A Real-Time Ultrasound Calibration System with Automatic Accuracy Control and Incorporation of Ultrasound Section Thickness

5 authors, including:

Mehdi Hedjazi Moghari
Harvard Medical School
54 PUBLICATIONS   **821** CITATIONS

SEE PROFILE

Randy Ellis
Queen's University
256 PUBLICATIONS   **3,377** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Medical Image Visualization View project

Phenol-based Soft Embalming View project

A Real-Time Ultrasound Calibration System with Automatic Accuracy Control and Incorporation of Ultrasound Section Thickness

Thomas Kuiran Chen[a]   Adrian D. Thurston[a]   Mehdi H. Moghari[b]   Randy E. Ellis[a,c]   Purang Abolmaesumi[a,b]

[a]School of Computing, Queen's University, Kingston, Canada;
[b]Department of Electrical and Computer Engineering, Queen's University, Kingston, Canada;
[c]Division of Orthopaedic Surgery, Kingston General Hospital, Kingston, Canada

ABSTRACT

This paper presents a real-time, freehand ultrasound (US) calibration system, with automatic accuracy control and incorporation of US section thickness. Intended for operating-room usage, the system featured a fully automated calibration method that requires minimal human interaction, and an automatic accuracy control mechanism based on a set of ground-truth data. We have also developed a technique to quantitatively evaluate and incorporate US section thickness to improve the calibration precision. The experimental results demonstrated that the calibration system was able to consistently and robustly achieve high calibration accuracy with real-time performance and efficiency. Further, our preliminary results to incorporate elevation beam profile have demonstrated a promising reduction of uncertainties to estimate elevation-related parameters.

Keywords: ultrasound imaging, real-time calibration, accuracy control, beam width, section thickness, computer-assisted surgery

1. INTRODUCTION

3D ultrasound (US) imaging has seen increasing applications in intraoperative guidance of computer-assisted surgery. Medical US is non-ionizing, compact, portable, relatively inexpensive, and capable of imaging in real time. To construct a high-resolution 3D US image of the patient's anatomy from a set of 2D images, a tracked 2D US probe that allows image acquisition in an unconstrained "freehand" motion is a useful alternative to direct volume imaging. Tracking is typically achieved by rigidly affixing the probe with a localizer traced by a position sensing system. However, knowing the position of the US probe alone is not adequate to determine the positions of the acquired 2D images. The relationship between these two coordinate frames can be calculated through the process of US probe calibration, where a homogeneous transformation is estimated to map the position of individual pixels from the US image frame to the US probe frame. With the latter being tracked in real time by the tracking system, we are able to obtain the physical positions of those pixels in the world coordinate frame, which is normally fixed to the imaging anatomy. Calibration is therefore a fundamental step and a single point of failure in a freehand US imaging system. A recent overview of the US calibration techniques may be found in Ref. 1.

Most current calibration techniques focus primarily on precision and accuracy but, for computer-assisted surgeries that may require calibration inside an operating room (OR), many other aspects must be considered.

First, intraoperative use needs a calibration procedure that is fast, robust and easy to perform. Overall, calibration techniques may be divided into either an iterative approach or a closed-form solution.[1] Iterative approaches are, in general, less robust than closed-formed solutions because of the non-guaranteed convergence, local minima, and sensitiveness to initial estimates. Also, to achieve a similar accuracy, iterative methods typically need more input data and computational time than closed-form techniques. For instance, calibration with the *Cambridge* phantom[2] (an iterative approach) would require at least 550 images to achieve acceptable

1

accuracy, as compared to around 6–30 images with N-wire phantoms[3–7] (closed-form solutions). On the other hand, closed-form solutions must accurately extract point-targets from an US image, which is often done manually (undesirable in an OR setting). Some successful attempts have been made by Lindseth *et al.*[4] and Hsu *et al.*[7] to automate segmentation on images acquired from N-wire phantoms, but their approaches all require human interferences to certain extent: i.e., the former requires finding an image point in a manually specified region as an initialization of the algorithm, and the latter needs to know the actual scale factors of the US image which have to be manually measured on the US machine.

Second, it is desirable to have automatic, real-time feedback of calibration accuracy in the OR. Calibration and validation are conventionally two-phase tasks that are isolated: a calibration is performed, followed by a validation, so the only way for a surgeon to improve an inaccurate calibration is to recalibrate (which is not only time-consuming but lacks an inherent accuracy check). Boctor *et al.* were among the first to introduce a real-time *in-vivo* quality control mechanism that monitored the consistency in calibration parameters through frequent recalibration in the background.[8,9] However, their validation on the calibration results was based on precision and not accuracy. It is important to note that a measure of precision is quite different from that of accuracy:[1] precision defines the repeatability and consistency of the system, while accuracy evaluates how much the output is away from a known "ground truth" (typically measured independently). Not relying on a ground truth, a high precision (a low variance in results) does not necessarily guarantee a high accuracy. For example, it is possible that a calibration system that achieves highly consistent results may include a systematic error that renders the system inaccurate.

A further challenge to the calibration and use of tracked US images is the US elevation beam width, also referred to as the section thickness.[10,11] This out-of-plane beam width not only causes false contours in the image that do not correspond to any real tissues,[12] but also leads to localization errors in calibration, image segmentation and registration. Hsu et al. were the first to improve the elevation estimate in calibration using a precisely positioned membrane.[7] But no quantitative beam profile was given and the procedure was difficult to work with.

Our research goal was to address all of the aforementioned concerns. We designed and tested a real-time US calibration system that needed minimal human interaction, and was equipped with automatic error retrieval and accuracy control based on a known ground truth. An important part of the system was a *Double-N* calibration phantom of a simple, sterilizable design. We also developed a *Bridge* phantom to quantitatively evaluate and incorporate the elevation beam profile to improve the calibration precision. Extensive validations were conducted to thoroughly investigate the accuracy, robustness, and performance of the calibration system.

