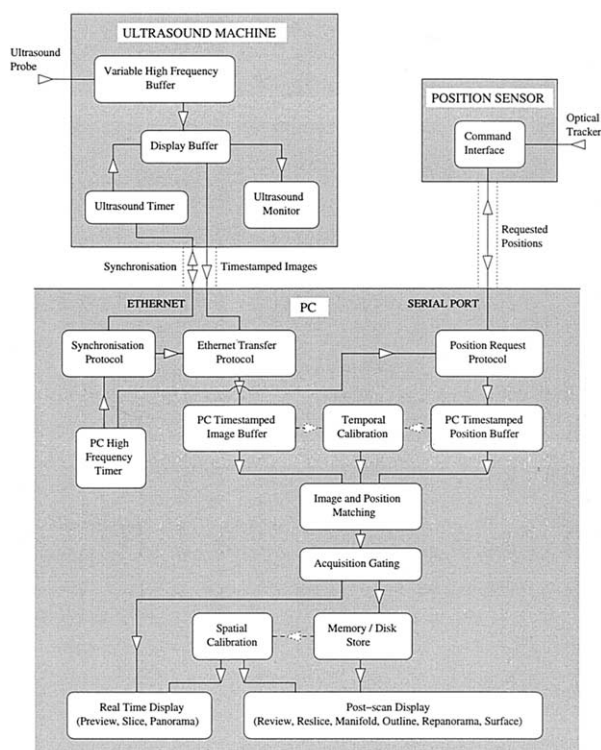


Fig. 4. High-definition freehand 3-D US system design; the main hardware and software components of the PC handling the acquisition and display of the B-scan images and the relevant parts of the US system.

to avoid this source of error and, instead, transfer the images in an uncorrupted digital format.

One solution would be to employ a custom digital link, like the one described by Berg et al. (1999). However, in the interests of cost, flexibility and rapid prototyping, it is preferable to avoid specialised hardware and to work, instead, with standard PC components. The Diasus, like many modern US machines, is equipped with an off-the-shelf PC front end that includes an Ethernet network interface. This raises the possibility of streaming the B-scans across the network; for a switched 100 Mbit/s link, and assuming one byte per grey-scale pixel, it should be possible to transfer large B-scans (up to  $724 \times 724$  pixels) at the full PAL frame rate of 25 Hz.

This solution proved to be both feasible and highly effective. Minimal changes to the Diasus internal software were required to send each B-scan to the Ethernet card at the same time as it was displayed on the screen. A suitable module was added to Stradx to receive the frames from the Ethernet instead of from the usual video framegrabber. The software for both ends of the link was developed in a few days, with no need for specialised hardware. Moreover, the resulting system is compatible with any US machine equipped with a PC front end.

Even though the Diasus is a purely grey-scale machine, the network transfer mechanism could work equally well with Doppler machines, provided each pixel is coded into a single-byte colour index before transmission<sup>6</sup>.

The one complication concerns the time-stamping of the B-scans. Each B-scan needs to be tagged with the time at which it was acquired, so that it can later be matched with a position read from the position sensor at the same time. Time-stamping the B-scans when they are received by the client PC would be unreliable because the time of acquisition is not deducible from the time of receipt; there is an unpredictable network latency to consider. So the time-stamps must be applied by the server (the US machine) and sent down the Ethernet link with each B-scan. In our case, the time-stamps are applied at the point at which B-scans are extracted from the display buffer, which runs at a constant rate of 30 Hz. The rate at which the scans are actually acquired is usually higher than this, dependent mainly on the number of focal points in the transmit path. Applying the time-stamps at the display buffer rather than the variable-rate high-frequency buffer introduces a delay that is accounted for by temporal calibration<sup>7</sup>.

Sending a suitably high-resolution time-stamp down the Ethernet link requires only a few extra bytes per B-scan, so there is no impact on the Ethernet frame rate. But, there is an issue with clock synchronisation; the clock in the US server will not be telling the same time as the clock in the client PC and will not even run at the same rate. There will be a certain amount of relative clock drift. Because the positions are time-stamped by the client PC, we need a way to translate a US server time-stamp into a client PC time-stamp, so the B-scan times can be compared directly with the position times.

A simple network time protocol was developed to achieve this. When the client PC first establishes Ethernet communication with the US machine, it requests a single time-stamp  $t_s$  from the server. When  $t_s$  is received, the client immediately takes a local time-stamp  $t_c$ . The difference  $\Delta_1 = t_c - t_s$  can be used to translate server time-stamps into client time-stamps. However,  $t_s$  might have been held up between the server and client by a network delay, causing  $\Delta_1$  to be bigger than it should be, so the exchange is repeated 10 times and the smallest  $\Delta_1$  is retained.

This simple procedure captures the time difference between the two clocks in an instant, but tells us nothing about the relative clock drift. The B-scan time-stamps

<sup>6</sup> 128 grey levels and 64 shades of red and blue are all that is needed to represent a colour Doppler US image.

<sup>7</sup> This delay is itself dependent on the actual rate of acquisition and will, therefore, change with the number of focal points.

will soon become inaccurate unless we account for this drift. So, the next time the client connects to the server (perhaps at the start of a 3-D US recording), the time-stamp exchange is repeated, yielding a new difference  $\Delta_2$ . Because we know when  $\Delta_1$  and  $\Delta_2$  were measured, we can estimate the relative clock drift and use this to correct subsequent server-to-client time-stamp translations. Furthermore, a new value of  $\Delta$  can be measured at the start of each 3-D recording, leading to better estimates of the drift over longer time-scales. The end result is accurate B-scan time-stamps, applied by the US machine but reliably translated into client time, ready for matching with time-stamped positions from the position sensor.

#### *Image to position matching*

Recently acquired B-scans are stored in a circular buffer in the client PC memory. A similar circular buffer holds recently acquired positions. B-scans are harvested from the circular buffer after a slight delay; that is, we start by processing a recent B-scan, though not the most recent one. This delay is important because we might not yet have a position to match with the most recent B-scan, depending on the latency of the position sensor<sup>8</sup>. The result of this calibration is used to correct the position time-stamps before they are stored in the circular buffer.

The time-stamp of the harvested B-scan is compared with the time-stamps in the position circular buffer. An accurate position for the B-scan is obtained by linear interpolation of the two position readings on either side of the B-scan time-stamp. At this point, the B-scan is tagged with its position and made available to Stradx's visualisation front end. The process continues with the next frame in the B-scan circular buffer; further details can be found in Prager *et al.* (1999).

#### *Acquisition modes*

In its default mode, Stradx processes B-scans sequentially, harvesting one frame after another from the B-scan circular buffer. Making use of every B-scan ensures the maximum possible acquisition rate, though this is not always desirable. For example, if the probe is being held stationary, the system will record endless B-scans on top of each other, repeatedly sampling the same slice of space. For this reason, Stradx offers motion-gated acquisition (Carr *et al.* 2000); after position tagging, B-scans are discarded if the probe has not moved by more than a user-defined threshold since the last retained B-scan. Motion-gated acquisition maintains an even spatial sampling density, independent of probe velocity.

A further acquisition mode is stable-gated acquisition. Again, the probe must have moved by a certain amount for a B-scan to be retained, but the probe must also be moving relatively slowly (another user-defined threshold), otherwise the B-scan is discarded. This ensures the acquisition of well-spaced B-scans free of motion blur artifacts. Such data sets are particularly useful for spatial calibration, as discussed in the following section.

## CALIBRATION

Temporal calibration is necessary to determine the offset between the position sensor time-stamps and the B-scan time-stamps. A matching B-scan and position reading will not necessarily have the same time-stamp, depending on the latencies of the two data streams. Particularly pertinent is the latency of the position sensor, because the position time-stamps are applied, not by the position sensor (which has no clock), but by the PC each time it requests a reading from the sensor. There is an unknown delay between the PC making the request and the sensor taking the reading.

