AN AUTOMATED ULTRASOUND CALIBRATION FRAMEWORK INCORPORATING ELEVATION BEAMWIDTH FOR TRACKED ULTRASOUND INTERVENTIONS

by

THOMAS KUIRAN CHEN

A thesis submitted to the
School of Computing
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
October 2012

Copyright © Thomas Kuiran Chen, 2012
Abstract

Image-guided surgeries employ advanced imaging and computing technologies to assist the surgeon when direct visualization is inadequate or unavailable. As modern surgeries continue to move toward minimally invasive procedures, tracked ultrasound (US), an emerging technology that uniquely combines US imaging and position tracking, has been increasingly used for intraoperative guidance in surgical interventions.

The intrinsic accuracy of a tracked US system is primarily determined by a unique procedure called “probe calibration”, where a spatial registration between the coordinate systems of the transducer (provided by a tracking device affixed to the probe) and the US image plane must be established prior to imaging. Inaccurate system calibration causes misalignments between the US image and the surgical end-effectors, which may directly contribute to treatment failure. The probe calibration quality is further reduced by the “elevation beamwidth” or “slice thickness”, a unique feature of the ultrasound beam pattern that gives rise to localization errors and imaging uncertainties.

In this thesis, we aim to provide an automated, pure-computation-based, intraoperative calibration solution that also incorporates the slice thickness to improve the calibration accuracy, precision and reliability. The following contributions have been made during the course of this research. First, we have designed and developed an
automated, freehand US calibration system with instant feedback on its calibration accuracy. The system was able to consistently achieve submillimeter accuracy with real-time performance.

Furthermore, we have developed a novel beamwidth-weighted calibration framework (USB-FW) that incorporates US slice thickness to improve the estimation of calibration parameters. The new framework provides an effective means of quality control for calibration results. Extensive phantom validation demonstrated that USB-FW introduces statistically significant reduction ($p = 0.001$) in the calibration errors and produces calibration outcomes that are less variable than a conventional, non-beamwidth-weighted calibration.

Finally, we were the first to introduce an automated, intraoperative Transrectal Ultrasound (TRUS) calibration technology for needle guidance in prostate brachytherapy. Our tests with multiple commercial TRUS scanners and brachytherapy stepper systems demonstrated that the proposed method is practical in use and can achieve high calibration accuracy, precision and robustness.
Acknowledgments

This dissertation is dedicated in loving memory of my mother, Youlan, who lost her fight to lung cancer in the midst of my journey toward this degree but had never lost her hope and faith in her son to pursue his dreams.

I owe my deepest gratitude to my mentors, Drs. Purang Abolmaesumi, Randy Ellis and Gabor Fichtinger, for their wisdom, knowledge, encouragement, friendship, and perhaps most profoundly, their great passion that has been inspiring me to carry on, especially during those most difficult moments of my life. And to my Ph.D. committee, Drs. Parvin Mousavi and Selim Akl, for their invaluable insights, advice and support throughout the course of this research.

I wish to sincerely thank the Natural Sciences and Engineering Research Council (NSERC) and Mathematics of Information Technology and Complex Systems (MITACS) for awarding me the great honors of Canada Graduate Scholarship (CGS 2006-2009) and MITACS Ph.D. Accelerate Fellowship (2010-2011), respectively. These scholarships have provided me and my family with years of essential financial support so that I can freely pursue my research adventures.

My gratitude also extends to all my friends and colleagues at Medical Image Analysis (MedIA) Laboratory, Laboratory for Percutaneous Surgery (PerkLab) and Medical Informatics (Med-i) Laboratory at School of Computing, Queen’s University,
for keeping me a cheerful and memorable company; and in particular, to Amir Tah-
masebi, Yenhui Gong, Mehdi Moradi, Andrew Lang, Mehdi H. Moghari, Lili Wang,  
Farhad Imani, Tamas Heffter, Andras Lasso, Tamas Ungi, and Csaba Pinter, without  
whom the days of research would have been a little too lonely!

A heartfelt thanks to Debby Robertson, our beloved Graduate Assistant at School  
of Computing, for her tireless support during my years of studies at the School, and  
her timely help to put every single piece of required documents together for the final  
submission of this thesis while I was helplessly away in British Columbia.

Nothing could express my gratefulness to my father, Bairong, my sister, Mingdi,  
and my uncle, Xingfeng, whose unwavering love and support are always there for me.

And personally, I would offer this thesis to my lovely wife, Hong, for every moment  
of joys and tears that we have shared together. None of this would have become even  
close to possible for me without her unyielding love, comfort and self-sacrifice.

Finally, this is for Eddie and Evelyn, my little angels, who give daddy the endless  
joy, the greatest laugh, and of course, those countless, sleepless nights!

Thomas Kuiran Chen

British Columbia, Canada

August 29, 2012.

"Enthusiasm is contagious. Be a carrier. - Susan Rabin".
Statement of Originality

This dissertation is the result and contribution of my own original work and, to the best of my knowledge and belief, contains no material previously published or authored by another person except where due references/acknowledgements are made. It has not been submitted in whole or in part for a degree at any other university. The following contributions were derived from this work.

Peer-reviewed Journal Publications:


Patent applications:


## Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>Two Dimensional Page 26.</td>
</tr>
<tr>
<td>3D</td>
<td>Three Dimensional.</td>
</tr>
<tr>
<td>AIUM</td>
<td>American Institute of Ultrasound in Medicine.</td>
</tr>
<tr>
<td>BMUS</td>
<td>British Medical Ultrasound Society.</td>
</tr>
<tr>
<td>CAS</td>
<td>Computer-assisted Surgery Page 1.</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography Page 1.</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration (an United States Federal Agency in charge of public health safety.</td>
</tr>
<tr>
<td>freehand</td>
<td>The ultrasound images are captured freehand, meaning as long as the US probe is properly tracked, the probe could move freely on top of the patient’s anatomy so that the images would cover the desired region of interest. Page 7.</td>
</tr>
<tr>
<td>iCAL</td>
<td>Intraoperative Calibration for Brachytherapy Page 102.</td>
</tr>
<tr>
<td>IGS</td>
<td>Image-Guided Surgery Page 1.</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>LRE</td>
<td>Line Reconstruction Error</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic Resonance</td>
</tr>
<tr>
<td>OR</td>
<td>Operating Room</td>
</tr>
<tr>
<td>PRE</td>
<td>Point Reconstruction Error</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>TRE</td>
<td>Target Registration Error</td>
</tr>
<tr>
<td>TRUS</td>
<td>Transrectal Ultrasound</td>
</tr>
<tr>
<td>US</td>
<td>Ultrasound</td>
</tr>
</tbody>
</table>
Table of Contents

Abstract ............................................................... i
Acknowledgements ..................................................... iii
Statement of Originality ............................................... v
Glossary ................................................................. vii
Table of Contents ....................................................... ix
List of Tables .......................................................... xi
List of Figures .......................................................... xii

Chapter 1:
   Introduction ....................................................... 1
   1.1 Clinical Background and Significance ........................ 1
   1.2 Current Challenges and Motivations ........................... 3
   1.3 Research Objectives ........................................... 7
   1.4 Contributions .................................................. 9
   1.5 Thesis Organization ............................................ 11

Chapter 2:
   Background and Literature Overview ............................ 14
   2.1 Tracked Ultrasound Imaging .................................. 14
   2.2 Ultrasound Probe Calibration .................................. 26
   2.3 Characterization of Ultrasound Beam Pattern ............... 38
   2.4 Use of Slice Thickness in Probe Calibration ............... 41

Chapter 3:
   Automated, Freehand Calibration With Accuracy Feed-
   back ................................................................. 42
3.1 Introduction ........................................ 43
3.2 Methods and Materials .............................. 43
3.3 Results ........................................... 72
3.4 Conclusion ........................................ 77

Chapter 4:
USB-FW: A Beamwidth-Weighted Calibration Framework 79
4.1 Introduction ........................................ 79
4.2 Methods and Materials .............................. 80
4.3 Results ........................................... 94
4.4 Conclusion ........................................ 99

Chapter 5:
Intraoperative Calibration for Prostate Brachytherapy 101
5.1 Introduction ........................................ 102
5.2 Materials and Methods .............................. 107
5.3 Results ........................................... 126
5.4 Conclusion ........................................ 135

Chapter 6:
Conclusions and Future Work ............................. 136
6.1 Summary of Novelty and Contributions .......... 136
6.2 Discussion and Future Work ........................ 139
6.3 Potential Use of Beamwidth in Other US-Guided Interventions 145

Bibliography ............................................. 147
List of Tables

1.1 Advantages of US compared to CT and MR. .......................... 2
3.1 Standard deviation (in pixels) in the segmentation outcome: automatic system vs. human operators. ...................................... 74
3.2 PRE for 50 independent calibration trials. ............................. 76
4.1 Reduction of 3D Point Reconstruction Errors (PRE) by USB-FW. .. 96
5.1 Experimental setup with multiple TRUS scanners. .................... 124
5.2 Experimental setup with multiple brachytherapy stepper systems. .. 125
5.3 Validation results of iCAL for 50 independent, real-time calibration trials. ................................................................. 127
5.4 Validation results of iCAL tested with different TRUS scanners. .... 129
5.5 Validation results of iCAL tested with different brachytherapy steppers. 131
5.6 TRE results of needle insertion to validate template-TRUS calibration accuracy. ............................................................... 132
List of Figures

1.1 Ultrasound beam profile of a linear-array transducer. ................. 12
1.2 Localization errors and artifacts caused by ultrasound slice thickness. 13

3.1 The design of the real-time automatic ultrasound calibration system. 44
3.2 Hardware configuration for image acquisition and tracking. .......... 46
3.3 A High-Level Object-Oriented Design Graph of the Calibration System in industry-standard Unified Modeling Language (UML). ............. 49
3.4 Front and side views of the Double-N phantom. ....................... 50
3.5 CAD drawing of the Double-N phantom. ................................ 52
3.6 Design specifications of the Double-N phantom. ....................... 53
3.7 Wire appearance determined by the ultrasound axial and lateral resolutions. ...................................................... 54
3.8 Appearance of nylon wires of different diameters at the ultrasound scan-plane focal depth. ........................................ 56
3.9 The multiple-step fully automatic segmentation algorithm. .......... 59
3.10 Frame transformations to solve for calibration parameters. .......... 63
3.11 Calculation of α-ratio based on the N-wire geometry. ............... 65
3.12 Real-time calibration accuracy feedback and control. ............... 70
3.13 Standard deviations in manual segmentation for 7 human subjects. . 74

4.1 Principles of slice-thickness measurement using an inclined plane. ... 81
4.2 Design of the real-time, freehand, N-wire-based US calibration system with USB-FW. .................................................. 85
4.3 Extracted US elevation beamwidth profiles. ............................. 95

5.1 TRUS-guided prostate-cancer brachytherapy: needles are inserted into the prostate through a template into a patient in the lithotomy position. 103
5.2 Design of the iCAL system: calibration phantom mounted on a standard brachytherapy stepper. ................................. 108
5.3 Workflow of the iCAL brachytherapy calibration procedure. .......... 109
5.4 iCAL phantom: CAD design, inner N-wires, complete assembly, and rubber-membrane seal. ........................................ 111
5.5 An example of the automated temporal calibration in iCAL. .......... 116
5.6 The unique N-wire geometry of the iCAL calibration phantom. . . . . 117  
5.7 Coordinate transformations from the template to the TRUS image. . 118  
5.8 The interactive graphics interface of iCAL showing the 3D overlay of  
the template grid to the TRUS image in real time. . . . . . . . . . . . . 121  
5.9 Automatically segmented N-Wires from the TRUS image of the iCAL  
phantom. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 126  
5.10 Needle insertion to validate template-TRUS calibration accuracy (TRE).133  
5.11 Results of the final, real-time 3D overlay of the template to TRUS. . 134
Chapter 1

Introduction

1.1 Clinical Background and Significance

Image-guided surgery (IGS), also referred to as computer-assisted surgery (CAS), employs advanced computing and imaging technologies to guide and assist the surgeon when his/her direct vision is inadequate or unavailable [63]. As modern surgeries continue to move toward minimally invasive or non-invasive procedures, tracked ultrasound (US), an emerging technology that uniquely combines both US imaging and position tracking, has been increasingly used for intraoperative guidance in surgical interventions [18].