2. METHODOLOGY

Figure 1(a) shows the design of our calibration system consisting five successive stages: US images were continuously acquired from the *Double-N* phantom. The position of the US probe was tracked by a camera system. The pixel locations of the cross section of N-wires in the US images (denoted as N-fiducials) were then extracted in real time by a fully automatic segmentation algorithm. The segmented N-fiducials and their corresponding physical coordinates collected in the phantom space, were fed to a closed-form formula to calculate the calibration parameters. Measured by fiducial registration error (FRE) against a known ground truth, the accuracy of the current calibration result was fed back to the control loop to determine if it was satisfactory. The FREs were updated, monitored and displayed in real time to the user. Once FRE converged or reached an acceptable level, an interactive interface was provided to the user to terminate the process and save the calibration result.

Figure 1(b-d) demonstrates our experimental setup for the calibration system: US images were generated by a General Electric (Mississauga, ON, Canada) Voluson 730 US machine (Figure 1(b)) then fed to an ATI (Markham, ON, Canada) AIW-7500 frame grabber. A Traxtal (Toronto, ON, Canada) VersaTrax Active Tracker was mounted on the US probe and tracked by a Northern Digital (Waterloo, ON, Canada) Optotrak Certus (Figure 1(c)). The US probe used in our experiments was a 192-element GE SP10-16 Wide Band Linear Array (Figure 1(d)) that operated at a central frequency of 12.5MHz with imaging depth of 3.4 cm and 4 focal zones activated. The frame rate for image acquisition was 24Hz. The central processor was a Dell Optiplex GX270 workstation (Intel Pentium 4 2.6GHz and 2GB SDRAM) running Microsoft Windows XP Professional.
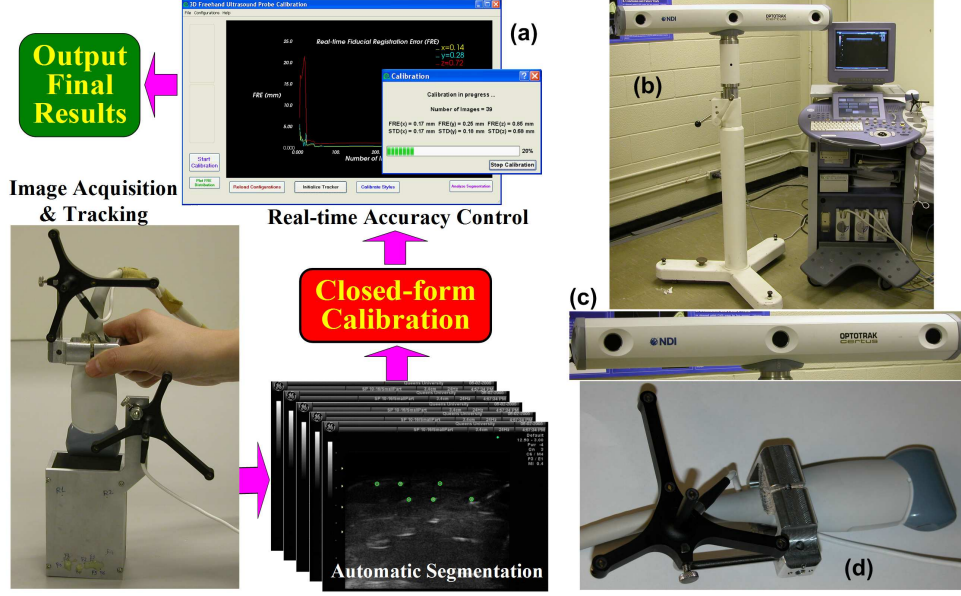
Figure 1. The design and hardware setup of the real-time automatic US calibration system.

Further, we improved the calibration precision by incorporating the US elevation beam width measured with a *Bridge* phantom (Figure 2(a)) that utilizes the section-thickness artifacts of an inclined plane.[13] Evaluation of US imaging quality and resolution in ultrasound physics and instrumentation[11] is commonly described in the axis of axial (direction of sound travel), lateral (in plane and perpendicular to the beam axis), and elevation (out of plane), all with respect to the crystal array (Figure 2(b)). The axial and lateral axes form the scan (imaging) plane while the lateral and elevation axes define the elevation plane. The US beam has finite width along both lateral and elevation directions (Figure 2(c)).
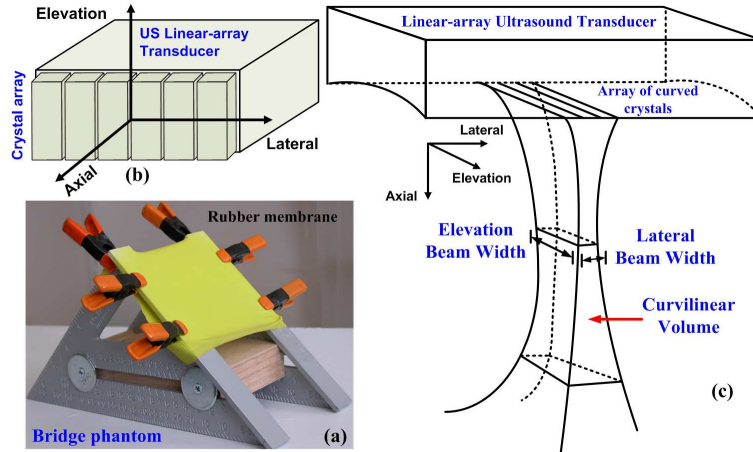

Figure 2. Bridge phantom and US beam profile in *Axial*, *Lateral* and *Elevation* axes.