Spatial calibration determines the size and location of the invisible B-scan plane relative to the position sensor reference frame. Because spatial calibration involves the analysis of matched B-scans and positions, it follows that poor temporal calibration will affect the accuracy of the spatial calibration; hence, temporal calibration must be performed first. This means that the temporal calibration procedure must be designed with no knowledge of the location or size of the B-scan.

#### *Temporal calibration*

Temporal calibration has previously been achieved by looking for sudden changes in B-scan and position sensor readings, caused, for example, by suddenly jerking the probe away from the skin surface, and matching them up in time (Meairs *et al.* 2000; Prager *et al.* 1999). For images acquired at 25 Hz, this gives a temporal resolution of only 40 ms. It is possible to improve on this value by comparing a stream of position sensor readings with positions derived from the B-scans themselves.

We have previously described an algorithm that can automatically detect horizontal lines in a B-scan, as in Fig. 5 (Prager *et al.* 1998). Such lines can easily be generated by scanning the bottom of a bath full of water. The line-detection algorithm can run sufficiently fast to give a real-time estimate (*i.e.*, at the PAL frame rate of 25 Hz) of the location of the base of the water bath in the B-scan. Note that this measurement only gives the variation in height of the probe above the (horizontal) surface and cannot detect horizontal movements. The dominant probe movement must, therefore, be vertical for

<sup>8</sup> The position sensor latency is accounted for by the temporal calibration procedure described later.

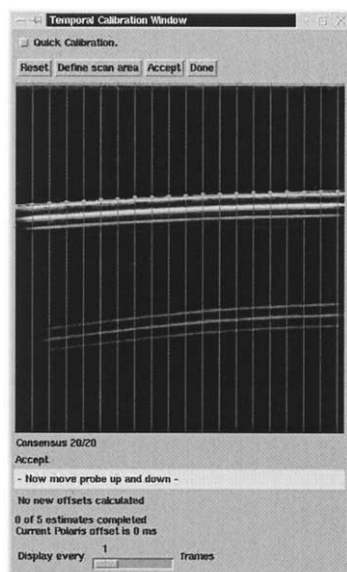
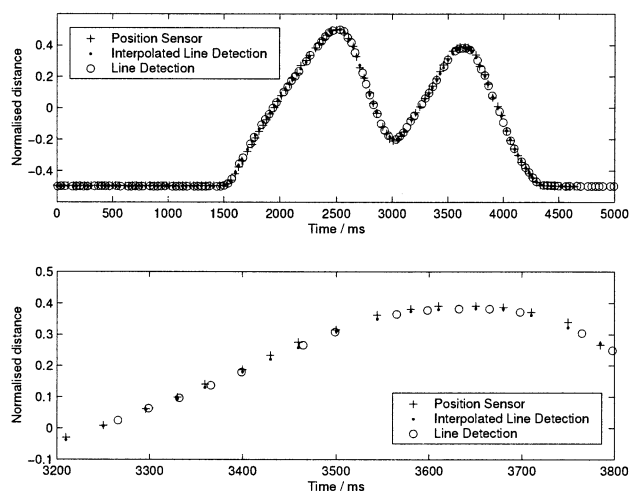


Fig. 5. Automatic line detection for temporal calibration. The key to temporal calibration is the ability to provide an automatic real-time estimate of the location of a horizontal echo in the B-scan that can subsequently be compared to the position sensor readings. This is achieved by looking for large gradients along the vertical lines, then fitting a line through this set of points.

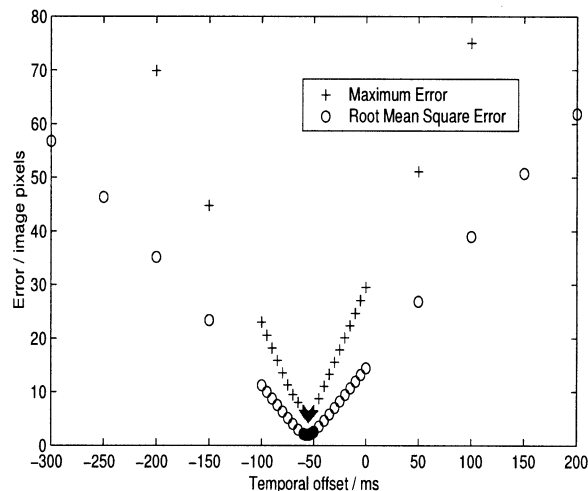
this measurement to correlate well with readings from the position sensor. This can be achieved with sufficient accuracy by moving the probe by hand.

At this stage, without a spatial calibration, neither the B-scan scale nor the orientation of the B-scan with respect to the position sensor coordinate system are known; we do not know how far the probe has travelled, nor which direction is up. However, assuming the probe was moved vertically, 3-D position sensor readings can be converted to vertical distances by taking the scalar product of each reading with the direction of maximal movement over the whole calibration sequence. The sign can be deduced by comparison with the distances derived from the B-scan. Finally, both the B-scan and position sensor distances are arbitrarily normalised to lie in the range  $-0.5$  to  $0.5$ , to allow for the unknown scale. To prevent this normalisation from badly conditioning the problem, only calibrations where the probe has moved a certain distance (as measured in the B-scan) are accepted for further processing.

A temporal calibration, therefore, proceeds as follows. After 1 s with the probe held steady, a further 4 s of B-scans are acquired, during which the probe is moved (by hand) up and down above the base of the water bath. After normalisation, this results in a stream of distance measurements from both the position sensor and the B-scan, as in Fig. 6a (shown at the correct temporal offset, rather than that at which they were acquired). These measurements are at different rates, approximately every 40 ms for the images and 30 ms for the positions. To correlate the two streams, the four closest distances derived from the images are interpo-



(a) Image and position sensor signals



(b) Signal correlation against temporal offset

Fig. 6. Temporal calibration of the B-scan image stream to the position stream. (a) For a given temporal offset, the position signal derived from the images is interpolated to give values at the same time points as each of the position estimates from the position sensor. (b) The (interpolated) image and position sensor position streams can then be correlated at various temporal offsets. A multiresolution approach is used to speed up the search for the minimum error between the streams.



lated with a cubic polynomial to give values that coincide (for a given temporal offset) with those from the position sensor.

The temporal offset is found that gives the minimum root-mean-square error between the distances derived from the B-scans and those derived from the position sensor. The correlation function is generally well behaved with respect to the temporal offset, as can be seen in Fig. 6b, so a simple multiresolution search can be used to find the optimum offset. The maximum error for a given offset is also calculated, and the result is disregarded if this is greater than 10 pixels (as measured in the B-scan); this gives some protection against calibrations where the probe was not moved with sufficient care.

The whole process is guided by prompts from the user interface as to when to hold the probe steady, when to move it, and whether or not the movement was sufficient and the eventual correlation good enough. Acceptable results can either be used individually or averaged to improve the precision. With a little practice, 10 or more calibrations can be performed within 1 min, leading to a temporal resolution that is significantly better than that of the original image and position streams.

#### *Spatial calibration*

Spatial calibration of the high-definition system is performed using the Cambridge phantom (Prager *et al.* 1998), which essentially offers calibration on a flat plane. The plane appears as lines in the B-scans; by constraining the 3-D reconstruction of these lines to be coplanar, the eight calibration parameters can be estimated. Moreover, because the lines can be detected automatically in the B-scans, as illustrated in Fig. 5, the calibration process is extremely rapid. However, the accuracy of this approach is limited by beam width effects; with the probe held at an oblique angle to the plane, reflections from the edge of the US beam can cause the plane to appear in the wrong position in the B-scan. The Cambridge phantom overcomes this problem by generating a “virtual” plane, with reflections coming only from the centre of the US beam. Full details of the phantom and its use are described by Prager *et al.* (1998). Here, we describe some refinements that increase the accuracy of the spatial calibration.