Medical US has significant advantages over other imaging modalities, such as Computed Tomography (CT) or Magnetic Resonance (MR) (Table 1.1).

First, medical US does not expose the surgical team or the patient to radiation. It is also safe to operate an US in the presence of metals, which are the primary materials used in conventional surgical equipment. In comparison, CT can provide high-quality visualization, especially in the examination of hard tissues. However, it
CHAPTER 1. INTRODUCTION

Table 1.1: Advantages of US compared to CT and MR.

<table>
<thead>
<tr>
<th></th>
<th>CT</th>
<th>MR</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ionizing Radiation</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ferromagnetic Risk</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Size</td>
<td>Takes up a room</td>
<td>~2 m³</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>1-2 ton</td>
<td>~200 kg</td>
<td></td>
</tr>
<tr>
<td>Portability/Mobility</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Real-time Imaging</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cost (USD)</td>
<td>1-3 millions</td>
<td>50-80k</td>
<td></td>
</tr>
</tbody>
</table>

emits ionizing radiation that may cause long-term health damage to the patient and the surgical team. MR does not emit ionizing radiation, but its powerful magnetic field attracts ferromagnetic (iron-containing) objects, causing them to move suddenly and forcefully. This magnetic field can pose a safety risk for the patient and limits the use of MR (e.g., it cannot be used with conventional surgical tools or if the patient has metallic implants within his body).

Second, US scanners are lightweight, compact and mobile. A typical medical US machine is less than 2 cubic meters in volume and weighs approximately 200 kilograms. In comparison, CT and MR scanners weigh more than a ton on average and occupy a large room (commonly located in a dedicated building) for operation. In addition, US machines are normally equipped with wheels to allow easy mobilization from one location to another. The smallest commercially available US machine resembles a cell phone (e.g., General Electric Vscan). Additionally, there are USB ultrasound machines that attach a US transducer to a PC or laptop via a USB port.

Furthermore, US technology is relatively inexpensive. CT and MR scanners cost millions of dollars per unit. In addition, the construction of CT and MR suites (rooms
or buildings to safely house and operate the machine) also adds significant cost to their operation and maintenance. These expensive technologies are unlikely to be universally adopted, thus limiting their ability to benefit most patients [18]. US is considered a potentially cost-effective alternative to CT and MR scanners. Even the most advanced US machines are currently commercially available for a fraction of the cost of a CT or MR machine.

Finally, US imaging is captured in real time. Modern US instrumentation instantly displays anatomic structures while the patient is being scanned.

All of these characteristics, when combined with precise position tracking, advanced image processing, and automated computer algorithms, make US an ideal tool for providing safe intraoperative guidance in the operating room (OR).

1.2 Current Challenges and Motivations

The idea of using diagnostic US for intraoperative guidance is revolutionary yet surprisingly simple; it involves a technology called "tracked US imaging" that embeds real-time position data into the US-image acquisition process. Tracking is typically achieved by rigidly affixing the probe to a localizer that is tracked in real time via a position sensing system [50]. This localization strategy avoids invasively mounting the tracking device directly on patient (as is typically performed in CT or MR interventions), reducing any additional trauma or pain to the patient that can prolong his recovery and hospitalization time.
1.2.1 Tracked Ultrasound Calibration

The patient’s position cannot be adequately determined from the localization of the US probe. The intrinsic accuracy of a tracked US system is solely determined by a unique procedure called “probe calibration”, for which a spatial registration between the coordinate systems of the transducer and the US image plane must be established prior to imaging. Inaccurate system calibration causes misalignments between the US image and the surgical end-effectors (such as a surgical drill in an orthopedic procedure or a template needle guide in prostate brachytherapy) and may directly contribute to treatment failure or patient morbidity.

Unfortunately, tracked US calibration is currently a laborious, manually intensive procedure. The calibration is also more qualitative than quantitative and requires a great deal of eyeballing and subjective judgments by the operator [50].

Calibration is only performed periodically (primarily due to the procedure’s inefficiency) and is mostly conducted outside of the operating room. This practice is based on the assumption that calibration parameters remain valid over time. However, in reality, calibration parameters can change during the equipment’s storage, transportation and setup.

Perhaps most critically, system calibration errors are difficult to detect. As a result, the surgeon has no assurance whether the system is functioning correctly at the start of the procedure. Most current calibration procedures do not contain a validation mechanism to verifies the equipment’s calibration accuracy in the operating room.

Finally, the current calibration procedure is a major recurring cost for care facilities, consuming manpower, time and money. For example, in prostate cancer
brachytherapy, one must first book a calibration room, de-commit a US scanner from clinical use, transport the equipment, prepare the supplies (needles, water tank, etc.), set up the system, collect and process data, log, analyze and document the results, dispose of all used supplies, pack away the brachytherapy system, and return the US scanner to the clinic. This workflow needs to be repeated from time to time.

1.2.2 Ultrasound Slice Thickness and Localization Errors

Probe calibration quality is also dependent on US slice thickness, commonly referred to as the elevation beamwidth or section thickness [33, 42]. This unique, inherent characteristics of an ultrasound beam pattern gives rise to localization errors and uncertainties in US imaging.

To understand the problem, one must first understand how an ultrasound resolution is defined. US imaging quality is commonly evaluated [33, 42] along the axial (direction of sound travel), lateral (in plane and perpendicular to the beam axis), and elevation (out of plane) axes, all of which are taken with respect to the transducer’s crystal array. The axial and lateral axes form the scan plane, whereas the lateral and elevation axes define the elevation plane. An US beam has a finite width in the lateral and elevation directions. These widths define the imaging resolution (Figure 1.1).

The lateral beamwidth is determined by focusing the beam in the scan plane. Such beam focus can be controlled electronically and is typically achieved by offsetting the firing of multiple crystals in a group by a very small time delay [33]. To control the lateral resolution, the number of electronic focal zones and their positions can be adjusted across the imaging depth of the scan plane, resulting in a small and relatively uniform lateral beamwidth. However, electronic focusing cannot be used in
the elevation direction because the elevation plane of a 1-D linear array or a curvilinear transducer does not contain multiple crystals. Consequently, the US beam is only focused mechanically in the elevation direction, either by curving the crystal or placing an acoustic lens in front of it. As a result, the elevation beamwidth is typically much larger than the lateral beam width and can only achieve a sharp focus when it is at a fixed axial distance from the transducer. More advanced 1.5D transducer arrays have been recently introduced to address this limitation in the elevation resolution. These advanced transducers have an increased number of crystal elements and can be electrically focused along the elevation axis [40]. However, the technologies are still in development and are not used in conjunction with the majority of clinically available diagnostic ultrasound scanners.

To generate an US image, three fundamental assumptions are made in the data-handling protocols [29, 33, 42]. First, the sound signal is assumed to propagate at a speed of 1540 meters per second in all tissues. Second, the sound signal is assumed to travel along a straight path in a tissue (this assumes no bending of the beam pattern). Finally, all of the reflected echoes are assumed to be from the central axis of the sound beam. This last assumption essentially implies that an US image is a mathematical plane with zero thickness, leading to artifacts and localization errors.

This concept is demonstrated in Figure 1.2. Three sound reflectors, $A$, $B$ and $C$, are located at the same axial distance from the transducer but have different elevational offsets. When the sound reflectors are interrogated by a sound beam, each reflector produces an echo that is detected by the transducer. However, due to the assumption that all echoes are received from the beam’s central axis, these reflections would be summed and interpreted as if they were from a single reflector on the central
axis. Hence, the US machine would not be able to resolve the three reflectors based on their elevational differences; instead, the reflectors would be displayed as a single point. This result leads to an imaging error. For example, if an object is physically located at position \( D \), the US machine would assume that it is on the beam’s central axis and would display a false contour (artifact \( E \)) in the image.

All of these problems lead to localization errors because an US image is essentially a projection from the object’s true spatial location to its assumed position along the beam’s central axis. The only assurance in US imaging is that the position error due to its slice thickness cannot be greater than the elevation beamwidth at any given axial depth.

### 1.3 Research Objectives

In this thesis, we aim to overcome the aforementioned obstacles in tracked US calibration by developing a fully automated, computationally based, intraoperative calibration technology that is intended for use in the operating room while a patient is prepared for surgery. Thus, the following research objectives were established.

Our first objective is to design and develop an automated, freehand US calibration solution with accuracy feedback and a control mechanism that requires minimal to no human interaction. The software system must be accurate, reliable, and sufficiently fast for intraoperative applications. As part of the calibration system, we will also design and prototype a calibration phantom that can readily facilitate an automated procedure. Equally important, the device should be easy to use, sterilizable, and cost-effective to manufacture and maintain. Lastly, to determine the validity of the calibration, an automated accuracy report needs to be in place to provide instant
feedback to the surgical team.

Second, we plan to quantitatively characterize US beam thickness and incorporate this information into the US calibration framework. Initially, we aim to develop a US beamwidth characterization system that allows quantitative measurement and analysis of the US slice thickness with a linear array transducer. We intend for this beamwidth profiling process to be run preoperatively. The process will characterize the designated ultrasound beam pattern. This information may offer insights into the localization errors that are introduced by US slice thickness. Next, we will investigate how US slice thickness can be effectively used to improve the calibration results of our US calibration process. Finally, we will perform phantom validation experiments to examine the impact of incorporating the US slice thickness into our calibration process. Our hypothesis is that incorporating slice thickness information, which is directly associated with the localization errors and uncertainties in tracked US imaging, will improve the accuracy, precision and reliability of our calibration results.

Our final objective is to develop a fully automated US calibration solution in a clinical application, specifically prostate brachytherapy. We will design and develop an intraoperative transrectal ultrasound (TRUS) calibration system for brachytherapy needle insertion, a definitive treatment for early-stage prostate cancer. The needle is used to implant small radioactive isotope capsules (seeds) into the prostate gland to kill the cancerous cells with radiation [81]. We will thoroughly investigate the accuracy, precision, robustness and speed/performance of the TRUS calibration technology using multiple commercially available TRUS scanners and brachytherapy systems.
1.4 Contributions

1.4.1 Automated, Freehand Calibration with Accuracy Feedback

_N.B.: This work was published in the Journal of Ultrasound in Medicine and Biology 2009 [16]_.

We have designed and developed a fully automated, freehand US calibration system with real-time feedback and calibration quality control. An important element of this system is its Double-N calibration phantom, featuring a simple, sterilizable design that is intended for operating room usage. The calibration system uses an automatic error retrieval and accuracy control mechanism based on a set of ground-truth data.

Extensive validations were conducted on a set of 10,000 images in 50 independent calibration trials to thoroughly investigate the accuracy, robustness, and performance of the calibration system. The results have shown that the calibration system was able to consistently, efficiently and robustly achieve high calibration accuracy in real-time performance.