The lateral (in-plane) beam width is determined by focusing in the scan plane which is electronically controllable and typically achieved by offsetting the firing of multiple crystals in a group by a very small amount of time delay[11](e.g., 10^{-9} second). On the other hand, since there is no multiple crystals in the elevation plane for a 1-D linear-array transducer to achieve electronic focusing, the US beam is only focused mechanically in the elevation direction, by either curving the crystal or placing an acoustic lens in front of the crystals. The elevation beam pattern are therefore solely determined by the size and curvature of the crystal or the acoustic lens, with sharp focus only possible at a narrow axial distance to the transducer. As a result, the beam profile for a linear array is not symmetrical. The beam width is typically much larger elevationally than laterally, forming a curvilinear volume (Figure 2(c)) that gives rise to more uncertainties and localization errors in the elevation

direction. Because the elevation beam profile is constant for a given transducer construction and operational frequency,[11] we measured the section thickness and incorporated it into our real-time calibration system to enhance the calibration precision using an Unscented Kalman Filter (UKF) algorithm.[14]

3. THE FULLY AUTOMATIC, REAL-TIME CALIBRATION SYSTEM

3.1 The *Double-N* Calibration Phantom

The *Double-N* phantom consisted of a front and a back plate connected by two side walls, forming an open-ended box (Figure 3(a,b)). For effective sterilization, the phantom could be quickly disassembled and reassembled using an *L*-Key screwdriver. To facilitate easy manufacture, ensure rigidity and durability, and reduce distortion in the phantom geometry during reassembling, the plates are designed in plain rectangle-shape and made of 5 mm-thick aluminum. The front plate had an extended arm to mount a spatial localizer for tracking purposes.
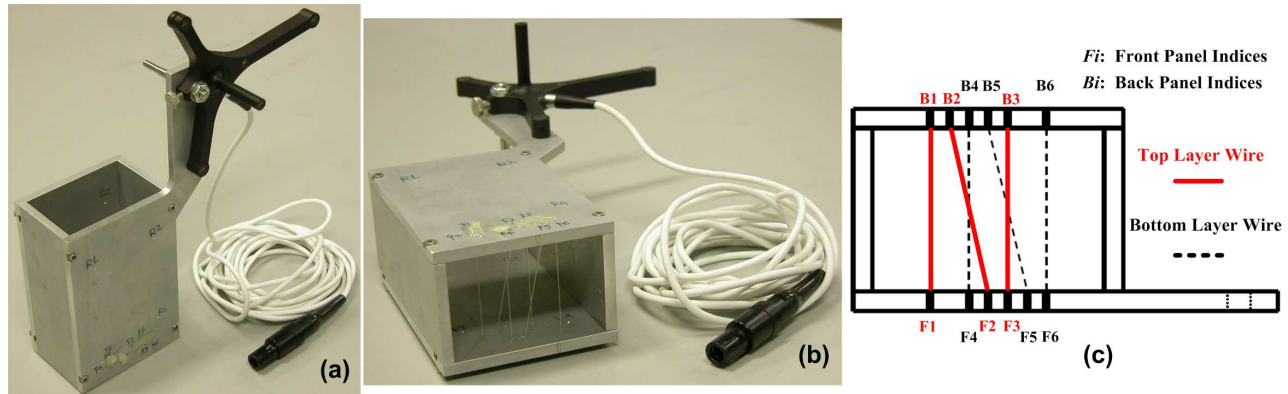

Figure 3. Front (a), side (b) views and design (c) of the *Double-N* phantom.

Figure 3(c) shows the design of the *Double-N* phantom. There are only two layers of N-wires, shifted horizontally to improve imaging and reduce the occurrence of reverberance artifacts that may arise if more than two layers are used.[15] With the nylon wire selected by empirically imaging various diameters, the *Double-N* phantom produces remarkably clean and well defined US images of N-fiducials that facilitated automated segmentation.

3.2 Fast, Robust, Fully Automatic Segmentation

Demanding no human interference at all, we developed a fast and robust algorithm to automatically segment images acquired from the *Double-N* phantom. The technique took advantage of two unique geometric features of N-fiducials in the image to assist the segmentation: the three *collinear* dots that form a N-wire intersection with the US image plane, and the two nearly *parallel* lines that pass through these two layers of dots.

Our segmentation algorithm had four stages with various image processing techniques involved. First, dominant speckles were removed by a series of morphological operations. Pixels were then clustered to yield possible candidates of dots. Now we probed for lines composed of three dots, and further narrowed the target down to a pair of two such lines close to parallel. Figure 4 shows an example of an accurate segmentation result given a test image abundant with speckles and noise.

3.3 Automatic Accuracy Feedback and Control

Direct evaluation of calibration accuracy can be challenging because of the lack of a reliable way to obtain the exact spatial relationship between the US image plane and the probe. Similar to the approach of Pagoulatos et al.,[3] we employed the *Double-N* phantom to test our calibration accuracy. We were able to fully automate the error evaluation in real time by iteratively reconstructing the N-wire positions in the physical phantom space using the current calibration results, and then compared them to a ground truth – the known wire locations from the *Double-N* phantom geometry. This is commonly referred to as "Fiducial Registration Error (FRE)",[1] because it shares the same principle as the use of fiducials in known positions to judge a registration error only that, instead of point fiducials, line fiducials from the construction of N-wires were used here.
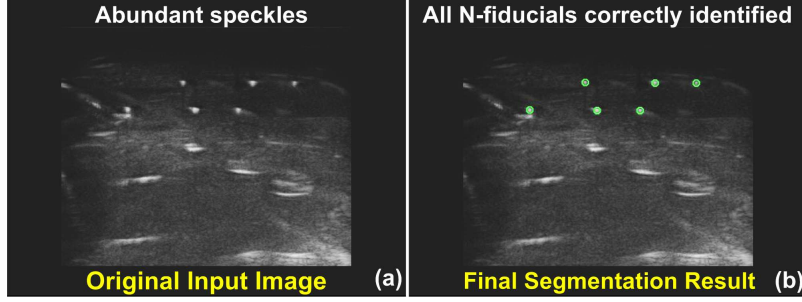
4

Figure 4. Robust and fully automatic segmentation.