First, the improved temporal calibration process, combined with the stable gated acquisition mode, significantly reduces temporal errors in the calibration process; the result is more accurate labelling of each calibration B-scan with its position. In addition, the high quality of the digitally-transferred images improves the accuracy with which the line-detection process can be performed. We also aim to keep the water fairly hot during the calibration; tap water at 50°C has the same speed of sound, 1540 m s<sup>-1</sup>, as is normally assumed for average human tissue. These steps improve the conditioning of

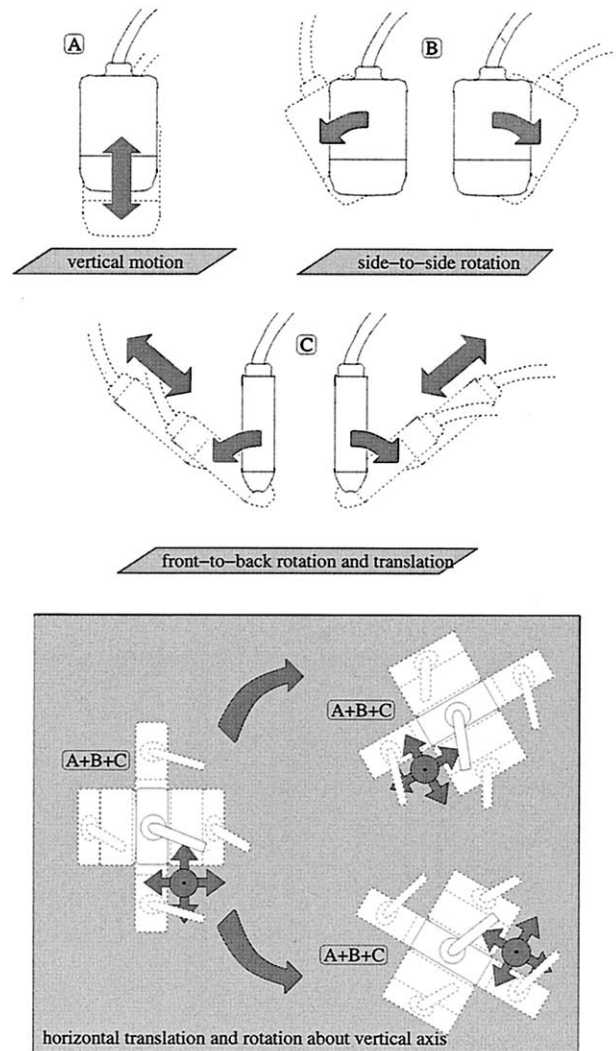


Fig. 7. Probe movement during spatial calibration. The sequence of movements A, B and C is repeated at three locations and orientations on the plane being imaged. This is necessary to define both the spatial calibration parameters and the location of the calibration plane.

the equations that must be solved to give the spatial calibration parameters, and ensure the suitability of the resulting parameters for normal clinical use.

Second, we have slightly adapted the pattern in which the probe is moved during calibration to constrain the calibration parameters reliably. This pattern, shown in Fig. 7, has been derived experimentally. Motion C is particularly important to distinguish between the  $z$  (out of plane) translation and rotation of the B-scan about a vertical axis.

#### *Depth change recalibration*

Although spatial calibration can be performed in around 10 min using this technique, it must be repeated

whenever the relative locations of the position sensor and the US probe, or the size or location of the B-scan in the video frame, change. This latter restriction implies that pan, zoom and depth changes require a new set of spatial calibration parameters and, hence, must be avoided during a scanning session; a constraint that is contrary to normal clinical practice. However, such changes do not affect the rotational calibration parameters; they only affect the translation ( $t_x$ ,  $t_y$ ,  $t_z$ ) and the scale ( $s_x$ ,  $s_y$ ). Indeed, changes to this subset of parameters are dependent on only three independent variables: these are the translation in the plane of the B-scan ( $b_x$ ,  $b_y$ ) and the zoom  $z$ .

In certain situations,  $b_x$ ,  $b_y$  and  $z$  can be estimated directly from the US image without the need for a full spatial recalibration. As long as the top and some of both sides of the B-scan data can be seen, we can estimate the location of the central top point of the B-scan ( $c_x$ ,  $c_y$ ) relative to the top left corner of the image, and the B-scan width ( $w$ ). This estimation is trivial for linear probes. For convex probes, the point ( $c_x$ ,  $c_y$ ) is the centre of curvature,  $w$  is the width at the top of the B-scan, and an additional parameter  $r$  is required to define the radius of curvature (Treece et al. 2002).

If these probe parameters were ( $c_{xo}$ ,  $c_{yo}$ ,  $w_o$ ) at the time the spatial calibration was performed, and are measured as ( $c_{xn}$ ,  $c_{yn}$ ,  $w_n$ ) after a change to the zoom or depth settings, then:

$$z = \frac{w_o}{w_n} \quad (1)$$

$$\Delta b_x = c_{xo} s_{xo} - c_{xn} s_{xn}, \Delta b_y = c_{yo} s_{yo} - c_{yn} s_{yn} \quad (2)$$

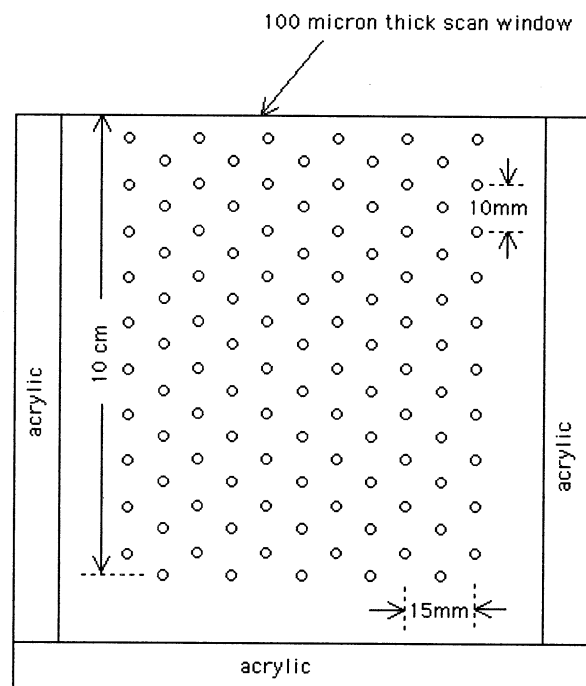
where ( $\Delta b_x$ ,  $\Delta b_y$ ) is the change in offset within the plane of the B-scan. Changes in the actual calibration parameters ( $t_x$ ,  $t_y$ ,  $t_z$ ,  $s_x$ ,  $s_y$ ) can be derived from these values<sup>9</sup>.

This means that, for most depth changes and some pan and zoom changes, the spatial calibration can be updated by analysing a single image from the US machine. Furthermore, this analysis (*i.e.*, the detection of the top and sides of the B-scan data) can be carried out by automatic image-processing algorithms (Treece et al. 2002).

## EXPERIMENTAL METHOD

### Ultrasound phantom

To establish the performance of the system, a highly accurate tissue-mimicking US phantom was required. Phantoms consisting of very thin nylon wires embedded



### PHANTOM WITH 110 2MM DIAMETER SPHERES

Fig. 8. Schematic of the US phantom, made from a tissue-mimicking substance containing a planar array of 110 precisely located, nonechogenic, 2 mm diameter spheres (Kofler and Madsen 2001; Madsen et al. 1998).

in a tissue-mimicking material are often used to determine the resolution of 2-D US machines. However, these phantoms are only designed to be scanned from one insonification angle. To test a freehand 3-D US system, we need a target that can be scanned from multiple angles, such as a small nonechogenic sphere. Phantoms containing such targets are used to determine lesion detectability for 2-D US systems. If we know the exact location of each sphere, then we can compare the known location with the apparent location from each of a sequence of freehand scans of the phantom and, therefore, deduce system accuracy in a (nearly) clinical setting.

Both the materials and the manufacturing process exist to make such a phantom, consisting of a grid of 2 mm diameter spheres that are located to an accuracy of  $\pm 0.1$  mm. A cubic phantom, with sides approximately 10 cm and spheres arranged as in Fig. 8, was designed and manufactured specifically for this purpose by Prof Madsen at the University of Wisconsin. The manufacturing process is described in Kofler and Madsen (2001) and the tissue-mimicking material used for the nonechogenic spheres in

<sup>9</sup> The spatial calibration translation ( $t_x$ ,  $t_y$ ,  $t_z$ ) is in the coordinate frame of the position sensor, not the B-scan.