1.4.2 USB-FW: A Beamwidth-Weighted Calibration Framework

_N.B.: This work was published in the Journal of Ultrasound in Medicine and Biology 2011 [14]_.

We have developed a novel real-time, freehand calibration system that systematically incorporates US slice-thickness profiles into a filtered, weighted-least-square
framework (USB-FW) to improve the tracked US’s reconstruction accuracy. An important component of this system is its slice-thickness measurement device, which aids in the extraction and quantitative characterization of the slice-thickness profile associated with the US imaging system. This device can be used to characterize beamwidth profile across a wide range of imaging depths.

Extensive experiments were conducted on a 10,000-image dataset to evaluate how our framework affects calibration accuracy. The results showed that 3D reconstruction errors were significantly reduced in all of the experiments. Furthermore, real-time testing demonstrated that the proposed method can perform effectively with a small number of input images. This result suggests great potential for using the device intraoperatively, when only a limited amount of data may be available. This new framework can thus enable efficient quality control of calibration accuracy during real-time operating room use.

1.4.3 Intraoperative Calibration for Prostate Brachytherapy

N.B.: This work was published in the Journal of Medical Physics 2011 [15].

We have designed and developed a fast, automated, pure-computation based, intraoperative TRUS calibration technology for brachytherapy systems. This technology is intended for use in the operating room while the patient is being prepared for surgery. The core of this invention is the mechanical coupling of a precision-made calibration phantom and a geometric replica of a standard brachytherapy template. The coupling effectively combines the phantom with the brachytherapy stepper system in one unit. Elimination of conventional calibration systems, which are typically lengthy and laborious, will make this brachytherapy system more accurate and consistent, as
well as much less expensive. Thus, two main goals can be achieved simultaneously with this technology.

Four types of independent tests with multiple commercial TRUS scanners and brachytherapy stepper systems were performed, and the results demonstrated that the proposed method has high calibration accuracy, precision and robustness, and can be practically useful for a current brachytherapy procedure.

1.5 Thesis Organization

This thesis is organized in manuscript format and contains six chapters:


Chapter 3 - Chapter 5 These four chapters correspond to the main contributions of this thesis:

- Automated, Freehand Calibration with Accuracy Feedback (Chapter 3),
- USB-FW: A Beamwidth-Weighted Calibration Framework (Chapter 4),
- Intraoperative Calibration for Prostate Brachytherapy (Chapter 5),

Chapter 6 Conclusions and Future Work This chapter concludes the thesis, summarizes current advancements, and describes future work associated with each contribution.
Figure 1.1: Ultrasound beam profile of a linear-array transducer in *axial*, *lateral* and *elevation* axes. The lateral beamwidth 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. The beam is only focused mechanically in the elevation direction, by either curving the crystal or placing an acoustic lens in front, with sharp focus only possible at a narrow axial distance from the transducer. As a result, the beam is much wider elevationally than laterally, forming a curvilinear volume that defines the US resolution along these directions.
Figure 1.2: Localization errors and artifacts caused by ultrasound slice thickness. US instrumentation assumes zero thickness of the image plane and therefore cannot differentiate sound reflectors (A, B and C) located at the same axial depth but with different elevation offsets. As a result, an object (D) located off the central beam axis elevationally will be falsely displayed.
Chapter 2

Background and Literature

Overview

2.1 Tracked Ultrasound Imaging

Tracked ultrasound imaging embeds real-time position information into the ultrasound-image acquisition process. As previously discussed, an image-guided surgical system needs to accurately localize the patient in the operating room, in order to register a preoperative plan to the patient for accurate surgical navigation. In tracked ultrasound imaging, the position of the targeted anatomy is provided by the intraoperative ultrasound images by mounting the tracking device directly onto the ultrasound transducer (instead of the patient) and tracked by a camera system in real time. In this section, we will review the standard components for tracked ultrasound imaging, including the state-of-the-art technologies for position tracking, image acquisition, and perhaps most crucially, synchronization between the two processes.
2.1.1 Real-time Position Tracking

In tracked ultrasound, positioning is typically achieved by affix trackers (or localizers) onto the ultrasound transducer and traced in real-time by the tracking system. The advancement in tracking technologies has greatly facilitated the accurate and fast surgical guidance in image-guided surgeries. There are three types of most commonly used tracking devices [63]: mechanical, optical and electromagnetic tracking systems.

Mechanical Tracking System

Mechanical trackers were the earliest tracking tools available for image-guided interventions. These devices are articulated mechanical arms with multiple, precision-engineered joints to calculate the tip position in the 3D space. Originally intended for neurosurgeries, mechanical trackers have very high accuracy (typically in the order of 0.1 mm [48]). However, their main disadvantages are: (a) they are unable to track multiple targets; (b) the equipment is typically cumbersome and can easily interfere with surgical field. Today the most frequent use of mechanical trackers are in the professional, large-scale manufacturing (e.g., car industry). One of the biggest, commercial providers for high-accuracy, mechanical tracking arms is FARO Technologies Inc. (Lake Mary, FL).

Optical Tracking System

Optical trackers are perhaps the most widely used tracking devices for image-guided surgeries (especially in orthopaedics) due to high accuracy and relatively large working volume [25]. The basic idea is to optically trace (with light) the pre-determined geometry of rigid markers by a camera system. Mathematically, a minimum of three
markers are required to compute the physical position and orientation of a rigid body in a 3D space, though more are desired to improve the tracking visibility and the measurement accuracy. Generally, there are three types of optical tracking systems, depending on different light sources they are using:

1. **Infrared tracking.** These are the Charge-Coupled Device (CCD) camera systems that detect the infrared signals. The markers in use here either emit infrared signals by embedded Light Emitting Diodes (LED) (active marker) or reflect the infrared light (using retro-reflective spheres) transmitted by the camera (passive marker). Widely used tools in this category are Polaris and Optotrak (Certus) Optical Tracking Systems, both manufactured by NDI (Northern Digital Inc., Waterloo, ON, Canada). Polaris (P4 model) is notably the most-used optical tracking system in clinical applications and research [24], which is equipped with two infrared sensors with a reported accuracy of 0.38 mm 3D root mean square (RMS) error at 60 Hz. The Optotrak Certus employs three sensors and has a reported accuracy of 0.1 mm root mean square (RMS) error in X and Y axes, and 0.15 mm in Z-axis, all measured at a stand-off distance of 2.25 m [74].

2. **Video or frame (visible light) tracking.** Unlike infrared-based tracking, this type of optical tracking devices are fully passive, using available visible light to detect and track objects of interest from a real-time video frame sequence. Typically two calibrated cameras are required to accurately localize a 3D object. The objects are marked using small checkered (checkbox-like) patterns. The leading, commercially available product based on this technology is MicronTracker2 series (Claron Technology Inc., Toronto, ON, Canada), which has a claimed
accuracy of 0.2-0.35 mm RMS at a distance of 40-100 cm. Bootsma et al. has found that their target registration accuracy using a MiconTracker system was slightly better than using a Polaris tracking system [10]. Finally, one attractive point of a video-based tracking system is its low price, e.g., at a similar accuracy, the price of MiconTracker is only a fraction of that an infrared-based system. This is largely contributed by the fact that the video tracking system is based on fairly mature techniques in the computer vision field and does not require highly sophisticated camera hardware, which subsequently reduces the manufacturing costs.

3. Laser tracking. This tracking mechanism is similar to that of infrared-based trackers except that the cameras and targets work with the laser spectrum (high frequency, focused light). The only commercially available product in the current market is laserBird2 (Ascension Technology Corp., Burlington, VT, USA), which has a claimed accuracy of 0.7 mm RMS and can track the target at a very high frequency (240 Hz).

Optical tracking systems are known for their high localization accuracy and the capability to fast track multiple targets. However, their major drawback is the line-of-sight issue: because the cameras have to actually “see” the target in order to track, an unblocked visibility between the camera and targets must be maintained at all time during tracking, which poses space constraints on the motion of the surgical tools. And as a result, optical tracker cannot be used inside human body for tracking. This challenge has led to the advancement of electromagnetic trackers, which have alleviated the need of line-of-sight and are able to track surgical tools like catheters and needle tips inside the body.
Electromagnetic Tracking System

The fundamental idea behind electromagnetic tracking is based on the Faraday's Law of Induction Theory, i.e., to generate voltage (and current) in a conductive material moving through a magnetic field. In an electromagnetic tracking system, a transmitter creates a pulsed, magnetic field of known geometry that is detected by individual electromagnetic field sensors (made of solenoids or coils). By measuring the magnitude of the induced electrical currents as each sensor moves within the magnetic field, the relative position of the sensor to the transmitter can be determined. The magnetic field may be generated by either an alternating current (AC) or direct current (DC) transmitter.

Compared to optical tracking systems, the primary benefits of using magnetic trackers are:

- No requirement of line of sight between the transmitter and the target;
- The sensor can be made very small (e.g., only 8 mm long and 0.5 mm in diameter [63]) so that it is feasible to embed the sensors in surgical tools and track them inside the human body.

However, their main disadvantages are:

(a) Both the AC and DC electromagnetic trackers are sensitive to metallic objects placed near to the transmitter or sensors, as well as to other magnetic noise generated by power sources and electrical devices, which would distort the magnetic field to certain extent and reduce the tracking accuracy. This limits the use of electromagnetic tracking inside the operating room because most of the surgical tools are made of metals and the environment also contains many metallic
devices. Nevertheless, the recently developed devices are far less affected by metal distortion than previous generations [63].

(b) As studied by Glossop [25] and Yaniv et al. [80], the localization accuracy (in the range of 1-3 mm RMS) of the electromagnetic tracking systems still falls behind that of optical trackers (typically sub-millimeter).

Commercial electromagnetic tracking systems include the Aurora system (Northern Digital Inc, Waterloo, ON, Canada) with a reported accuracy in the range of 0.9-1.4 mm RMS, and the most recent 3D Guidance Tracker family of driveBAY, trakSTAR, and medSAFE (Ascension Technology Corp., Burlington, VT, USA) with a reported accuracy of 1.4 mm RMS.

Recently, Yaniv et al. reported the results of a comprehensive study on the performance of electromagnetic tracking systems with multiple factors to be considered in a clinical setup [80].

2.1.2 Image Acquisition

For tracked US to integrate position information into the US images, the image data needs to be transferred to an external computer for postprocessing. There are three common ways to acquire (transfer) the ultrasound images:

1. Analog video output to a frame-grabber installed on the computer;

2. Direct export of raw RF data (before image formation) to a storage media;

3. Digital output and transfer in DICOM format.
Analog Acquisition

Analog data output provides the most common interface that would work for any ultrasound hardware and a standard external PC. Modern diagnostic ultrasound equipment offers an analog video stream output in the format of composite, S-video, or a component-video (which has the highest video quality). Via proper cabling, the analog video can be connected to a standard video-capturing card (frame-grabbers) installed on the host computer to be digitally sampled and saved for further processing. The main benefits of analog data acquisition are:

(a) Standard setup that can work on virtually all modern ultrasound scanners and PCs; no special hardware requirement on either the ultrasound scanner or the PC.

(b) Best software portability (compatibility) of image-guided applications, i.e., the same application/software package can work with different ultrasound scanners in the market;

(c) Real-time image acquisition at 30 frames per second (fps) typically.