Since FRE is an absolute distance between two positions in space, it remains invariant to frame transformations. Hence, for easier visualization of the error distributions, we converted the FRE from the phantom space (H) to the physical transducer frame defined with respect to the alignment of crystal arrays (C) as in Figure 2(b), to monitor its change in *axial*, *lateral* and *elevation* components. FREs were updated and displayed in real time to the user via a graphical interface (Figure 5(a)). After all FRE curves either converged or fell below a desired threshold, the calibration system provided an interactive interface (Figure 5(b)) for the user to terminate the process and output the final calibration result.
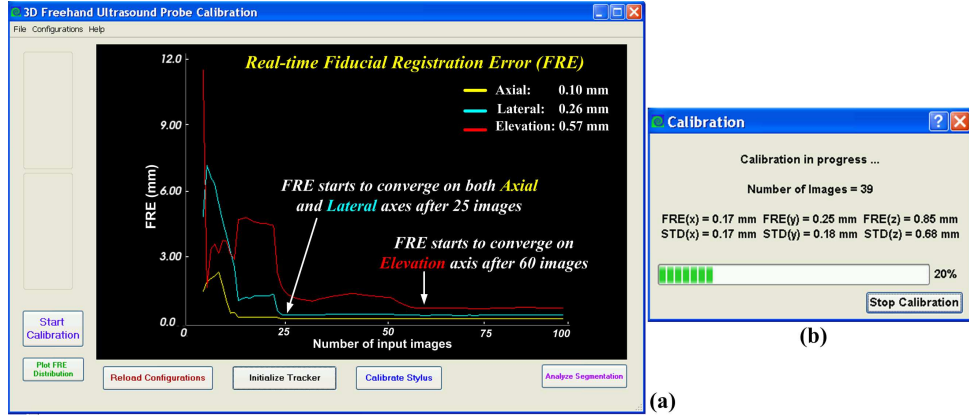

Figure 5. Real-time calibration accuracy feedback and control.

4. EVALUATION & INCORPORATION OF US ELEVATION BEAM PROFILE

4.1 Quantitative Evaluation of US Elevation Beam Profile

To quantitatively evaluate the US elevation beam width, we adopted a method based on the principle originally proposed by Goldstein.[13] The basic idea is to take advantage of the section-thickness artifacts generated when a sound beam interrogates a diffusive (non-specular) inclined plane.

Figure 6(a) shows a cross-section view of a diffusive inclined surface being interrogated by a linear-array transducer in the elevation plane. As we have discussed earlier, a typical US beam is mechanically focused in the elevation plane by either curving the crystals or using an acoustic lens. Because of the finite elevation beam width, the sound beam will intersect the inclined surface from A to B in the elevation plane. The diffusive nature of the surface AB would scatter the sound energy omnidirectionally, so some signals would be detected by the transducer. As the sound wave propagates, the first echo would be generated when the wave hits position A, which is the closest axial distance to the transducer; the last echo would be from B, the position most distant to the transducer. However, because the US machine algorithms always assume that all echoes received are from reflectors located on the central beam axis,[11] the echoes from AB would be interpreted as if they were from CD on the central beam axis. As a result, the US machine would display a thick echo band with its height the length of CD. Furthermore, if the angle of inclination is 45 degrees (i.e., $\theta = \pi/4$), the length of CD (the axial height

5

of the echo band) would actually be equal to the effective elevation beam width (section thickness) at that axial depth.[11,13]
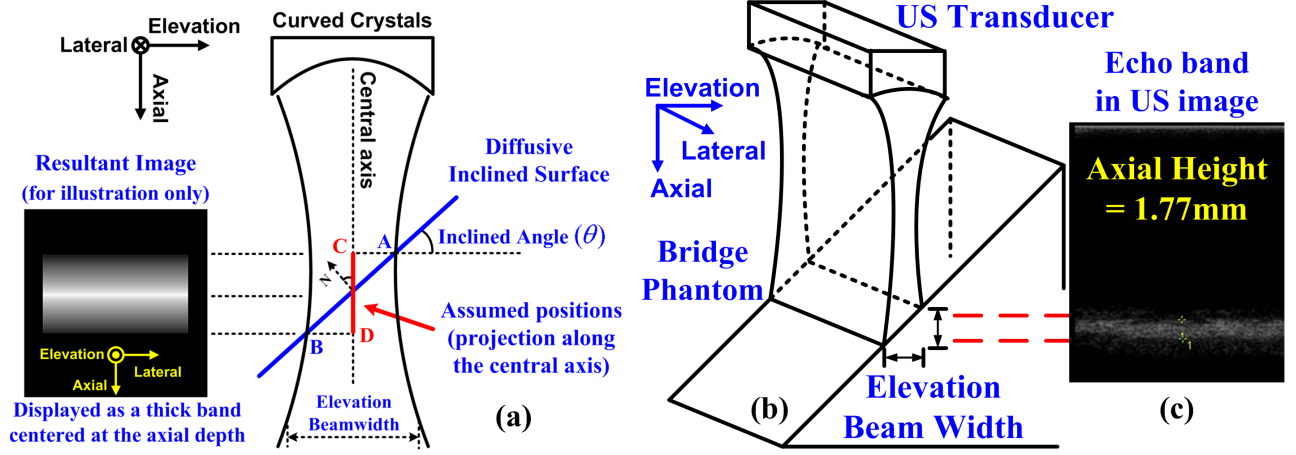

Figure 6. Principles of section-thickness measurement using an inclined plane.