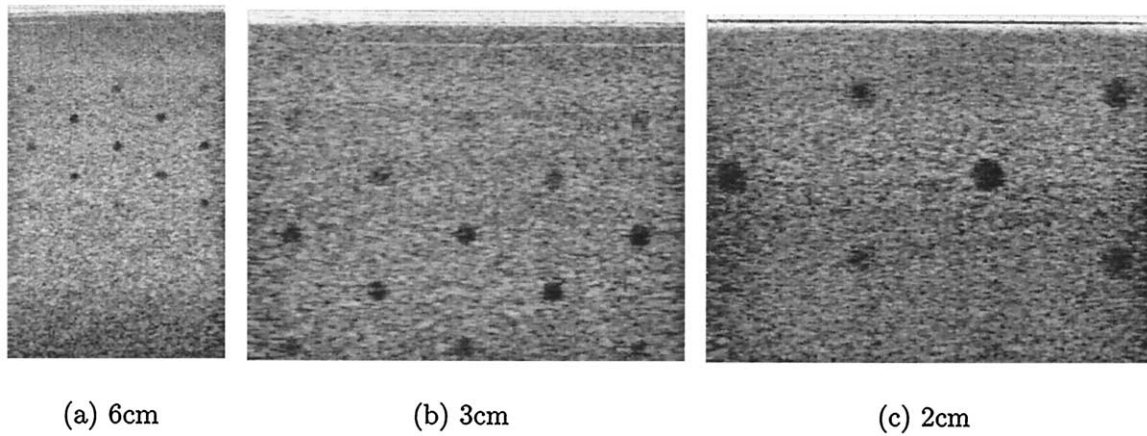


Fig. 9. Typical scans of the US phantom at varying depth settings; (a) and (b) scanned using a 5–10 MHz probe and (c) using a 10–22 MHz probe.

Madsen *et al.* (1998). The phantom has a water bath at its top so that it can be scanned from varying directions.

Figure 9 shows some typical B-scans of this phantom, at different frequencies and depth settings. The propagation speed and attenuation coefficient for the background material are  $1540 \text{ m s}^{-1}$  and  $0.55 \text{ dB cm}^{-1} \text{ MHz}^{-1}$  respectively, and for the spheres are  $1541 \text{ m s}^{-1}$  and  $0.53 \text{ dB cm}^{-1} \text{ MHz}^{-1}$ , measured at  $22^\circ\text{C}$  and 8 MHz. Close matching of these values ensures that the spheres generate only minimal distortion to the underlying image. Nevertheless, the backscatter of the sphere material is 40 dB lower than that of the background; hence, the spheres are clearly visible as black circles. Note that the spheres are not necessarily visible across the entire depth of the image; this is due to the beam thickness being greater than 2 mm at very shallow or very deep locations. The US machine settings (for in-

stance, time-gain compensation and transmit focusing) were optimised for each depth setting, to maximise the visibility of the spheres without reducing the frame update rate below 25 Hz. Typically, four or five focal points were used, arranged across the midregion of each B-scan. Having optimised these settings, they then remained constant for all experiments using the same depth setting.

For each test, the phantom was scanned with five patterns, shown in Fig. 10a to e. These represent the five basic motions of the probe during freehand 3-D US acquisition. Patterns (b) and (c) actually generate extended field-of-view data; although not strictly 3-D, these sort of data are freely available with freehand 3-D systems (Gee *et al.* 2003). A combined motion, as in Fig. 10f, was used to test the repeatability of mounting the position sensor on the probe.

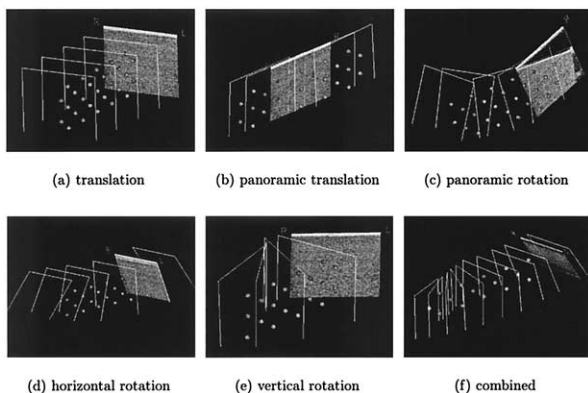


Fig. 10. Scanning patterns. A few sample B-scans, displayed as white “goal posts,” are shown to demonstrate the scanning pattern; typically 400 to 500 B-scans were acquired in each sequence.

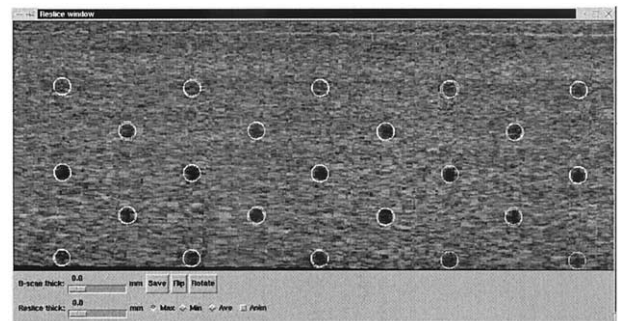


Fig. 11. Reslice through the phantom data set. Two white circles are shown for each sphere. One circle shows the location of each sphere as derived from the B-scan data. The other is a regular array of spheres with relative locations, as in Fig. 8, with global location and orientation optimised to match as closely as possible the set of data-derived spheres.

### Automatic detection of spheres

To be able to compare the location of a sphere as determined from a freehand 3-D data set with its true value, we need an accurate method of locating the centre of a sphere contained in a set of sequential B-scans. The number of measurements required (over 5000) prohibited the use of entirely manual sphere location. Hence, an automatic method was used to refine the location of each sphere centre, given an initial approximate location from the user clicking in a B-scan. The same approximate location could be used to initialise the algorithm for all sets of spatial calibration parameters (derived by the spatial calibration method outlined earlier, not from the location of the spheres), thus significantly reducing the number of manual interactions required.

For each candidate centre point on each B-scan within the sphere diameter (2 mm) of the initial ( $x, y$ ) point, specified by the user, the B-scan was multiplied locally with a 2-D mask, radius 1 mm, of the form  $m(x, y) = (1 - \sqrt{x^2 + y^2})$ , ( $0 \leq m \leq 1$ ). The point of minimum summed response to this mask was considered to represent the centre of the intersection of the sphere in any given B-scan. The 3-D location of the sphere centre was then calculated by assuming that the scans were approximately parallel in the locality of the sphere, and interpolating between the centres in the three B-scans with the minimum responses, using the same interpolation strategy as employed for subpixel interpolation in Treece et al. (2002).

Although the geometry of the phantom is known, the exact location of each sphere in the position-sensor coordinate system is not. Sphere locations derived from the US data were, therefore, compared to a regular grid of spheres, spaced as in Fig. 8 and globally aligned to the measured sphere centres using a nonlinear optimisation algorithm, Levenberg–Marquardt (More 1977). A typical result of this process is shown in Fig. 11.

There are several important points to note regarding this method of analysis.

- Only the alignment of the grid of spheres was optimised, not the scale parameters.
- The grid was optimised to all the visible sphere centres (*i.e.*, all those that could be detected by the algorithm described above); there were typically 20 or more in each scan.
- By definition, such analysis will result in a total mean location error in each of the  $x$ ,  $y$  and  $z$  dimensions of zero.
- The variance of the error will be slightly less than the population variance. Six parameters (the location and orientation of the plane containing the spheres) were derived from the data; therefore, we must multiply the mean-square-error by  $\frac{3n}{3n-6}$  ( $n$  being the number of

Table 1. Accuracy of automatic sphere detection

Frequency (depth)	Mean error (mm) 3-D	95% Confidence limits (mm)			
		$x$	$y$	$z$	3-D
5–10 MHz (3 cm)	0.16	$\pm 0.16$	$\pm 0.19$	$\pm 0.29$	$<0.35$

The mean error and 95% confidence intervals ( $\pm 95\%$ ) are given for sphere detection accuracy, estimated from 10 sphere location measurements.

points) to give an unbiased estimate of the variance.