However, there are two notable disadvantages of the analog output however:

- The image quality is lower compared to a digital format [33]. As we have explained in the previous section, the ultrasound machine stores the scan-converted image data digitally in its internal memory. When outputting to the analog video port, the ultrasound machine needs to perform a digital-to-analog (D/A) conversion of the signal, which is then converted back to digital by the video-capturing card on the host computer. This double conversion results in degradation in image quality of the original, digital image.
• There is a frame-rate mismatch between the ultrasound machine and the video output [50]. The ultrasound scanner operates at a certain frame rate depending on the imaging settings (typically in the range of 10 to 100 fps). The analog video on the other hand has a standard output of 25 fps (PAL) or 30 fps (NTSC). So in practice, frames are either dropped if the ultrasound scans at higher-than-analog frame rate, or duplicated if the ultrasound frame rate is lower. The former diminishes the ultrasound scan’s temporal resolution, i.e., a loss of the original, more densely sampled data scans.

RF Data Acquisition

As discussed in the previous section, the real-time processing units of an ultrasound scanner go through multiple stages of signal processing on the reflected or scattered echoes, including single and multiple line RF processing, filtering, envelop detection, compression, preprocessing, scan conversion, etc. before they are stored in the image buffer for final display on the monitor. These processes aim to make the data representation compact which is essential for real-time ultrasonic imaging, however, at an expense of loss of information in the original RF signals. Certain applications like tissue characterization make use of information embedded in the original RF data, in which case direct access to the raw RF data is preeminent [54].

Primarily for research purposes, some new diagnostic ultrasound scanners in the market, such as SonixTOUCH Research (built on the company’s OPENSonix platform) (Ultrasonix Medical Corp., Burnaby, Canada), give users access to the raw, real-time data before the image formation, including pre-beamformed digital RF data.
from individual channels, beamformed RF data, envelope detected data and interpolated image data [21, 70].

**DICOM Acquisition**

DICOM stands for “Digital Imaging and Communications in Medicine”, an international standard established in 1993 by the American College of Radiology (ACR) and National Electrical Manufacturers Association (NEMA) to standardize the handling, storing, printing, and transferring information in medical imaging [59]. DICOM provides a universal file format to store medical image data including patient information, and a TCP/IP-based network protocol to communicate between different systems. The idea is for all manufacturers that conform to DICOM standard to produce medical hardware that would easily and efficiently share and exchange data in the same format (called “a picture archiving and communication system (PACS)”), regardless of what imaging device is used to generate the data (e.g., CT, MRI or US).

DICOM has been widely supported by all major medical imaging device manufacturers and quickly adopted by hospitals and medical imaging facilities worldwide. Most of the current diagnostic ultrasound scanners in the market support DICOM and allow the users to transfer the acquired image data from the local ultrasound machine to a remote computer (called DICOM server) via a standard TCP/IP network.

**2.1.3 Synchronization between Imaging and Tracking**

The localization accuracy of the tracked ultrasound imaging relies on accurate and reliable determination of the position of the ultrasound images acquired from the patient. However, since the images and the tracked probe positions are generated by
two separate hardware equipment (the ultrasound scanner and the tracking system),
a proper synchronization between the two must be established to correctly associate
each acquired ultrasound image with its corresponding positional data. This process
is also commonly referred to as a “temporal calibration”.

When an ultrasound image is acquired and its corresponding probe position recorded,
both data can be time-stamped. However, this time-stamping process itself introduces
certain delay, because of the difference in processing speed between the ultrasound
machine and the tracking system, and the necessary time required for data transfer
(from ultrasound machine and the tracking device to computer). Finding this delay
or lateness (called “latency”) is the primary goal for the synchronization.

In general, there are have been two types of technologies in the state of the art to
find the temporal latency: an active synchronization or a passive one.

**Active Synchronization**

In active synchronization [1, 2, 4], a number of distributed, real-time, multi-threaded
software processes are involved to establish an initial synchronization state between
the tracking system and the image acquisition component, and then a continuous
image acquisition process can carry on. Afterwards, the synchronization program
may work as a background process (a software daemon) to perform automatical re-
sync in a predefined interval or upon user’s request.

Barry et al. invented a hardware device to send syncing request to the tracking
system every time that an ultrasound image was acquired [4].

Barratt et al. developed a software program to trigger the synchronization of
image capturing and tracking at a predetermined interval based on the ECG wave in
a cardiac cycle [1, 2].

The major benefits of active synchronization are its high accuracy, real-time performance, and no assumption of a constant latency. Further, there is no image-based processing or computation required which is an added advantage considering image processing on ultrasound is a challenging task and may introduce uncertainties and errors into results. Active synchronization is therefore the common approach adopted in industry where efficiency and accuracy are preeminent. However, the drawbacks are the added system complexity and the requirement for advanced software tools and maintenance.

**Passive Synchronization**

Unlike the active approach, a *passive* synchronization is typically performed “offline” and involves a great deal of postprocessing on the image data acquired from a specifically designed ultrasound phantom. The general idea behind it is to introduce some form of abrupt change in the motion of the ultrasound transducer that would also result in a traceable difference in the ultrasound image. By identifying and matching this difference in the positional data and in the ultrasound image, a latency between the tracking and image acquisition could be determined.

One of the simplest techniques [49, 66] was to first hold the ultrasound probe static (while imaging the skin) for a moment and then remove it out of scene quickly - this introduced a sudden changes in both the image and the positional data that could be detected to compute an offset (latency) between the two. The benefit of this approach were simplicity and no requirement for an additional phantom, however at a cost of manual extraction of the features in the images (due to the tissue’s
irregular appearance) and therefore lack of accuracy. This challenge has led to the later development of methods to typically include a specially designed phantom to facilitate easy automation and improve the latency estimate [26, 32, 58, 71, 76].

With a single-wall phantom imaged in a water bath, Treece et al. was able to automatically segment the line (reflections from the bottom of the phantom) and identify the distance changes from the line in both the US images and the position data while moving the probe up and down. The temporal latency can be identified as the minimal root-mean-square (RMS) error between the two distance measurements [76].

In Gobbi et al.’s work [26], by scanning a cross-wire phantom while moving the ultrasound probe laterally (from side to side), he performed a motion analysis to identify the primary moving axis in both the ultrasound images and tracking data with Principle Component Analysis (PCA). The temporal offset was then computed by minimizing a least-mean-square difference between the two motion signatures.

Nakamoto et al. proposed a simple design of point-based phantom to estimate the latency by least-square minimization of the distance between the point position in the acquired ultrasound image and its projected physical position in the 3D space [58]. A notable drawback of this approach however, is that it requires an ultrasound probe calibration (discussed in the next section) to be performed first (in order to localize the point target).

Also using a wall phantom, Rousseau et al. provided a more robust, flexible alternative to Treece’s approach [76] to improve the line detection with a Hough-Transform-based algorithm [71]. Because Hough Transform can detect lines in any orientations, one major advantage of the method was that it imposed no specific probe motion constraints. The temporal latency was computed by matching the position
sensor signals with the line parameters.  

Finally, all the aforementioned technologies assumed that the temporal latency is a constant offset and does not change during image acquisition. Gooding et al. broke this assumption by treating the temporal offset a variable that best aligned a set of 2-D images within the reconstructed volume [32]. Their approach aimed to minimize a registration error with respect to latency for each scan. Like Nakamoto’s approach [58], one disadvantage of this method is the requirement of a spatial, transducer calibration in order to reconstruct the 3D volume for registration purposes.

2.2 Ultrasound Probe Calibration

Affixing a tracking device onto the ultrasound transducer ascertains an accurate localization of the ultrasound probe. However, knowing the position of the ultrasound 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 a process known as ultrasound probe calibration, where a homogeneous transformation is estimated to map the position of individual pixels from the ultrasound image frame to the ultrasound 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. Calibration is therefore a fundamental step and a single point of failure in a freehand, tracked ultrasound system.

In this section, we will review some fundamental aspects of the calibration process as well as the current state of the art in the calibration technologies. A recent and comprehensive overview of the US calibration techniques could be found in [50].
2.2.1 A General Calibration Formula

The calibration procedure is typically conducted by ultrasound scanning an artificial object with known geometries, referred to as the “phantom” - exception to this were the recently proposed phantomless (also referred to as self-calibrating) calibration techniques [3, 8], where images from actual patient were used instead of a specific calibration phantom. The fundamental idea behind a phantom-based calibration is to identify features in both the acquired images and in the physical phantom space (which is known to us by construction). With both the position of the transducer and the phantom tracked by a localization system, an equation can then be built to converse between these two coordinate systems.

To solve for the calibration parameters, a general approach adopted is to employ least-mean squares to minimize the distance between the features of interest (either points or lines) in the image space and the phantom space. If an exact correspondence of the features between the two spaces can be established, a closed-form solution is generally preferred; if on the other hand, the precise location of features in the phantom space is unknown (which is the case for some phantom designs), then a method based on iterative regression must be used [50].

2.2.2 Phantom Designs and Methods

Many phantom have been crafted for calibration. Regardless of the versatility in design, all phantoms share one thing in common: their purpose is to introduce a set of unique features that could be systematically recognized in both the phantom geometry and the ultrasound images. Based on the underlying principle, the feature appearance in the image and whether or not transducer alignment is required, the
phantom designs can be roughly divided into the following categories [50]: single-point or cross-wire phantoms, multiple-point phantoms, three-wire phantom, wall phantoms, and N-wire (or Z) phantoms.

**Single and Multiple Point (Cross-wire) Phantoms**

These phantoms employed either point targets or cross wires to form single or multiple dot appearance in ultrasound images [1, 4, 20, 41, 45, 55, 49, 67, 77, 78]. The dots were manually extracted from the ultrasound images and their physical positions in the phantom place identified by either the construction of the point targets or the ends of the wires that made the crossing knots.

One exception to this is the method proposed by Muratore et al. where instead of using a phantom, they directly scanned the tip of a pre-calibrated stylus probe tracked in realtime by a localizer [55].

One critical requirement for this type of technologies was a proper alignment of the ultrasound image plane with the point targets or wire crossings, which was also their major disadvantage because:

- (a) It is difficult to manually align multiple targets in one image, unless specially engineered tools are used [55], which may also add to the complexity of the system;

- (b) Accurate sampling of the ultrasound image plane using a point target (or wire crossing) is always challenging with the presence of ultrasound elevation beamwidth (subsection thickness), which may introduce localization errors [28] and subsequently diminish the calibration accuracy.
Since an exact correspondence can be established between the image points and their respective positions in the phantom, both iterative [1, 4, 20, 41, 49, 55, 67] and closed-form [45, 77, 78] methods were used to solve for the calibration transformation.

**Three-Wire Phantom**

The three-wire phantom alleviated the aforementioned alignment issue by introducing three orthogonal wires crossing at one single point, which effectively established a Cartesian coordinate system of its own [67]. The wires were labeled as the three base axes of the coordinate frame and can be scanned one at each time, with the ultrasound probe oriented freehandedly at different angles to the wire. The calibration algorithm must be informed which wire appeared in which image, in order to construct correct equation with respect to different wires.

As there was no way to tell precisely where the ultrasound image plane intersected the wires, this calibration technology can only use an iterative approach to determine the calibration parameters.

**Wall Phantoms**

Wall phantoms were the first automated and commercially available calibration technologies. The most representative works are the single-wall phantom [67], the membrane phantom [44] and the Cambridge phantom [67] (which was patented and made to the commercial market in 1997).