Based on this principle, we designed the *Bridge* phantom (Figure 2(a)) to form a 45-degree inclined plane that was made of 0.35 mm-thick rubber membrane. Rubber was an excellent choice for the inclined surface because it has an acoustic impedance (1.81M Rayles) roughly similar to that of water (1.48M Rayles), making it a weak but adequate sound reflector in water. The membrane was a diffusive reflector that would backscatter echoes to the transducer, which was positioned perpendicular to the horizon (Figure 6(b)). Hence, by measuring the axial height of the echo band in the display (Figure 6(c)), we were able to obtain the section thickness at that specific axial depth. Consequently, we made a series of such measurements at various depths to estimate the relevant elevation beam profile.

4.2 Incorporation of US Elevation Beam Profile using UKF

To incorporate the section-thickness information, we developed a method based on Unscented Kalman Filter (UKF)[14] to calculate the calibration parameters. Our goal here is to incorporate the elevation beam width into the UKF algorithm, while determining the calibration parameters, namely, the rotation matrix, $\mathbf{R}_{(\theta_x, \theta_y, \theta_z)}$, the translation vector, $\mathbf{T}_{(t_x, t_y, t_z)}$, and the scaling matrix $\mathbf{S}_{(s_x, s_y)}$, so that we can minimize the calibration error, serving as our cost function, $\mathcal{C}$:

$$\mathcal{C}_{(\theta_x, \theta_y, \theta_z, t_x, t_y, t_z, s_x, s_y)} = E[|\mathbf{p}_{pr} - \mathbf{R}_{(\theta_x, \theta_y, \theta_z)} \mathbf{S}_{(s_x, s_y)} \mathbf{p}_{us} - \mathbf{T}_{(t_x, t_y, t_z)}|^2] \tag{1}$$

For the sake of observation, we define a state vector, $\mathbf{x}$, that houses all the calibration parameters to estimate:

$$\mathbf{x} = [t_x, t_y, t_z, \theta_x, \theta_y, \theta_z, s_x, s_y]^T \tag{2}$$

In addition, let us further define a state model to iteratively update the state vector:

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathcal{N}(0, \mathbf{\Sigma_Q}), \tag{3}$$

where $\mathcal{N}(0, \mathbf{\Sigma_Q})$ represents a zero mean Gaussian random vector with covariance matrix $\mathbf{\Sigma_Q}$.

Finally, to minimize the cost function in Equation 1, we also need to define an observation model:

$$\mathbf{p}_P^{1:i} = \mathbf{R}_{(\theta_\mathbf{x}, \theta_\mathbf{y}, \theta_\mathbf{z})} \mathbf{S}_{(s_x, s_y)} \mathbf{p}_U^{1:i} + \mathbf{t}_{(t_x, t_y, t_z)} + \mathcal{N}(0, \mathbf{\Sigma_R}^{1:i}) \tag{4}$$

where, $i = 1, ..., N$, $\mathbf{p}_P^{1:i} = [\mathbf{p}_P^{1T}, ..., \mathbf{p}_P^{iT}]^T$ is the observation vector composed by our data positions ($^P X$) in the US probe frame, $\mathbf{p}_U^{1:i} = [\mathbf{p}_U^{1T}, ..., \mathbf{p}_U^{iT}]^T$ is the input vector made of the data positions ($^U X$) in the US image frame, and $\mathcal{N}(0, \mathbf{\Sigma_R}^{1:i})$ is the point localization error for points 1 to i.

6

The last term in Equation 4 should interest us the most because it is where the information about the elevation beam width comes into play. $\mathcal{N}(0, \mathbf{\Sigma}_{\mathbf{R}}^{1:i})$ is modeled by a zero mean Gaussian 3×1 random vector with the covariance matrix $[\sigma_x^2 \; \sigma_y^2 \; \sigma_z^2] \cdot I_{3 \times 3}$, where I is the identity matrix and σ_x^2, σ_y^2 and σ_z^2 are the variances of noise in data along the lateral, axial and elevation axes of the US scan plane, respectively. Hence, what $\mathcal{N}(0, \mathbf{\Sigma}_{\mathbf{R}}^{1:i})$ really does in Equation 4 is to estimate the point-localization errors in the data sets from the US image frame and then add their influence to the data sets in the US probe frame. Because the point-localization errors in question mainly come from the elevation axis (the z-axis in the equations), with their range directly proportional to the effective elevation beam width, it is reasonable to assume the quantitative elevation beam profile may offer a good estimate of the variance of the errors (i.e., σ_z^2) at the corresponding axial depths in the elevation direction. Hence, we may embed this variance information into Equation 4 (which is essentially a nonlinear function of the state vector $\mathbf{x}$), and then employ the UKF algorithm to estimate $\mathbf{x}$ (the calibration parameters). The UKF algorithm would iterate through the following steps until it converges to a final solution:

1. Predict the state vector, $\mathbf{x}$, and its covariance matrix from the state model (Equation 3);
2. Append a point from the US image coordinate frame, $\mathbf{p}_U^i$, to the previously collected points in that frame, $\mathbf{p}_U^1 \cdots \mathbf{p}_U^{i-1}$, to estimate their corresponding positions in the US probe coordinate frame by using the predicted state vector in Step #1;
3. Compute the error in the estimated positions of the points in the probe coordinate frame to update the state vector and its covariance matrix.