Normally, we would only need to multiply by  $\frac{n}{n-1}$  to allow for the use of the mean value in each dimension.

- Given the adjustment above, the variance of the error gives an unbiased estimate of the location accuracy. It does not reveal any systematic errors in locating a feature with respect to the position-sensor coordinate system (which is what might be required for a US-guided biopsy, for instance). However, it does reveal the accuracy of locating a feature with respect to another feature within the data set, which is the important quantity for diagnostic medical imaging. Distance and volume measurement accuracy can also be derived from this result.

## RESULTS

Wherever confidence limits are given in the following analysis, these are derived from unbiased estimates of the population statistics, allowing for the quantity of measured data and the number of parameters derived from these data during the calculation. In each case, the coordinate system is aligned with the phantom, so that  $x$  is along the rows of spheres,  $y$  down the columns and  $z$  out of the plane of the spheres.

Three probe frequencies and depth settings were used in the following experiments. The 10–22 MHz probe was the highest frequency probe available and 2 cm was the most shallow depth setting. The 5–10 MHz probe was also tested at depth settings of 3 cm and 6 cm; these collectively cover the working range for musculoskeletal and arterial scans. The B-scan pixel size was approximately 0.05 mm (2 cm depth), 0.07 mm (3 cm depth) and 0.14 mm (6 cm depth).

### Automatic sphere detection

Before investigating the freehand system accuracy, the attainable accuracy of the 2-D US machine, the tissue-mimicking phantom and the sphere detection process were ascertained. The US phantom was scanned by mounting the 5–10 MHz probe in a rigid clamp attached to the movable section of a precision vice. With the

Table 2. Variation of spatial calibration parameter values

Frequency (depth)		$t_x$	$t_y$ (mm)	$t_z$	$\alpha$	$\beta$ (°)	$\gamma$	$s_x$ (mm/pixel)	$s_y$
5–10 MHz (6 cm)	mean	122.16	−80.47	−2.43	−111.17	−6.14	−1.11	0.1383	0.1382
	rms	0.13	0.11	0.25	0.03	0.25	0.35	0.0005	0.0003
	± 95%	± 0.27	± 0.21	± 0.50	± 0.05	± 0.51	± 0.71	± 0.0010	± 0.0007
5–10 MHz (3 cm)	mean	122.19	−80.04	−1.73	−111.32	−6.30	−0.88	0.0693	0.0699
	rms	0.14	0.18	1.06	0.06	0.39	0.64	0.0004	0.0004
	± 95%	± 0.28	± 0.36	± 2.13	± 0.12	± 0.77	± 1.28	± 0.0007	± 0.0007
10–22 MHz (2 cm)	mean	100.69	−68.46	0.76	−107.52	−3.89	0.29	0.0461	0.0461
	rms	0.10	0.12	0.20	0.04	0.18	0.62	0.0003	0.0002
	± 95%	± 0.20	± 0.23	± 0.39	± 0.08	± 0.35	± 1.25	± 0.0007	± 0.0003

20 spatial calibrations were performed for each probe frequency and depth setting. The mean, root-mean-square (rms) error and 95% confidence intervals (± 95%) are given for each of the estimated spatial calibration translations ( $t_x$ ,  $t_y$ ,  $t_z$ ), rotations ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) and scale factors ( $s_x$ ,  $s_y$ ).

B-scan plane parallel to the plane containing the array of spheres, the probe was moved in a direction orthogonal to this plane and the movement of the vice jaw (and, hence, the probe), measured with a micrometer. B-scans were recorded every 0.002" (0.0508 mm) so that each sphere was spanned by approximately 40 B-scans.

The location of all the detectable spheres was measured and compared with the true geometric arrangement, as given in Fig. 8. Results are shown in Table 1. The 3-D confidence limit of <0.35 mm includes errors in the formation of each B-scan, the location of the spheres in the phantom and the sphere-detection process. The greater error in the  $z$  (out of plane) dimension is due to the beam width and, possibly, also the construction of the phantom.

#### Spatial and temporal calibration precision

The precision of the temporal and spatial calibration procedures was analysed by comparing the results of

repeated calibrations. A total of 40 temporal calibrations were performed, with the same US machine settings, giving a root-mean-square deviation from the mean value of 4.9 ms, and a 95% confidence limit of ± 10.0 ms. This is clearly much better than the period between US frames (40 ms) or between position sensor readings (30 ms). The precision is improved further by using the average of 10 such calibrations; this was done before each of the spatial calibrations in the following experiments.

The precision of the spatial calibration was tested by repeating the calibration process 20 times for each probe frequency and depth setting. The position sensor remained attached to the probe throughout these experiments, but the probe was reattached to the Cambridge phantom each time. The temperature of the water bath was held approximately constant by replacing some of the water between calibrations.

Table 2 shows the mean, root-mean-square varia-

Table 3. Variation in B-scan location due to the spatial calibration

Frequency (depth)		Mean error (mm) 3-D	95% Confidence limits (mm)			
			$x$	$y$	$z$	3-D
5–10 MHz (6 cm)	centre	0.44	± 0.37	± 0.18	± 0.93	<0.78
	top left	0.29	± 0.27	± 0.21	± 0.50	<0.50
	top right	0.48	± 0.25	± 0.30	± 0.95	<0.84
	bottom left	0.54	± 0.52	± 0.18	± 1.08	<0.94
	bottom right	0.67	± 0.50	± 0.34	± 1.42	<1.18
5–10 MHz (3 cm)	centre	1.03	± 0.32	± 0.30	± 2.27	<1.80
	top left	0.94	± 0.28	± 0.36	± 2.13	<1.65
	top right	1.01	± 0.24	± 0.35	± 2.29	<1.77
	bottom left	1.07	± 0.48	± 0.28	± 2.35	<1.87
	bottom right	1.14	± 0.38	± 0.36	± 2.45	<1.99
10–22 MHz (2 cm)	centre	0.21	± 0.20	± 0.14	± 0.45	<0.36
	top left	0.21	± 0.20	± 0.23	± 0.39	<0.37
	top right	0.24	± 0.24	± 0.15	± 0.47	<0.42
	bottom left	0.25	± 0.22	± 0.20	± 0.53	<0.44
	bottom right	0.26	± 0.20	± 0.18	± 0.60	<0.46

The table shows the effect of the spatial calibration parameter variation in Table 2 on the location of the centre and corners of the B-scan. The mean 3-D error and 95% confidence intervals are given in each case.

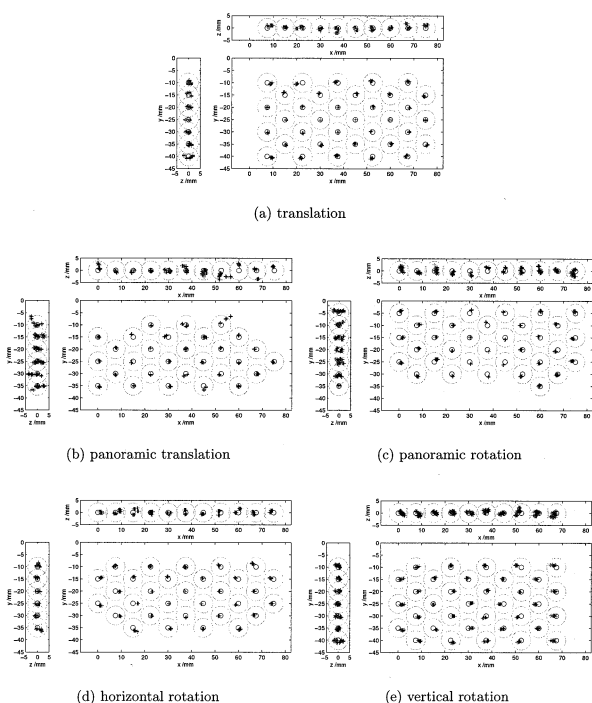


Fig. 12. System accuracy: 5–10 MHz probe, 6 cm depth. (a) to (e) the distribution of sphere locations for each of the 20 spatial calibrations in Table 2 applied to each of the scanning patterns in Fig. 10a to e. Bold circles show the actual sphere location and size; the location errors and grey circles have been magnified by a factor of 4 to show the distribution more clearly.

tion and 95% confidence limits for the eight spatial calibration parameters. The translations ( $t_x$ ,  $t_y$ ,  $t_z$ ) and

rotations ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) are defined in the coordinate system of the position sensor, not the B-scan plane. Due to the orientation of the position sensor on the probe,  $t_x$  and  $t_y$  were, nevertheless, approximately in the plane of the B-scan, and  $t_z$  orthogonal to it.  $\alpha$  represented rotation about an axis approximately normal to the B-scan plane,  $\beta$  about an approximately vertical axis and  $\gamma$  about an approximately horizontal axis. The scale factors ( $s_x$ ,  $s_y$ ) are defined within the B-scan plane.