All wall phantoms produced a line-alike object in the image by scanning a diffusive, flat surface (e.g., the bottom of a box or a rubber membrane), which was more appealing to automation than a point-based calibration because a line contains more
(redundant) information than a point for an automatic segmentation to discover reliably. Also, the requirement for a diffusive surface, as discussed in previous sections, is crucial to backscatter the ultrasound signals to the transducer for detection.

Like the three-wire phantom, no one-to-one correspondence of features can be established in wall phantoms, therefore only an iterative method was adopted for calibration.

**N-wire (Z) Phantoms**

N-wire (Z) phantoms are among the most easy-to-use, truly freehand calibration technologies available today [13, 16, 35, 46, 62, 84]. Except by scanning a set of wires (arranged in a N-shape) in a completely unconstrained motion, no alignment or any other imaging conditions are imposed.

The fundamental principle was originally proposed by Brown *et al.* [11] to construct a stereotactic head frame for use with computed tomographic (CT) scans in neurosurgeries, and later introduced to ultrasound calibration by Comeaul *et al.* [19] and Pagoulatos *et al.* [61]. The idea is to precisely match ultrasound scan of a set of N-shape wires to their corresponding, physical position in the phantom geometry by using the *Theorem of Similar Triangles*.

Since the freehand data acquisition was efficient and fast with the N-wire phantoms, closed-form solution was exclusively used and successful automation of the calibration process aimed for intraoperative use have been reported [16, 35, 46].
CHAPTER 2. BACKGROUND AND LITERATURE OVERVIEW

2.2.3 Iterative and Closed-form Solutions

Classification

In general, the calibration problem is solved by establishing an overdetermined system (the number of equations are greater than the number of unknowns) to find a set of parameters that yields the least-mean-square error (minimum residual error) to all equations. There are two types of methods used, either iterative or closed-form (non-iterative) [50]:

- **Iterative** approach: when an exact correspondence (the physical position) to the features extracted from the ultrasound images cannot be found in the phantom geometry, the calibration problem has to be solved iteratively, i.e., an initial estimate of calibration parameters is first “guessed” to start an regression process with the errors repeated evaluated in each iteration until it is sufficiently smaller than a pre-determined tolerance level (a threshold). The most commonly used optimization algorithm here is the “Levenberg-Marquardt” algorithm for non-linear least-square fitting [53].

- **Closed-form** solution: when both the image points and their corresponding positions in the phantom are known to us, the calibration parameters can be solved in a closed-form (i.e., non-iteratively) [23].

The following summaries the use of iterative or closed-form solutions by different calibration technologies:

- Single/Multiple Point (Cross-wire) Phantoms: Used both iterative and Closed-form methods;
• Three-wire Phantoms: Used only iterative methods;

• Wall Phantoms: Used iterative solutions predominantly, but a closed-form method was recently proposed.

• N-wire (Z) Phantoms: Used closed-form solutions exclusively.

Iterative Versus Closed-form Approaches

Iterative approaches are, in general, less robust than closed-formed solutions due to the non-guaranteed convergence, local minima, and sensitiveness to initial estimates [23].

Efficiency-wise, to achieve a similar accuracy, iterative methods typically need more input images than closed-form techniques. For instance, calibration with the Cambridge phantom [68] (an iterative approach in the wall phantom group) would require an input of at least 550 US images, as compared to less than 30 images with a typical N-wire phantom [13, 16, 46, 62, 84]. Using a Sandwich phantom, Boctor et al. [6] proposed a closed-form calibration method that requires as few as only three poses of US images to achieve a stable solution. More recently, Najafi et al. developed a closed-form solution for a wall-phantom based calibration using differential measurement of the line slope instead of the absolute position of the line itself [57].

Manual Versus Automated Procedures

Even though closed-form calibration methods are generally preferred to iterative approaches, one major challenge limiting their intraoperative usage is the difficulty to fully automate the segmentation on the US images. The fundamental idea behind a
closed-form calibration approach is to map a set of geometric features (e.g., points, line slopes, etc.) in the US images to their corresponding physical positions on the calibration phantom. These points therefore need to be accurately extracted (or segmented) from the US images for their pixel locations. However, given the generally poor visibility and the abundance of speckles in US images, automatic extraction of individual points from the image remains difficult, lacking accuracy and robustness. Therefore, most of the current closed-form techniques choose manual over automated procedures simply to ensure the segmentation accuracy, which nevertheless compromises the performance and is undesirable in the OR.

There have been a few attempts made to automate the segmentation on images acquired from N-wire phantoms [16, 35, 46]. Lindseth et al. [46] were among the earliest trying to automate the segmentation process in their closed-form calibration system using a bead phantom, a Diagonal phantom, and a pyramid N-wire phantom. They introduced some spatial constraints (geometry patterns) unique to individual phantoms to help simplify the segmentation process. However, their method (for the Diagonal and N-wire phantoms) is not completely automatic because it requires finding an image point in a manually specified region as an initialization of the algorithm. Also, their segmentation results have a much larger variation compared to a manual method, indicating the accuracy and robustness of the algorithm still leave much to desire.

In one of the recent developments, Hsu et al. proposed a real-time automatic segmentation method using a modified N-wire phantom [35]. They introduced a 1 mm-thick translucent 40-degree shore-A-silicone rubber membrane clamped under tension on top of the N-wires. Since this membrane appears to be a straight line in
the image that can be automatically segmented with relative easiness, the positions of the N-wires are hence at known depths beneath the membrane by the phantom’s precise construction. Nevertheless, the assembly of this membrane adds more complexity to the phantom design and several manual interferences are demanded in their process: layers of search regions need to be specified in order to locate wires below the membrane, and to do so, the algorithm requires to know the actual scale factors of the US image which have to be manually measured using tools provided by the US machine.

In contrast to making the phantom clouded with wires or become more complicated, we focused our effect toward another direction: a much simplified N-wire-based phantom with only two layers of wires (called Double-N phantom) to facilitate easy automation. Based on this design, we were able to develop a fully automatic, realtime calibration system with no human interventions at all [16]. Through extensive tests on 10,000 freehand images, we found that to converge to a good accuracy, our system always required fewer than 60 images with an average of two data points (corresponding to two sets of N-wires) each, as compared to 30 images with an average of four N-wire sets per image [62, 83, 13] or six images with an average of 19 N-wire sets per image [35]. In our work and related works, we observed 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 interventions in the image-segmentation stage. We have traded off the number of images needed (more) against human intervention (zero) and phantom complexity (minimum).
2.2.4 Calibration Outcome Evaluation

The calibration outcome is typically evaluated based on two criteria: *precision* and *accuracy*. It is important to note that a measure of precision is quite different from that of accuracy [50]:

- Precision evaluates the repeatability and consistency of a system behavior (e.g., the variance in the system output running independently for many times given the same input), while

- Accuracy addresses how much the system is deviated from a known “ground truth” or “gold standard” typically measured independently.

Not relying on a ground truth, a high precision (indicated by a high consistency and low variance in results) does not necessarily guarantee a corresponding high accuracy, and it is possible that the calibration system may risk having a high consistency in the results while at the same time being trapped inside a local minimum (e.g., deviated by a systematic error) without even knowing that.

For example, it is possible that a calibration system that achieves highly consistent results may include a systematic error that renders the system inaccurate.

**Evaluation of Precision**

Precision examines the tightness (or diverging pattern) of repeated measurements on the same target using the tracked ultrasound. Statistical analysis of calibration precision involves the computation of the standard deviation, root-mean-square errors, confidence interval or a range for repeated measurements.

One of the most commonly adopted methods was to ultrasound scan a point target
(e.g., a crossing wire) from various possible angles and at different imaging depth, and then project each of the segmented point to the world coordinate system using the existing calibration parameters. This would ultimately form a cloud of reconstructed points for which the centroid and deviation may be calculated [13, 20, 41, 45, 62, 68].

**Evaluation of Accuracy**

Direct evaluation on calibration accuracy is typically challenging due to the lack of a reliable way to obtain the exact spatial relationship between the US image plane and the probe. A common walk-around is to reconstruct a cloud of points from the US image to the world coordinate system using the estimated calibration parameters, and then compared to their known physical locations (the gold standard) to compute the mean residual error [50]. This type of evaluation is typically referred to as “3D point reconstruction error (PRE)”, and can be extended to more complex structures: i.e., to scan a specially designed phantom [46] or even simpler, the calibration phantom itself [16, 62].

The PRE of ultrasound calibration, reportedly in a range from 0.6 mm to 4.9 mm [16, 36, 37, 46, 55, 62, 65, 83], provide a sound estimate on how accurately a tracked US would reasonably perform in a clinical application.

**Realtime Quality Assurance of Calibration Accuracy**

One essential element that the current calibration systems typically lack for intra-operative use is automated real-time feedback and control of accuracy. Calibration and validation are commonly two-phase tasks and remain isolated to each other in conventional techniques. First, a calibration is performed followed by a validation;
hence, the only way for the surgeons to possibly improve an inaccurate calibration outcome reported by the validation procedure is to recalibrate, which is not only time-consuming but also, more critically, lacking assurance to warrant a satisfactory result in the repeated procedure.

Another significant advantage of automatic error retrieval is that the accuracy test could be conducted quickly for an extensive data set on a large number of experimental conditions to thoroughly investigate and validate the calibration system.

Boctor et al. were among the first to address real time quality control in calibration [6, 7, 8]. With a Sandwich phantom, they used a Bootstrapping method to loop a closed-form calibration algorithm and a real-time validation procedure to minimize the standard deviation of a 3D reconstruction error [6]. Further, Boctor et al. introduced an in-vivo quality control mechanism that monitors the consistency in calibration parameters through frequent recalibration in the background [7, 8]. There is, however, one drawback of this quality-control mechanism: their evaluations on the calibration results are primarily based on precision, instead of accuracy.

To alleviate this problem, we have developed an automatic feedback and control system of the calibration accuracy based on 3D point reconstruction error that was calculated in real-time against a known ground truth during the calibration process [16]. This error was then updated and displayed in instantly to the user via a graphical interface. After the error converged or fell below a desired threshold, the calibration system provided an interactive dialog interface for the user to terminate the process and output the final calibration result.
2.3 Characterization of Ultrasound Beam Pattern

The best way to approach an error is to trace the origin of it - the invisible US beamwidth. As such, many methods and devices have been invented to detect and measure the US beam pattern. Some of these measurements are quantitative, but most methods aim to provide a qualitative assessment on the general US imaging quality.

2.3.1 Measurement of Axial Resolution

The axial resolution in US imaging is the minimum separation (distance) of two objects along the axial axis that could be distinguished by two separate echoes [33]; it also determines the smallest object detectable by the sound wave in the wave propagation direction. In theory, the axial resolution is equal to half the spatial pulse length (SPL), which is the product of the US wavelength and the number of cycles. Because the number of cycles is typically fixed by the transducer design, the axial resolution is proportional to the sound wavelength and is independent of the image depth.

In general, the higher the operational frequency of the transducer, the smaller the wave length and the better the axial resolution. For example, for a broadband linear-array transducer operating at a central frequency of 12.5 MHz, the axial resolution is approximately 0.1 mm; thus, any object of a dimension greater than 0.1 mm along the axial direction should be resolvable by the sound wave.
2.3.2 A Hydrophone Approach

One of the earliest and perhaps the most accurate methods available to measure the US beamwidth in both lateral and elevation directions involves the use of hydrophone (an US detection device) to detect the US intensity in an open tank of water [33]. The hydrophone is typically mounted on a highly sophisticated and calibrated mechanical device allowing it move in and out of the US beam at various depth to measure the sound intensity. The size of the hydrophone therefore needs to be smaller than the beamwidth itself (typically in the order of 1 mm) and the displacement should be accurately measured by a micrometer in order to sample the entire US field.