From one perspective, one can say that UKF is essentially a weighted least-squares method. However, it differs from a typical least-squares solution in that it also takes into account the variance of the input data during optimization and can also compute the calibration parameters and variances simultaneously.

5. RESULTS AND VALIDATIONS

5.1 Segmentation Accuracy and Robustness

For a test of segmentation accuracy, we randomly selected 100 images acquired from the *Double-N* phantom and visually inspected them for the locations of the six N-fiducials. This gave us a total of 600 independent measurements as the ground truth to compare to the results suggested by the automatic segmentation. The algorithm correctly identified all 600 N-fiducials positions (100% recognition rate). More importantly, the results were precise to one tenth of a pixel, whereas manual segmentation had a precision of only one pixel.

Performance-wise, aside from being able to accurately and robustly extract the N-fiducials, the segmentation algorithm is fast and runs in real time. Implemented in C++, and without hardware or software optimization, it only took the algorithm on average 0.17 seconds to segment a single image.

5.2 Calibration Accuracy, Robustness and Performance

One advantage of our fast segmentation algorithm was that it enabled us to easily validate the calibration system on a large set of US images acquired under various experimental conditions.

- We tested the calibration system on a set of $10,000$ images, acquired from the *Double-N* phantom in freehand motion. The images were captured from a wide range of angles and depths that the US probe could be physically placed within the phantom.
- Data were randomly divided into 50 groups of 200 images in each. This established 50 independent trials for testing. To avoid bias, the data for validation were separated from those of calibration.
- For validation, we estimated the FRE from the first 80 images in each set, computing the means and standard deviations per set.
- For calibration, images were taken from the other 120 images in each set and then fed to the calibration system to calculate the calibration parameters.

Figure 7 shows the mean of FREs along the axial, lateral and elevation axes for all 50 calibration trials. Means and standard deviations of FREs and the best and worst cases of all trials are summarized in Table 1.

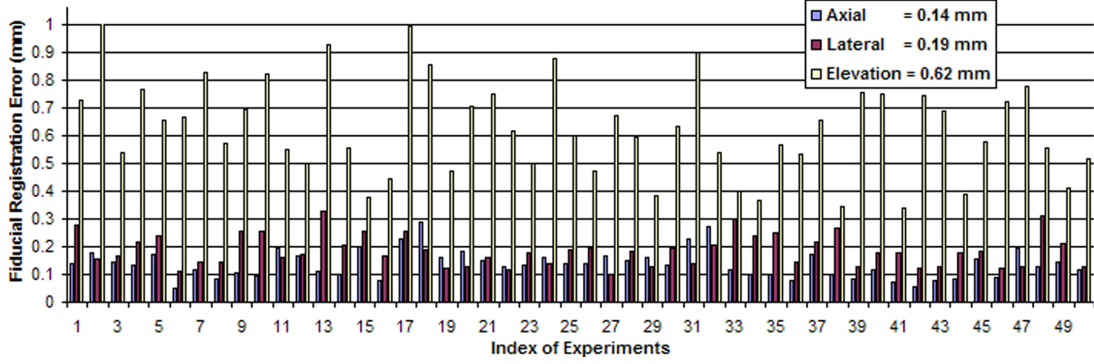Observations on the calibration accuracy and robustness results include:

Figure 7. Comparison of converged FREs for all 50 calibration trials.
Table 1. FRE for 50 independent calibration trials.

Number of Trials:	Fiducial Registration Error (mm)			
50	**axial**	**lateral**	**elevation**	**norm**
Mean (μ)	0.14	0.19	0.62	0.66
Standard Deviation (σ)	0.12	0.22	0.64	-
Minimum (best case)	0.05	0.10	0.34	0.36
Maximum (worst case)	0.29	0.33	1.02	1.11

1. In 49 out of the 50 independent experiments, convergence was to sub-millimeter accuracy in all of the axial, lateral and elevation axes; the remaining set had a FRE slightly above 1 mm. This indicates the calibration system was able to achieve high accuracy in 98% of cases.

2. The averaged FRE of all 50 independent trials was 0.66mm (Table 1), which is consistent with accuracy reported by related work with N-wire phantoms.[3, 5, 7]

3. The elevation axis had a much larger mean FRE (more than three times larger) than that of the axial and lateral directions, and also a larger standard deviation of FRE. The axial direction always had the smallest overall errors in both mean and standard deviation.

The obvious differences of FRE distributions among the axial, lateral, and elevation axes were the direct results of how the US resolution varies *axially*, *laterally*, and *elevationally* (Figure 2(c)):

a. Ultrasound images have the finest axial resolution among all axes.[11] This resolution is determined by the sound wavelength and does not change with the image depth. For our broadband linear-array transducer with a central frequency of 12.5 MHz, the axial resolution is in the order of 0.1 mm. This explains why the FRE in axial axis has a value of 0.14 ± 0.12 mm, which is in the order of the axial resolution and the smallest among all axes.

b. The lateral resolution, on the other hand, is defined by the lateral width of the sound beam, which is generally larger than the axial resolution and varies with the image depth. Fortunately, thanks to the use of multiple crystals in modern linear-array transducers, the sound beam can be electronically focused and controlled in the scan plane in either single or multiple focal zones, which results in a relatively narrow lateral beam width (reportedly 1-3 mm[11]). This translates to a larger (than axial) FREs along the lateral axis at 0.19 ± 0.22 mm.

c. The US elevation resolution was the dominant influence on calibration accuracy. Similar to the lateral direction, the elevation width (section thickness) of the US beam determined the elevation resolution. However, as we discussed earlier, because there were no multiple crystals in the out-of-plane direction to provide electronic focusing, the beam was only focused mechanically in the elevation axis (by either curving the crystal or placing an acoustic lens in front). This resulted in a much larger elevation beam width, which contributed to the largest uncertainties and errors among all axes.