It is immediately apparent from the results that the parameters in the plane of the B-scan ( $t_x$ ,  $t_y$ ,  $\alpha$ ,  $s_x$ ,  $s_y$ ) are much better determined than those out of the plane. In the worst case, the in-plane parameters are defined to within  $\pm 0.36$  mm,  $\pm 0.12^\circ$  and  $\pm 0.001$  mm/pixel, whereas those out of the plane are only defined to within  $\pm 2.13$  mm and  $\pm 1.28^\circ$ . This is a consequence of the US beam width, which, away from the foci, can exceed 4 mm for the 10–22 MHz probe and 8 mm for the 5–10 MHz probe. Use of the Cambridge phantom, as described earlier, limits the effect of beam width on the spatial calibration; which is why the out-of-plane accuracy is significantly better than the width of the US beam, but it does not suppress it entirely.

Of particular note in this respect is the combination of the out of plane translation  $t_z$  and rotation about a horizontal axis  $\gamma$ . These parameters can combine to generate movements of the B-scan plane that have only a small effect on the location of the calibration plane within the B-scan. Only with motion C in Fig. 7 is the image of the plane sensitive to combined changes in these two parameters, which is why this motion is an essential part of the calibration protocol. This is a par-

Table 4. Measurement accuracy

Frequency (depth)		Mean error (mm)		95% Confidence limits (mm)				
		2-D	3-D	$x$	$y$	$z$	2-D	3-D
5–10 MHz (6 cm)	Fig. 12a	–	0.27	$\pm 0.41$	$\pm 0.25$	$\pm 0.45$	–	$<0.51$
	Fig. 12b	0.19	–	$\pm 0.39$	$\pm 0.35$	–	$<0.38$	–
	Fig. 12c	0.24	–	$\pm 0.40$	$\pm 0.38$	–	$<0.49$	–
	Fig. 12d	–	0.32	$\pm 0.54$	$\pm 0.31$	$\pm 0.40$	–	$<0.60$
	Fig. 12e	–	0.35	$\pm 0.64$	$\pm 0.26$	$\pm 0.42$	–	$<0.65$
5–10 MHz (3 cm)	Fig. 13a	–	0.19	$\pm 0.33$	$\pm 0.21$	$\pm 0.25$	–	$<0.37$
	Fig. 13b	0.14	–	$\pm 0.26$	$\pm 0.21$	–	$<0.29$	–
	Fig. 13c	0.21	–	$\pm 0.41$	$\pm 0.27$	–	$<0.42$	–
	Fig. 13d	–	0.32	$\pm 0.62$	$\pm 0.26$	$\pm 0.35$	–	$<0.61$
	Fig. 13e	–	0.33	$\pm 0.64$	$\pm 0.21$	$\pm 0.45$	–	$<0.62$
10–22 MHz (2 cm)	Fig. 14a	–	0.22	$\pm 0.22$	$\pm 0.31$	$\pm 0.37$	–	$<0.44$
	Fig. 14b	0.17	–	$\pm 0.29$	$\pm 0.32$	–	$<0.37$	–
	Fig. 14c	0.22	–	$\pm 0.43$	$\pm 0.36$	–	$<0.49$	–
	Fig. 14d	–	0.25	$\pm 0.39$	$\pm 0.35$	$\pm 0.32$	–	$<0.50$
	Fig. 14e	–	0.22	$\pm 0.27$	$\pm 0.36$	$\pm 0.30$	–	$<0.46$

Summary results are shown for each of the graphs in Figs. 12, 13 and 14. Unbiased confidence limits are calculated, given the number of spheres in each case.  $x$ ,  $y$  and  $z$  are measured in the plane containing the phantom spheres, as for the graphic results. 2-D confidence limits are given for panoramic scans and 3-D confidence limits otherwise.

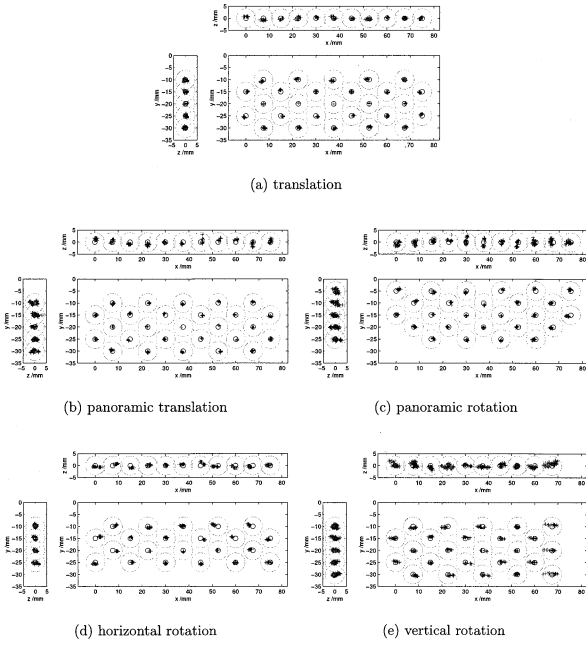


Fig. 13. System accuracy: 5–10 MHz probe, 3 cm depth. (a) to (e) are as in Fig. 12 and, once again, the location errors have been magnified by a factor of 4.

ticular problem for B-scans that are shallow, but have large beam widths, which may explain the results for the 5–10 MHz probe on a 3 cm depth setting, where the out-of-plane parameters are particularly poorly determined compared with those in the B-scan plane.

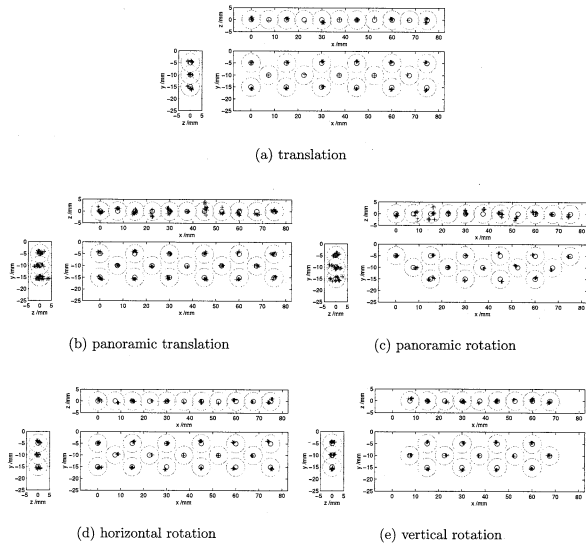


Fig. 14. System accuracy: 10–22 MHz probe, 2 cm depth. (a) to (e) are as in Fig. 12 and, once again, the location errors have been magnified by a factor of 4.

Table 5. Probe mounting repeatability: 10–22 MHz probe, 2 cm depth

Frequency (depth)	Mean error (mm) 3-D	95% Confidence limits (mm)			
		x	y	z	3-D
10–22 MHz (2 cm)	0.35	$\pm 0.53$	$\pm 0.41$	$\pm 0.53$	$<0.69$

The Table shows the variation in location errors for 20 spatial calibrations as in Fig. 2, each applied to 10 scans using the pattern in Fig. 10f. The optical tracker was remounted between each scan.