However, this type of measurement is expensive, time consuming and impractical to be widely used in a clinical facility. For this reason, other simple, image-based methods and devices (commonly referred to as “test objects”) have been developed to provide fast, mostly qualitative evaluation of the US resolutions.

2.3.3 Measurement of Lateral Beamwidth

For a linear crystal array, the multiple-element transducer electronically focused and swept the US beam in the field of view by sequentially firing the crystals in groups. The sound beam had a finite width laterally so, if an object was smaller than the beam width, the object would reflect numerous echoes back to the transducer as the sound beam swept across it. Since the image formation and representation of the echoes closely and continuously follow the motion of the beam, all these echoes would register a short line-alike shape in the display, with the length of the line directly proportional to the lateral beam width at that image depth. Therefore, because of the lateral beam width, point-shaped small objects were misrepresented and appeared
as short lines in the image. The physical length of the line is equal to the effective lateral beamwidth at that certain axial depth.

The underlying principle has been widely adopted to devise test objects for measurement of the lateral beamwidth and resolution [33, 42]. One of the earliest such test objects was constructed by American Institute of Ultrasound in Medicine (AIUM) in 1974 to perform various quality assurance (QA) tests on diagnostic US imaging, which included a qualitative assessment of the lateral resolution [27].

Currently, some of the most widely used, commercially available QA phantoms are manufactured by ATS Labs Inc. (Bridgeport, CT) as models ATS 520 (hydrogel) and ATS 539 (urethane) [22].

2.3.4 Measurement of Elevation Beamwidth (Slice Thickness)

To measure the elevation beamwidth or section thickness, the only widely used technique was based on a principle originally proposed by Goldstein more than 20 years ago [28]. This method and many of its variants [17, 33, 69, 73, 75] take advantage of the section-thickness artifacts generated when a sound beam interrogates a diffusive (non-specular) inclined plane. This principle has been subsequently adopted in the current industry-standard tissue-mimicking phantom commercially available through ATS Labs Inc. (Bridgeport, CT) as model 538N [33, 42], for US quality control purposes.
2.4 Use of Slice Thickness in Probe Calibration

In the past decade, only a few authors [6, 17, 36] have presented methods to improve the accuracy of US calibration beyond the US slice-thickness limit. Boctor et al. developed an image-based method to estimate the out-of-plane parameters of the US calibration by maximizing the similarity between the US image and the geometrical model of their sandwich phantom [6]. They introduced a real-time in vivo quality control mechanism that monitored the consistency of the calibration parameters.

In our preliminary study, we incorporated the US beamwidth into the calibration using an Unscented Kalman Filter (UKF) algorithm [17], which reduced the variance in the parametric estimate. However, both aforementioned approaches only improved the calibration precision, not the accuracy. It is important to note that a measure of precision is quite different from that of accuracy [50]: Precision defines the repeatability and consistency of the system, whereas accuracy evaluates how far the output is away from a known “ground truth” (typically measured independently). A low variance in the results, which is synonymous with high precision, does not guarantee a high accuracy, which can only be measured using a ground truth. For example, it is possible that a calibration system that achieves highly consistent results may include a systematic bias that renders the system inaccurate.

With a N-wire phantom, Hsu et al. were among the first to show an actual reduction in the elevation error of calibration by scanning a precisely mounted planar membrane at approximately the same oblique angle on each side [36]. However, the work assumed a uniform section thickness - this is not the case for a mechanically focused elevation beam pattern, as the beamwidth varies significantly with axial depth.
Chapter 3

Automated, Freehand Calibration
With Accuracy Feedback


Statement of originality: In this work, the complete calibration system design and development, the calibration phantom design, the real-time calibration algorithm, the automated calibration accuracy evaluation method, and the design of all scientific validation experiments are the sole innovations and contributions of T. K. Chen. A. D. Thurston contributed to the image segmentation algorithm. P. Abolmaesumi and R. E. Ellis were the principal investigators who funded and supervised the project.
3.1 Introduction

This chapter describes a fully automatic, real-time, freehand ultrasound calibration system. The system was designed to be simple and sterilizable, intended for operating-room usage. The calibration system used an automatic error retrieval and accuracy control mechanism based on a set of ground-truth data. An important part of the system was a Double-$N$ calibration phantom of a simple, sterilizable design.

The remainder of this chapter is organized as follows. First, an overview of the system design is given. Then, the design and specifications of the Double-$N$ calibration phantom are presented, along with an in-depth analysis of how the ultrasound resolution would play a major role in our selection of $N$-wires. Then, our experimental setup in image acquisition and position tracking, as well as the coupling mechanism between these two processes are introduced. Further, the robust and fully automatic segmentation algorithm is explored step by step. We then discuss the details of our closed-form calibration solution, and illustrate the automatic error retrieval and real-time accuracy control for the calibration system. Finally, extensive validation results are demonstrated, followed by a brief summary in the end.

3.2 Methods and Materials

3.2.1 System Overview

This section presents a high-level overview of our proposed ultrasound calibration system, along with a description of its hardware setup.
Figure 3.1: The design of the real-time automatic ultrasound calibration system. US images are acquired from *Double-N* calibration phantom and tracked in real time; N-wire pixel locations are then automatically extracted by the segmentation algorithm; Segmented points and their corresponding physical coordinates in the phantom are fed to a closed-form method to obtain the calibration parameters; Measured in PRE and displayed in real time to the user, the calibration accuracy is fed back to the control loop to determine if it is satisfactory; Once PRE converges or reaches satisfaction, an interface is provided to the user to stop the process and output the calibration result.
Design of the Real-time Automatic Calibration System

Figure 3.1 shows the design of our calibration system that consists of five successive stages.

1. Serving as input, ultrasound images were continuously acquired from the Double-$N$ phantom. The position of the ultrasound probe was tracked by a camera system.

2. The pixel locations of the cross section of N-wires in the ultrasound images (denoted as N-fiducials) were then extracted in real time by a fully automatic segmentation algorithm.

3. The successfully segmented N-fiducials, together with their corresponding physical coordinates collected in the phantom space, were fed to a closed-form formula to calculate the calibration parameters.

4. Measured by point reconstruction error (PRE) against a known ground truth, the accuracy of the current calibration result was fed back to the control loop to determine whether or not it was satisfactory. The PREs were updated, monitored and displayed in real time to the user.

5. Once PRE converged or reached an acceptable level, an interactive interface was provided to the user to terminate the process and save the calibration result.

Hardware Configurations

Figure 3.2 shows our hardware configuration for the automatic real-time calibration system.
Figure 3.2: Hardware configuration for image acquisition and tracking. Images were generated by a GE Voluson 730 Expert ultrasound machine ((a)-right). The ultrasound probe was a 192-element GE SP10-16 wide band linear-array transducer operating at a central frequency of 12.5MHz, with imaging depth of 3.4 cm and 4 focal zones activated (c). The image-acquisition frame rate was set at 24Hz. A 4-marker Traxtal VersaTrax Active Tracker was mounted on the ultrasound probe (red circle in (a)-right) and tracked in real time by a NDI Optotrak Certus Optical Tracking System ((a)-left) which employed three infrared sensors (b).
• Images were generated by a ultrasound machine (Voluson 730 Expert, General Electric Canada, Mississauga, ON, Canada), shown in Figure 3.2(a)-right, then fed to a frame grabber (ATI All-In-Wonder 7500, ATI Technologies Inc., Markham, ON, Canada).

• The ultrasound probe used in our experiments was a 192-element linear-array transducer (SP10-16 Wide Band, General Electric Canada, Mississauga, ON, Canada) that operated at a central frequency of 12.5 MHz with imaging depth of 3.4 cm and 4 focal zones activated (Figure 3.2(c)). The frame rate for image acquisition was 24 Hz.

• A four-marker optical target (VersaTrax Active Tracker, Traxtal Inc., Toronto, ON, Canada) was mounted on the ultrasound probe and tracked in real time by a camera (Optotrak Certus Optical Tracking System, Northern Digital Inc., Waterloo, ON, Canada), shown in Figure 3.2(a)-left. Certus employs three sensors (Figure 3.2(b)) to track infrared signals emitted by the optical targets, and has a reported root-mean-square (RMS) accuracy of 0.1 mm in X and Y axes, and 0.15 mm in Z-axis, all measured at a stand-off distance of 2.25 m [38].

• The central processor was a desktop workstation (Dell Optiplex GX270, Dell Canada, North York, ON, Canada) with an Intel Pentium 4 2.6GHz CPU and 2GB SDRAM running Microsoft Windows XP Professional.
Software System Design and Specifications

From a software-architecture point of view, the calibration system was designed and developed using a multiple-component-based object-oriented methodology. It encompassed four essential system components: ultrasound image acquisition and tracking, automatic segmentation, closed-form calibration, and real-time accuracy feedback and control. A number of open-source software libraries were employed, including the Visualization Toolkit (VTK), QT framework, Vision Numerics Library (VNL), and Microsoft DirectShow. Figure 3.3 demonstrates the high-level component-based system design and implementation in unified modeling language (UML) [9].

3.2.2 The Double-\textit{N} Calibration Phantom

Material and Constructions

The Double-\textit{N} phantom consisted of a front and a back plate connected by two side walls, forming a simple open-ended box and measured as 105 mm $\times$ 70 mm $\times$ 50 mm in dimensions (Figure 3.4).

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 the reassembling procedure, the plates are designed in plain rectangle-shape and made of 5 mm-thick aluminum. There are six holes on both the front and back plate to mount the upper and lower layers of N-wires that are 10 mm apart. Each set of N-wires is 20 mm wide (measured as the distance between the two parallel wires), which best accommodates the size of the ultrasound transducer in use. The ends of nylon wires are affixed using silicon glue which not only ensures sufficient tension of the wires but is also quickly
CHAPTER 3. AN AUTOMATED REALTIME CALIBRATION SYSTEM

Figure 3.3: A High-Level Object-Oriented Design Graph of the Calibration System in industry-standard Unified Modeling Language (UML). The calibration software system contains four essential components: ultrasound image acquisition and tracking, automatic segmentation, closed-form calibration, and real-time accuracy feedback and control. A number of open-source frameworks were employed, including the Visualization Toolkit (VTK), QT framework, Vision Numerics Library (VNL), and Microsoft DirectShow.
Figure 3.4: Front and side views of the *Double-N* phantom. The phantom consists of only a front and a back plate plus two side walls to form a simple cubic pipe, measured as 105 $mm \times 70\ mm \times 50\ mm$ in dimensions. The plates are designed in plain rectangle-shape and made of stainless steel at 5 mm thickness. The front plate has an extended arm to mount a spatial localizer for tracking purposes.
detachable (when heated) for easy sterilization. The front plate had an extended arm to mount a spatial localizer for tracking purposes. Figure 3.5 shows the CAD drawing of the phantom design.

Figure 3.6 illustrates the N-wire structure 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 [13]. With the nylon wire selected by empirically imaging various diameters, the Double-N phantom produces remarkably clean and well defined ultrasound images of N-fiducials that facilitated automated segmentation.

**Selection of Nylon Wires**

The typical cross-sectional appearance of a nylon wire in a ultrasound image is a small dot. Two important acoustic features make nylon an ideal candidate for N-wires:

- It has an acoustic impedance (2.9 MRayl) that is roughly that of distilled water (1.48 MRayl) and, as a result, a reflection coefficient of 0.32; this makes nylon an acceptable sound-reflective material in water.
- Nylon can be manufactured at a diameter comparable to ultrasound wavelengths, which are typically less than 0.5 mm for medical ultrasound [33], resulting in small but well-defined dots in the image.