In terms of performance, the calibration system was also reasonably fast. Each of the 50 trials converged in an average of 12.5 seconds, sufficiently fast for many applications in an operating room.

5.3 Quantitative US Elevation Beam Profile

Using the *Bridge* phantom, we measured the elevation beam width at 34 different axial depths for our broadband linear-array probe at a central frequency of 12.5MHz and 16MHz. Figure 8(a) shows seven samples of such measurements: the height of the echo band represents the elevation beam thickness at that depth. More measurements covering the axial depth from 4.4mm to 31mm outline the overall elevation beam profile (Figure 8(b)).
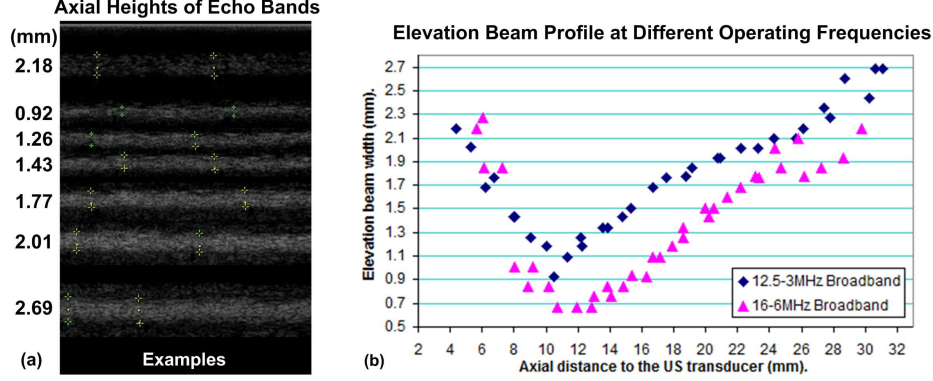

Figure 8. US elevation beam profile at different central operating frequencies.

We have made several key observations through the experiments:

1. The beam in the elevation plane focused sharply only at one fixed axial distance with a short focal length, e.g., at 12.5MHz, the beam was most narrow (0.92mm) at about 10.5mm axial distance away from the transducer. This means the elevation focal depth is at 10.5mm with a minimum section thickness of 0.92mm. On the other hand, for the same frequency, the maximum elevation beam width was 2.69mm at an axial depth of 30.6mm, which is located almost at the bottom of the effective imaging region.

2. Focusing in the scan plane (which is electronically controllable) had no influence on the beam profile in the elevation plane. At a fixed axial distance, modifying the focal length or the number of focal zones in the the scan plane only changes the clarity of the echo bands, not their heights. The elevation beam profile therefore remains unchanged. (N.B. these observations are in agreement with those of Goldstein[13] and can be explained by the fact that the focusing of US beam is performed mechanically[11]).

3. The operational frequency of the transducer also played a role in the elevation beam profile. At a higher central frequency of 16MHz, the elevation focal length was longer than at 12.5MHz (Figure 8(b)). This agreed with the basic US physics that the higher the frequency, the more the near field gets extended for a focused sound beam.[11] This observation is important because it suggests that a re-calibration may be necessary if the transducer's operational frequency is changed.

4. Finally, what Figure 8(b) essentially tell us is an error map in an US image, which outlines the distribution of errors if the US probe is to be used to locate an object in the space. The range of these localization errors are directly proportional to the elevation beam width varying with the axial depth. Knowing this error map would enable us to preventively reduce localization errors in image acquisition. For example, we may have the choice to image with regions that are closer to the elevation focal depth, which is known to introduce less errors, and to avoid regions with larger section thickness (e.g., the greater-depth region of the image in Figure 8(b)) that would most likely lead us to greater localization errors and uncertainties.

5.4 Incorporating the Elevation Beam Profile with the UKF Algorithm

Table 2 reveals our testing results of incorporating the elevation beam width into calibration using UKF. Our preliminary results showed a significant reduction in the variance (proportional to the standard deviation) by an averaged 46% in the elevation-related calibration parameters, which include the translation parameter in the elevation axis, the rotation parameter with respect to the lateral axis, and the rotation parameter with respect to the elevation axis. These results were very promising because

Table 2. Reduction in variance of parameters when incorporating the elevation beam width (EBW).

Incorporation of EBW Information	Standard Deviation in Elevation-Related Calibration Parameters		
	translation (mm) along elevation	rotation (rad) about lateral	rotation (rad) about axial
NO	0.20	0.78	0.91
YES	0.10	0.48	0.45
Reduction (%)	50%	38%	51%

- a reduction in variance for parametric estimate demonstrated an improvement on the overall calibration precision in a repeated procedure; and
- it indicated that knowing the elevation beam profile did help to reduce the uncertainties (that were originally introduced by the elevation beam profile itself) in US calibration, and allowed the UKF-based calibration algorithm to converge faster to the true calibration solution than a typical least-squares approach.

Investigations are currently underway to thoroughly validate these findings, and to explore possible means to enhance the calibration accuracy by further reducing the mean and variances of FREs.

6. DISCUSSION

The design of the *Double-N* phantom facilitated fast, accurate calibration with minimal operator interaction. The system required more images than did related systems, and the phantom design could be improved. Two notable limitations are the lack of temporal calibration and an automated procedure to measure the section thickness.