Table 3 shows the effect of this parameter variation on the location of the corner points and centre of the B-scan. The coordinate system in this table is aligned to the B-scan plane. As expected, the in-plane precision is significantly better than the out-of-plane ( $z$ ) precision. The bottom right of the B-scan, being farthest from the position sensor, is also slightly less well constrained. Nevertheless, a  $(x, y, z)$  precision at the centre of the B-scan of  $(\pm 0.20, \pm 0.14, \pm 0.45)$  is achieved for the highest resolution case, giving a 3-D confidence limit better than  $<0.5$  mm.

#### System accuracy

To assess the accuracy of the entire system for diagnostic US examinations, the phantom was scanned in each of the scanning patterns of Fig. 9a to e. For each probe frequency and depth setting, the five sweeps were recorded after the first 10 spatial calibrations. Sphere detection was then performed for all visible spheres in each of the five sweeps, with each of the 20 spatial calibration parameter sets.

The results are presented in graphic form in Figs. 12, 13 and 14, and summarised in Table 4. Each graph shows an orthographic projection of the distribution of (visible) sphere centres, so that the relative errors in each dimension, and across the depth of the B-scan, can be clearly seen. To make the location errors more visible, they have been magnified by a factor of 4. The small bold circles (representing the spheres) are drawn to scale, whereas the grey circles are also magnified by a factor of 4 for comparison with the location errors. Note that the scanning patterns in (b) and (c) are panoramic, so we do not expect to be able to estimate the  $z$  location from these scans with a great degree of accuracy. The  $z$  projections are given in these graphs for completeness only and 2-D, rather than 3-D, confidence limits are quoted in these cases.

As expected, the worst-case 3-D confidence limit improves with increasing probe resolution and decreasing depth setting, so that, with the 10–22 MHz probe, the point location accuracy within one scan is  $<0.5$  mm for any scanning pattern. This is similar to the precision of the spatial calibration, which implies that the spatial

Table 6. Accuracy of depth change recalibration

Frequency (depth)		Mean error (mm)		95% Confidence limits (mm)				
		2-D	3-D	x	y	z	2-D	3-D
5–10 MHz (6 cm)	Fig. 10a	–	0.30	$\pm 0.47$	$\pm 0.36$	$\pm 0.45$	–	<0.56
	Fig. 10b	0.16	–	$\pm 0.29$	$\pm 0.28$	–	<0.33	–
	Fig. 10c	0.19	–	$\pm 0.30$	$\pm 0.34$	–	<0.38	–
	Fig. 10d	–	0.29	$\pm 0.47$	$\pm 0.26$	$\pm 0.41$	–	<0.55
	Fig. 10e	–	0.39	$\pm 0.71$	$\pm 0.27$	$\pm 0.57$	–	<0.72
5–10 MHz (3 cm)	Fig. 10a	–	0.21	$\pm 0.29$	$\pm 0.26$	$\pm 0.29$	–	<0.39
	Fig. 10b	0.18	–	$\pm 0.28$	$\pm 0.31$	–	<0.37	–
	Fig. 10c	0.27	–	$\pm 0.50$	$\pm 0.40$	–	<0.56	–
	Fig. 10d	–	0.37	$\pm 0.70$	$\pm 0.37$	$\pm 0.35$	–	<0.69
	Fig. 10e	–	0.32	$\pm 0.56$	$\pm 0.32$	$\pm 0.42$	–	<0.61

Results are as in Table 4; however, the spatial calibrations used for the 3 cm depth setting were derived from those calculated for the 6 cm depth setting, and *vice versa*.

calibration is relatively unbiased. In all cases, scanning pattern (a) produced the most accurate results; in this orientation, the out-of-plane parameters of the spatial calibration have only a very limited effect on the geometry of the reconstructed data set. An error in  $t_z$ , for instance, would tend to move the entire data set bodily, which is irrelevant for this analysis. However, this is not the case for the rotations in scanning patterns (d) and (e). In a similar fashion, the panorama involving simple translation (b) generated more accurate results than that involving rotation (c).

In nearly all cases, the location errors were well within the radii of the spheres, particularly in the central region of the B-scan where the beam was focused. The

variation of accuracy with probe settings is, however, somewhat less than might have been expected from the threefold reduction in depth. This implies that the dominant errors are those not related to probe frequency and depth setting; possible culprits include the accuracy of the position sensor and temporal effects in the acquisition path.

#### Repeatability of position sensor mounting

The scanning pattern of Fig. 10f was used to assess the effect of reattaching the position sensor to the probe mount shown in Fig. 3. Ten scans were recorded using the highest resolution probe and lowest depth setting, removing and replacing the position sensor between each

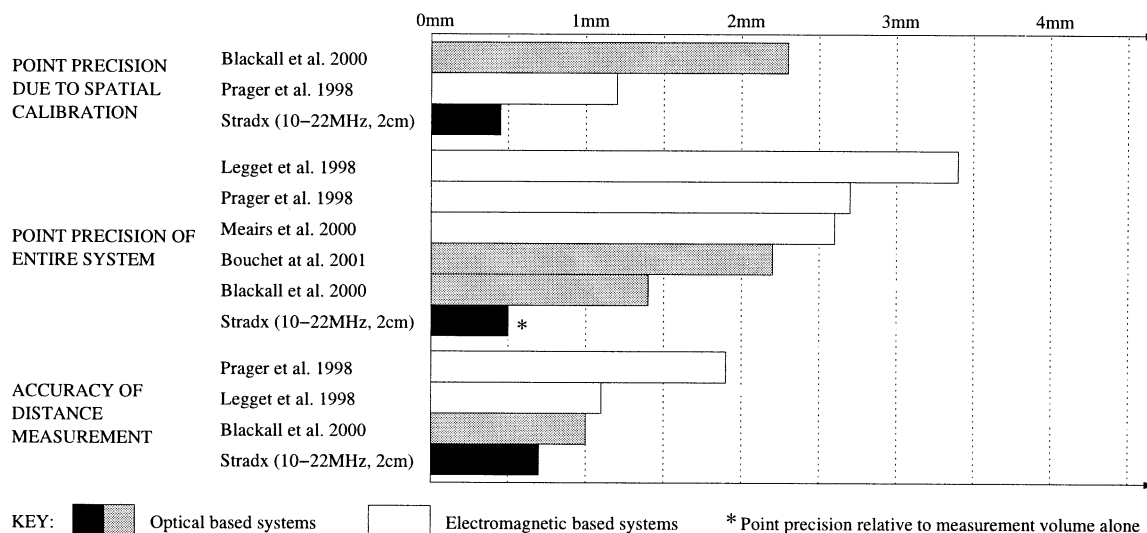


Fig. 15. Comparison of the 10–22 MHz Stradx system with other cited freehand 3-D US systems. The bar chart shows the 3-D confidence limits for various parameters of the systems. In most cases, this parameter has been estimated from alternative quoted values, as described in the Introduction. As a result, the comparison must be treated with some caution.



scan. These scans were then analysed as before, using each of the 20 spatial calibrations calculated for this probe frequency and depth setting.

The results are summarised in Table 5. The accuracy of the system is degraded only slightly by remounting the position sensor, from a 3-D confidence limit of <0.50 mm to <0.69 mm.

#### *Accuracy of depth change recalibration*

The system accuracy experiments were repeated for the 5–10 MHz probe on a 3 cm and 6 cm depth setting but, in each case, using the spatial calibration from the alternate depth setting, readjusted using the depth change recalibration algorithm. Table 6 shows the results in the same format as presented in Table 4. System accuracy is only slightly compromised by the use of this fast recalibration technique. For the 6 cm depth setting, the 3-D confidence limit is degraded from <0.65 mm to <0.72 mm and, for the 3 cm setting, from <0.62 mm to <0.69 mm.