We found that the image appearance of the nylon wire is largely determined by the axial and lateral resolutions in the ultrasound scan plane [33]. In general, the higher the operational frequency of the transducer, the smaller the wave length and the better the axial resolution. For our broadband linear-array transducer operating at a central frequency of 12.5 MHz, the axial resolution was approximately 0.1 mm;
Figure 3.5: CAD drawing of the Double-N phantom. There are six holes on both the front and back plate to mount the upper and lower layers of N-wires that are 10 mm apart.
Figure 3.6: Design specifications of the Double-N phantom. The phantom is 105 mm × 70 mm × 50 mm in dimensions. There are six holes on both the front and back plate to mount the upper and lower layers of N-wires that are 10 mm apart. Each set of N-wires is 20 mm wide (measured as the distance between the two parallel wires), which best accommodates the size of the ultrasound transducer in use.
thus, any object of a dimension greater than 0.1 mm along the axial direction should be resolvable by the sound wave. Figure 3.7 shows our experiment of imaging a nylon wire at 0.203 mm diameter across different image depths (the pixel resolution of the ultrasound image is 0.1 mm/pixel). Because the diameter of the nylon wire was greater than the axial resolution, the axial height of the wire was correctly represented in the image and remained the same at various depths because the axial resolution did not change with depth.

![Figure 3.7: Wire appearance determined by the ultrasound axial and lateral resolutions. The axial height of the dot is determined by the axial resolution which does not change with the image depth. The lateral width of the dot is directly proportional to the lateral resolution (the lateral beam width) at that depth and is depth-dependent.](image)

Similarly, the lateral resolution defines the ability to tell apart two objects located side-by-side in the lateral direction (perpendicular to the sound propagation in the
scan plane). Unlike the axial resolution, lateral resolution is determined by ultrasound lateral beam width and varies with the axial depth [42]. For a linear crystal array, the multiple-element transducer electronically focused and swept the ultrasound beam in the field of view by sequentially firing the crystals in groups. The sound beam had a finite width laterally so, if an object was smaller than the beam width, the object would reflect numerous echoes back to the transducer as the sound beam swept across it; this registered a short line-alike shape in the display, with the length of the line directly proportional to the lateral beam width at that image depth. Therefore, because of the lateral beam width, point-shaped small objects were misrepresented and appeared as short lines in the image. This phenomenon is evident in Figure 3.7, which shows the sonographic cross-sections of a nylon wire at various depths. As can readily be seen, the ideal circle-shaped cross-section of the nylon wire is distorted (except when at the focal depth).

The diameter of the nylon wire also has a significant effect on imaging. We compared the overall influence of both axial and lateral resolutions on nylon wires at different diameters, ranging from 0.203 mm up to 0.457 mm, all at the scan-plane focal depth (Figure 3.8). For our broadband transducer operating at a central frequency of 12.5 MHz (with a wavelength around 0.1 mm), we selected the nylon wire of 0.356 mm diameter as having the best overall definition (most circular) in both axial and lateral axes. Note that the choice of diameter may vary when another transducer with different operational frequency (wavelength) is used. We found in our experiments that the general rule of thumb is to use wire of diameter that is slightly larger than the operational ultrasound wavelength, then to choose a wire with the lateral width (at the image plane focal point) closest to the axial height in its image
appearance (or in other words, the dot appears to be more circular in shape).

Figure 3.8: Appearance of nylon wires of different diameters at the ultrasound scan-plane focal depth. The 0.356 mm wire gives the best (most circular) image definition.

3.2.3 Image Acquisition and Position Tracking

Experimental Setup for Image Acquisition & Tracking

Ultrasound images were acquired from the Double-N calibration phantom in a clean-water bath. Both the transducer and the phantom were rigidly affixed with optical targets and tracked in real time by the camera system. To approximate the speed of sound in tissue (1540 m/s), the water temperature was raised up to 37 degrees Celsius, in which sound travels at about 1570 m/s ([33], page 7). The ultrasound transducer was held freehand such that the acquired image would display the cross-section of the N-wires.
Position Tracking Coupled with Image Acquisition

Image acquisition and position tracking were performed with distinct hardware that were closely coupled by assigning position data to each captured image.

- The program was triggered by the image acquisition component when a new image was updated to its buffer; it would then attempt to read the position of the tracked ultrasound probe from the tracking component.

- If that attempt failed, the program would discard the image and go back to Step-1 to wait for a new image. This procedure was repeated until the position data were successfully retrieved.

- The system stored both the image and the position data, ready for the next acquisition.

As a result, an image was always tagged with data that recorded the positions of the optical targets affixed to the ultrasound probe and the Double-N phantom. Along with the image, these positions served as inputs to the calibration pipeline.

3.2.4 Fast, Robust, Fully Automatic Segmentation

This section details the ultrasound image segmentation method we developed to robustly extract the N-fiducials.

The Uniqueness in N-fiducial Geometry

The typical appearance of an N-fiducial is a single small dot in the ultrasound image, which is challenging for automatic segmentation (as opposed to a line object, which is
easier to segment and is the basis of wall-phantom-based calibration techniques [50]). A major difficulty is how to accurately and robustly recognize the N-fiducials in the presence of speckle, which has similar image intensities and shapes. The basic idea was to utilize two unique geometric features of N-fiducials in the image to assist the segmentation:

- the three collinear dots that form a typical N-wire intersection with the ultrasound image plane, and
- the two nearly parallel lines that pass through these two layers of dots.

The Automated Segmentation Algorithm

Our segmentation algorithm had four major stages involving various image processing techniques.

Stage 1. Morphological Operations

Two grayscale morphological operations [31] were sequentially applied to remove speckles and artifacts in the ultrasound image (Figure 3.9(a)).

The first operator removed any objects that were much larger than an N-fiducial, which effectively eliminated large speckles and artifacts. Since standard morphological operations can only remove objects that are smaller than a particular shape, we first find those big objects, and then subtract them from the original image. The morphological shape used for this operation was a rectangle bar of one pixel high and 18 pixel wide. The image was first eroded then dilated with this structural element. The combination of erosion and then dilation is commonly referred to as an opening operation in image processing [31]. This opening operation was applied
Figure 3.9: Results of the fully automatic segmentation algorithm. (a) Original ultrasound image with exemplified poor visibility and abundance of speckles and noises; (b) Results of the first morphological operation which removed speckles larger than N-fiducials; (c) Results of the second morphological operation that removed speckles smaller than N-fiducials; (d) Output of the pixel clustering where the dot locations were identified (in green circles); (e) Results of the line finding that discovered sets of lines (in red lines) passing through three identified dots; (f) The final segmentation after parallel-line searching where the two lines passing through the true N-fiducials were correctly extracted. The green circle indicated the segmented positions of the N-fiducials.
three more times with the structural element rotated 45, 90 and 135 degrees clockwise, respectively, which removed large structures along these orientations. Theoretically, more orientations could be used, but in practice we found the above was sufficient to remove large noise while preserving the dots (N-fiducials) of our interest. The final result was then subtracted from the original image. The outcome of the first morphological operation is shown in Figure 3.9(b). Take a note that the large speckles or artifacts presented in the original image were gone, left alone only the small ones in the background.

The second operator removed speckles that were too small to be N-fiducials. We have found, through our experiments with nylon wires at various diameters, that the N-fiducials in the image have a minimal diameter of at least 6 pixels (with an imaging resolution of 0.1 mm/pixel). Therefore, here the morphological shape was a circle at 2-pixel radius (or 0.4 mm in physical diameters) to effectively eliminate speckles that were smaller than that size, while making the remaining dots more uniform in shape. Similar to the first operation, an opening operation was applied with the morphological operation at four different orientations. After the second morphological operation, any remaining speckles in the resultant image were of a similar size to N-fiducials (Figure 3.9(c)).

**Stage 2. Pixel Clustering**

The image pixels were clustered to precisely identify the circle-shaped dots. The algorithm iterated through the image looking for a set (cluster) of pixels with similar intensities, and then explored the neighbouring pixels looking for more alike pixels. The search continued until all pixels in the image have been accounted for or a black boundary is reached. Afterwards, for each located cluster, a score was computed
that was the linear summation of the intensity at each pixel within the cluster. This approach effectively drove both the size and image intensity of the cluster to influence its score, a measure that would find its use in the later line-finding stage. Note if the image intensity is very low, the algorithm will reject the segmentation outcome due to its low score - a safeguard measure to reduce false-positive detections. Upon completion, all dots would be properly segmented and scored. Keep in mind that the identified dots here included not only the desired N-fiducials, but also similarly shaped and sized speckles (Figure 3.9(d)).

Stage-3. **Collinearity Discovering**

At this stage, we began to harness the special geometric relationships between the N-fiducials: we searched for collinearity of lines, i.e., sets of three dots that were collinear. Possible candidates were those segmented and scored dot locations from the clustering stage. A set of three distinct dots were accepted for further processing if the error in drawing a line through them was among the smallest in the overall ranking. Nevertheless, it would be unnecessary and inefficient to comb through all the candidates given the observation that majority of the segmented dots were those remaining speckles with typically lower intensity and smaller size than the N-fiducials. Hence, to speed up the search, we only processed and discovered those individual lines (drawn in red) passing through sets of three segmented dots that were higher ranked during pixel clustering (Figure 3.9(e)).

Stage-4. **Parallel-Line Searching**

Finally, we located a pair of detected lines that were parallel to each other—the second unique relationship between N-fiducials. The simplest way to do that was to
discard any pair of lines with an orientation difference that was greater than a small threshold. In our experiments, we chose an acceptable angle difference to be less than 2 degrees, which has proven to be more than sufficient in getting rid of the outliers. As a final result of segmentation, the two lines passing through the true N-fiducials were all correctly identified, as in Figure 3.9(f).

3.2.5 The Closed-form Calibration Method

A Closed-form Solution for Calibration Parameters

A closed-form solution is any formula that can be evaluated using a fixed number of standard operations. In the case of calibration, the unique N-wire geometry of the Double-N phantom provided a series of 3D spatial frame transformations, from which we solved for the calibration parameters in a single closed-form equation (Figure 3.10).

Let $^A X$ and $^B X$ denote a 3D position $X$ expressed in coordinate frame $A$ and $B$, respectively. $^A T_B$ then represents a homogeneous transform [72] that maps $^A X$ to $^B X$, as in:

$$^B X = ^{^A A} T_B . ^A X. \tag{3.1}$$

The sole objective of a ultrasound probe calibration is to determine $^P_I T$, the homogeneous transform that brings a position from the ultrasound image frame ($U$), to the ultrasound probe frame ($P$). To start, we acquired a set of 2D ultrasound images from the Double-N phantom. The intersection point of a wire and the ultrasound image plane would display a gray-intensity dot in the image, which could be expressed in both the ultrasound image frame ($U$) and the Double-N phantom frame ($H$) as
Figure 3.10: Frame transformations to solve for calibration parameters. The geometry of Double-N calibration phantom helps to bridge the segmented points in the ultrasound image frame with their physical positions in the phantom space, through which a closed-form equation could be established.
\( U \) and \( H \), respectively. From Figure 3.10, we have

\[
H = T_P T_U X. \tag{3.2}
\]

For non-singular homogeneous transforms, \( (B^{-1}A) = A^{-1}B \), so Equation (3.2) is equivalent to

\[
P_U T_U X = P_T T_H T_H X. \tag{3.3}
\]

On the right side of the Equation (3.3), \( P_T \), the transform from the tracker frame to the ultrasound probe frame, was known from the optical target mounted on the probe. \( T_H \), the transform from the phantom frame to the tracker frame, could be obtained by registering the Double-N phantom geometry to the optical target affixed to the phantom. On the left side, \( U \) could be measured as the N-fiducial positions in the ultrasound image frame. Hence, if we could locate \( H \), \( U \)'s corresponding positions in the Double-N phantom frame, we could solve for \( P_U \) using a least-mean-square method. To find \( H \), we used the Double-N phantom’s N-wire geometry.