To converge, our system always required fewer than 60 images with an average of 2 data points (2 sets of N-wires) each, compared to 30 images with an average of 4 N-wire sets per image[3,5,6] or 6 images with an average of 19 N-wire sets per image.[7] In this work and related work, we observe that the key value seems to be 120 data points – this is the minimum reported number for high-accuracy calibration. Other phantom designs have provided more data per image, but at the cost of complexity and requiring human interaction in the image-segmentation stage. We have traded off the number of images needed (more) against human interaction (zero) and phantom complexity (minimum).

The *Double-N* phantom described here was a preliminary design and can be improved upon. We found the phantom to be small, limiting freehand motion. The design was optimized for our higher-frequency US probes and is not suitable for low frequencies. Also, although the phantom was designed with sterilization in mind, it was not submitted for bioengineering approval and may require modification prior to operating-room use.

A minor technical limitation of our system is that we did not perform temporal calibration., i.e., we did not tightly synchronize the tracking with the image capture.[1] This is a relatively straightforward potential improvement.

Finally, our measurement of US section thickness was manual and time-consuming, which is impractical for intraoperative use. Maintaining the 45-degree alignment between the Bridge phantom and the US scan plane also posed a challenge in the current design. We are working on a new design and an advanced software system to fully automate the procedure and to achieve more accurate and repeatable measurements.

7. CONCLUSIONS

We have presented a real-time US calibration system with automatic accuracy control and an incorporation of the US elevation beam profile, including several features crucial for intraoperative surgical use:

1. a calibration phantom that facilitates the same level of calibration accuracy as that of conventional N-wire phantoms, with a simple, low-maintenance and sterilizable design;

2. a fully automatic segmentation and calibration algorithm that requires no human interaction beyond acquiring the US images;

3. an automatic error computation and control method to ensure calibration accuracy; and

4. a quantitative evaluation and incorporation of US elevation beam profile to improve the calibration outcome.

Experiments were conducted on a data set of $10,000$ images to thoroughly investigate and validate the accuracy, robustness and performance of the calibration system. These results demonstrated that the calibration system was able to consistently and robustly achieve very high accuracy with real-time efficiency. Work is currently underway to thoroughly verify the performance of the system under actual operating-room conditions. We are also investigating more efficient means to automatically evaluate and validate the US beam profile, which may be utilized to reduce errors in not only US calibration, but also US image segmentation and registration.

ACKNOWLEDGMENTS

REFERENCES

1. L. Mercier, T. Langø, F. Lindseth, and L. D. Collins, "A review of calibration techniques for freehand 3-D ultrasound systems," *Ultrasound Med Biol* **31**(2), pp. 143–165, 2005.

2. R. W. Prager, R. N. Rohling, A. H. Gee, and L. Berman, "Rapid calibration for 3-D freehand ultrasound," *Ultrasound Med Biol* **24**, pp. 855–869, Jul 1998.

3. N. Pagoulatos, D. R. Haynor, and Y. Kim, "A fast calibration method for 3-D tracking of ultrasound images using a spatial localizer," *Ultrasound Med Biol* **27**(9), pp. 1219–1229, 2001.

4. F. Lindseth, G. A. Tangen, T. Langø, and J. Bang, "Probe calibration for freehand 3-D ultrasound," *Ultrasound Med Biol* **29**(11), pp. 1607–1623, 2003.

5. W. Y. Zhang, R. N. Rohling, and D. K. Pai, "Surface extraction with a three-dimensional freehand ultrasound system," *Ultrasound Med Biol* **30**(11), pp. 1461–1473, 2004.

6. T. K. Chen, P. Abolmaesumi, D. R. Pichora, and R. E. Ellis, "A system for ultrasound-guided computer-assisted orthopaedic surgery," *Comput Aided Surg* **10**(5), pp. 281–292, 2005.

7. P.-W. Hsu, R. W. Prager, A. H. Gee, and G. M. Treece, "Real-time freehand 3d ultrasound calibration," in *Proc. of SPIE Medical Imaging: Ultrasonic Imaging and Signal Processing*, S. Y. Emelianov and S. A. McAleavey, eds., **6513**, pp. 6513081–6513088, (San Diego, CA, USA), February 2007.

8. E. M. Boctor, I. Iordachita, G. Fichtinger, and G. D. Hager, "Real-time quality control of tracked ultrasound," *Lecture Notes in Computer Science* **3749**, pp. 621–630, 2005.

9. E. M. Boctor, I. Iordachita, G. Fichtinger, and G. D. Hager, "Ultrasound self-calibration," in *Proc. of SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Display*, K. R. Cleary, R. L. Galloway, and Jr., eds., **6141**, pp. 61412N1–61412N12, (San Diego, CA, USA), March 2006.

10. F. W. Kremkau, *Diagnostic Ultrasound: Principles and Instruments*, W.B. Saunders Company, Philadelphia, 6th ed., 2002.

11. W. R. Hedrick, D. L. Hykes, and D. E. Starchman, *Ultrasound Physics and Instrumentation*, Elsevier Mosby, Missouri, fourth ed., 2005.

12. A. Goldstein and B. L. Madrazo, "Slice-thickness artifacts in gray-scale ultrasound," *Clin Ultrasound* **9**(7), pp. 365–375, 1981.

13. A. Goldstein, "Slice thickness measurements," *J Ultrasound Med* **7**, pp. 487–498, Sep 1988.

14. M. H. Moghari, T. K. Chen, and P. Abolmaesumi, "Freehand ultrasound calibration using the unscented kalman filter," in *Proc. of SPIE Medical Imaging: Ultrasonic Imaging and Signal Processing*, S. Emelianov and W. F. Walker, eds., **6147**, p. 61470X, (San Diego, CA, USA), February 2006.

15. T. K. Chen, "A system for ultrasound-guided computer-assisted orthopaedic surgery," Master's thesis, Queen's University, Kingston, Ontario, Canada, January 2005.