### SUMMARY

Figure 15 shows a comparison of the results for the highest definition system (10–22 MHz probe and 2 cm depth setting) and the other systems cited in the introduction. The distance measurement accuracy for the system presented in this paper is approximately 2 times the point location accuracy because it is a measure of difference between two identical distributions.

As has been previously explained, the other systems' accuracies were assessed in slightly different circumstances and quoted in different forms. Although every effort has been made to convert each of them to a comparable form, the values should be regarded as indications of system accuracy only.

### CONCLUSIONS

Our system can be used to locate points within a freehand 3-D data set to an accuracy of < 0.50 mm, using a 10–22 MHz probe on a 2 cm depth setting. The accuracy with which distances can be measured within a data set is approximately  $\pm 0.7$  mm. This accuracy can be achieved by using the temporal and spatial calibrations outlined in this paper and, subsequently, leaving the probe settings and position sensor mounting unchanged. It is valid for all practical freehand scanning patterns. The necessary calibration procedures can be performed in only 10 to 15 min from mounting the position sensor on the probe and connecting the PC to the US machine.

System accuracy for probes covering a wide range of frequencies and depth settings appropriate for musculoskeletal and arterial scans is within <0.65 mm. Faster

spatial calibrations are possible if only the depth setting or certain pan and zoom settings have changed, in which case the system accuracy is degraded only slightly, to <0.72 mm. If spatial calibration is only performed each time the probe mount is attached to the probe (rather than each time the position sensor is attached to the mount), then the highest definition system accuracy drops slightly from <0.5 mm to <0.69 mm.

The achievable accuracy of this system is significantly better than that of any other such system presented in the literature. This is still the case if the position sensor is remounted or the spatial calibration recalculated based on depth changes alone; both of which are options that improve the system's usability. This performance owes much to the quality of the position sensor and the US machine, but also to the careful design and calibration of the freehand acquisition system.

*Acknowledgments*—This work was carried out under an EPSRC grant (GR/N21062). Dynamic Imaging Ltd. provided a modified US machine to enable digital data acquisition.

### REFERENCES

- Anderson ME, McKeag MS, Trahey GE. The impact of sound speed errors on medical ultrasound imaging. *J Acoust Soc Am* 2000; 107(6):3540–3548.
- Bailey JE, Bude RO, Tuthill T. US artifacts: Effects on out-of-plane US images reconstructed from three-dimensional data sets. *Radiology* 2001;218:592–597.
- Barratt DC, Davies AH, Hughes AD, Thom SA, Humphries KN. Optimisation and evaluation of an electromagnetic tracking device for high-accuracy three-dimensional ultrasound imaging of the carotid arteries. *Ultrasound Med Biol* 2001;27(7):957–968.
- Barry CD, Allott CP, John NW, et al.. Three-dimensional freehand ultrasound: Image reconstruction and volume analysis. *Ultrasound Med Biol* 1997;23(8):1209–1224.
- Belohlavek M, Foley DA, Seward JB, Greenleaf JF. 3D echocardiography: Reconstruction algorithm and diagnostic performance of resulting images. *Proc SPIE* 1994;2359:680–692.
- Berg S, Torp H, Martens D, et al.. Dynamic three-dimensional freehand echocardiography using raw digital ultrasound data. *Ultrasound Med Biol* 1999;25(5):745–753.
- Birkfellner W, Watzinger F, Wanschitz F, et al.. Systematic distortions in magnetic position digitizers. *Med Phys* 1998;25(11):2242–2248.
- Blackall JM, Rueckert DCR, Maurer J, et al. An image registration approach to automated calibration for freehand 3D ultrasound. In: *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MIC-CAI) 2000. Lecture Notes in Computer Science. Vol. 1935. 2000:462–471.*
- Bouchet LG, Meeks SL, Goodchild G, et al.. Calibration of three-dimensional ultrasound images for image-guided radiation therapy. *Phys Med Biol* 2001;46:559–577.
- Cardinal HN, Gill JD, Fenster A. Analysis of geometrical distortion and statistical variance in length, area, and volume in a linearly scanned 3-D ultrasound image. *IEEE Trans Med Imaging* 2000;19(6):632–651.
- Carr J, Prager RW, Treece GM, et al.. Design of a clinical freehand 3D ultrasound system. In: *Medical imaging 2000 (Ultrasonic imaging and signal processing). Proc SPIE. San Diego, CA 2000:3982:14–25.*
- Chuang ML, Hibberd MG, Beaudin RA, et al.. Patient motion compensation during transthoracic 3-D echocardiography. *Ultrasound Med Biol* 2001;27(2):203–209.

- Edwards WS, Deforge C, Kim Y. Interactive three-dimensional ultrasound using a programmable multimedia processor. *Int J Imaging Syst Technol* 1998;9(6):442–454.
- Fenster A, Downey DB, Cardinal HN. Three-dimensional ultrasound imaging. *Phys Med Biol* 2001;46:R67–R99.
- Fornage BD, Atkinson EN, Nock LF, Jones PH. US with extended field of view: Phantom-tested accuracy of distance measurements. *Radiology* 2000;214(2):579–584.
- Gee AH, Prager RW, Treece GM, Berman L. Narrow-band volume rendering for freehand 3D ultrasound. *Comput Graph* 2002;26(3):463–476.
- Gee AH, Prager RW, Treece GM, Berman L. Engineering a freehand 3D ultrasound system. *Pattern Recognition Lett* 2003;24:757–777.
- Goldstein A. Errors in ultrasound digital image distance measurements. *Ultrasound Med Biol* 2000;26(7):1125–1132.
- Kofler JM Jr, Madsen EL. Improved method for determining resolution zones in ultrasound phantoms with spherical simulated lesions. *Ultrasound Med Biol* 2001;27(12):1667–1676.
- Legget ME, Leotta DF, Bolson EL, et al.. System for quantitative three-dimensional echocardiography of the left ventricle based on a magnetic-field position and orientation sensing system. *IEEE Trans Biomed Eng* 1998;45(4):494–504.
- Madsen EL, Frank GR, Dong F. Liquid or solid ultrasonically tissue-mimicking materials with very low scatter. *Ultrasound Med Biol* 1998;24(4):535–542.
- Meairs S, Beyer J, Hennerici M. Reconstruction and visualization of irregularly sampled three- and four-dimensional ultrasound data for cerebrovascular applications. *Ultrasound Med Biol* 2000;26(2):263–272.
- More JJ. The Levenberg–Marquardt algorithm: Implementation and theory. In: Watson A, ed. *Numerical analysis. Lecture Notes in Mathematics* 630. Springer-Verlag, 1977; 105–116.
- Palombo C, Kozakova M, Morizzo C, et al.. Ultrafast three-dimensional ultrasound—application to carotid artery imaging. *Stroke* 1998;29:1631–1637.
- Prager R, Gee A, Treece G, Berman L. Freehand 3D ultrasound without voxels: Volume measurement and visualisation using the Stradx system. *Ultrasonics* 2002;40(1–8):109–115.
- Prager RW, Gee AH, Berman L. Stradx: Real-time acquisition and visualisation of freehand 3D ultrasound. *Med Image Analysis* 1999;3(2):129–140.
- Prager RW, Rohling RN, Gee AH, Berman L. Rapid calibration for 3-D free-hand ultrasound. *Ultrasound Med Biol* 1998;24(6):855–869.
- Rohling RN, Gee AH, Berman L. A comparison of freehand three-dimensional ultrasound reconstruction techniques. *Med Image Anal* 1999;3(4):339–359.
- Treece GM, Prager RW, Gee AH, Berman L. Fast surface and volume estimation from non-parallel cross-sections, for freehand 3-D ultrasound. *Med Image Anal* 1999;3(2):141–173.
- Treece GM, Prager RW, Gee AH, Berman L. Correction of probe pressure artifacts in freehand 3D ultrasound. *Med Image Anal* 2002;6(3):199–215.
- Wang S, Chen RK, Cardinal E, Cho K-H. Joint sonography. *Radiol Clin North Am* 1999;37(4):653–668.