Finding \( H \) Using the Unique N-wire Geometry

Locating \( H \) given \( U \) can be done by using the distinctive N-wire geometry of the Double-N phantom. The fundamental principle was first introduced by Brown et al. [11] to construct a stereotactic head frame for use with computed tomographic (CT) scans to provide guidance in neurosurgeries, and was later applied to closed-form ultrasound calibration [19, 61].

When an ultrasound image was acquired from the N-wire phantom in a clean-water bath, both the ultrasound transducer and the phantom were oriented in such that the resultant image would show the cross-section of the wires. For the Double-N phantom (Figure 3.10), there were six wires (labeled 1 to 6) that corresponded to
six N-fiducials in the image, comprising two sets of N-wires (denoted as NW$_{1-3}$ & NW$_{4-6}$).

To illustrate, consider NW$_{1-3}$, where the ultrasound image plane intersects the wires 1, 2 and 3 at position $X_1$, $X_2$ and $X_3$, respectively (Figure 3.11(a)), producing three bright corresponding dots in the ultrasound image (Figure 3.11(b)).

Because the Double-N phantom’s front and back plates ($AB$ and $CD$ in Figure 3.11(a)) were parallel to each other, the ratio $\alpha$ between the designated line segments is a consequence of the similar triangles formed by the N-wires:

$$\alpha = \frac{\|A - X_2\|}{\|A - D\|} = \frac{\|X_1 - X_2\|}{\|X_1 - X_3\|},$$

(3.4)

where the form $\|P_i - P_j\|$ represents the Euclidean distance between two points $P_i$.
and $P_j$. The position $X_2$ in the phantom frame ($H$) can be calculated as

$$^{H}X_2 = ^{H}A + \alpha \cdot (^{H}D - ^{H}A).$$  \hspace{1cm} (3.5)

Bear in mind that the same set of points ($X_1$, $X_2$, and $X_3$) also appear in the ultrasound image, albeit in a different coordinating frame and in units of pixels. However, because $\alpha$ is a ratio and thus remains invariant to frame transforms and units, its value could also be calculated directly in the ultrasound image frame (Figure 3.11(b)), as

$$\alpha = \frac{\|^{U}X_1 - ^{U}X_2\|}{\|^{U}X_1 - ^{U}X_3\|}. \hspace{1cm} (3.6)$$

Combining Equation (3.5) and Equation (3.6) yields

$$^{H}X_2 = ^{H}A + \frac{\|^{U}X_1 - ^{U}X_2\|}{\|^{U}X_1 - ^{U}X_3\|} \cdot (^{H}D - ^{H}A), \hspace{1cm} (3.7)$$

where $^{H}A$ and $^{H}D$ were measurable from the Double-$N$ phantom geometry, with $^{U}X_1$, $^{U}X_2$, and $^{U}X_3$ automatically extracted by the segmentation algorithm. As a result, we were able to calculate $^{H}X$ directly from $^{U}X$ in the acquired ultrasound images of the phantom.

The aforementioned shows how we found one $^{H}X$ position using one set of N-wires. In theory, the more points we collected for the least-mean-squares method to solve for $^{P}T$ in Equation (3.3), the more accurate the calibration outcome would be. In practice we acquired two sets of N-wires of the Double-$N$ phantom in a single image, using multiple freehand images. Note that $\alpha$ is unique for individual N-wire, so Equation (3.7) needed to be evaluated for each set of N-wires to yield a corresponding $^{H}X$. Due to the fast, automatic segmentation algorithm, all process could be done in real time.
Measuring the Double-N Phantom Geometry

We employed a pre-calibrated Stylus probe (HP005 Integrated, Traxtal Inc., Toronto, ON, Canada) to measure the phantom geometry. First, we probed a set of physical landmarks on the Double-N phantom that were precisely machined with a known geometry; then we registered this geometry to the optical target that was affixed to the phantom, by which the phantom coordinate frame was defined. The entire probing process was guided by an interactive graphics interface provided to the user and required a few seconds to accomplish.

A General Form to Solve for the Calibration Parameters

Generally, if we use the form, $^H X(r_i, c^1_i, c^2_i)$, to represent the $i^{th}$ point in the phantom frame that we have collected from the N-shape at row $r$ and between columns $c^1$ and $c^2$; and $^U X(r_i, c^1_i, c^2_i)$, the corresponding point in the ultrasound image frame, and with $\alpha_i$ calculated from Equation (3.6), we could rewrite Equation (3.7) as:

$$^H X(r_i, c^1_i, c^2_i) = A(r_i, c^1_i) + \alpha_i \cdot (D(r_i, c^2_i) - A(r_i, c^1_i)).$$ (3.8)

Equation (3.3) could then be rewritten in a general form

$$P^T U^T \cdot ^U X(r_i, c^1_i, c^2_i) = P^T T^T D^T H^T \cdot ^H X(r_i, c^1_i, c^2_i).$$ (3.9)

Bear in mind that both $^U X(r_i, c^1_i, c^2_i)$ and $^H X(r_i, c^1_i, c^2_i)$ are $4 \times 1$ column vectors in homogeneous format, so Equation (3.9) would be particularly useful for implementation because we can construct $^U X = [U^x \ U^y \ U^z \ 0]^T$ and $^H X = [H^x \ H^y \ H^z \ 1]^T$ in matrix
format using $U_X(r_i, c_i^1, c_i^2)$ and $H_X(r_i, c_i^1, c_i^2)$ as their respective column vectors:

$$U_X = \begin{pmatrix} U_X(r_0, c_0^1, c_0^2), & U_X(r_1, c_1^1, c_1^2), & \ldots, & U_X(r_i, c_i^1, c_i^2), & \ldots \end{pmatrix}$$

$$= \begin{pmatrix} u_{x0} & u_{x1} & \ldots & u_{xi} & \ldots & u_{xN} \\ u_{y0} & u_{y1} & \ldots & u_{yi} & \ldots & u_{yN} \\ 0 & 0 & \ldots & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 1 & \ldots & 1 \end{pmatrix}$$

(3.10)

and

$$H_X = \begin{pmatrix} H_X(r_0, c_0^1, c_0^2), & H_X(r_1, c_1^1, c_1^2), & \ldots, & H_X(r_i, c_i^1, c_i^2), & \ldots \end{pmatrix}$$

$$= \begin{pmatrix} h_{x0} & h_{x1} & \ldots & h_{xi} & \ldots & h_{xN} \\ h_{y0} & h_{y1} & \ldots & h_{yi} & \ldots & h_{yN} \\ h_{z0} & h_{z1} & \ldots & h_{zi} & \ldots & h_{yN} \\ 1 & 1 & \ldots & 1 & \ldots & 1 \end{pmatrix}$$

(3.11)

Note the difference in the 3rd rows between matrices of $U_X$ and $H_X$. Points in the US image frame do not have a z-coordinate, so without losing generality we use all zeros for their 3rd components.

In all, what Equation (3.9), (3.10) and (3.11) have established is an overdetermined system for $P^T$, which can be solved using a straightforward implementation of least mean squares.

3.2.6 Automatic Accuracy Feedback and Control

Automatic Evaluation on the Calibration Accuracy

Direct evaluation of calibration accuracy can be challenging because of the lack of a reliable way to obtain the exact spatial relationship between the ultrasound image
plane and the probe.

A widely used measure has been the 3D reconstruction error \([50]\) of a known object \([16, 36, 47, 46, 55, 62, 78]\). In particular, it is possible to scan a validation phantom \([46]\) or conveniently the calibration phantom itself \([62]\). We used the latter approach, first scanning the phantom and then comparing the reconstructed cross sections of the wires to their known locations in the phantom geometry. This was the Point Reconstruction Error (PRE):

\[
\|PRE\| = \|H_X - H^T T \cdot T^T U \cdot U^T X\|.
\]

(3.12)

where \(P^U\) was the calibration parameter to examine, \(T^P\) was the transformation from the US probe frame to the tracker frame given by the tracking device mounted on the probe, \(H_T\) was the transformation from the tracker frame to the phantom frame known by the tracking device affixed onto the phantom, \(U X\) was the identified position of the wire in a US image, and \(H_X\) was the corresponding wire location known by the calibration phantom design (the ground truth).

PRE is a Euclidean distance between two points in space, so it remained invariant to frame transformations. Hence, for easier visualization of the error distributions, we converted the PRE from the phantom space \((H)\) to the physical transducer frame defined with respect to the alignment of crystal arrays \((C)\) as in Figure 3.12(a), by series of rigid frame transformations:

\[
C_{PRE} = C^p_T \cdot P^T \cdot T^H \cdot H_{PRE}.
\]

(3.13)

where \(C^p_T\) is constructed using \(P^U\) but with unit scaling (so that the units of PRE remains in meters).

Using the aforementioned segmentation algorithm, this error measurement was
Figure 3.12: Real-time calibration accuracy feedback and control. (a) The definition of the physical transducer frame with respect to the crystal alignments, in axial, lateral and elevation axes; (b) The graphical user interface of the calibration system displaying in real time the FRE converging curves with respect to the number of input images, along the axial, lateral and elevation axes, respectively; (c) The dialogue interface provided to the user to terminate the calibration process and output the current calibration results if the FRE starts to converge or falls below a satisfactory threshold.
performed by automatically extracting the wire positions; one significant advantage of this automatic error retrieval was that the process could be performed quickly for a large set of data acquired in a variety of experimental conditions to examine the calibration results. This enabled us to rapidly test the calibration system through many trials, computing the errors in real time.

Because the same phantom geometry was utilized for both calibration and accuracy assessment, we separated the data used for PRE calculation from those used for calibration to avoid a biased assessment.

**Real-time Update, Feedback and Control of PREs**

PREs were updated and displayed in real time to the user via a graphical interface (Figure 3.12(b)). There were three pieces of critical information conveyed to the user during the calibration:

- The vertical axis was the PRE value, calculated on-the-fly during each iteration, and the horizontal axis showed the number of input images used to estimate these calibration parameters. We expected to see that with more and more input images, the calibration error becomes smaller and smaller (which is clear in this example).

- Three color-coded curves were displayed, each representing the PRE components along the axial, lateral and elevation axes. This gave the user a visual inspection of how the error was distributed along these three directions. In the example illustrated, it can be seen that the elevational component of PRE (the red curve) had larger values and greater dynamics than the other two components. Also, errors in the elevation axis dominated the overall calibration
error.

- With fewer input images initially, the PRE curves fluctuated significantly at the beginning but eventually smoothed down to a stable state as more data was acquired. This showed the converging pattern of the PREs as the number of input images increased. In the example illustrated, both axial and lateral components of the PRE started to converge after around 25 input images but it took 60 images for convergence in the elevation axis. In general, when all the PRE curves started to converge, it was usually a good indication that the calibration error had reached a stable minimum and the user could stop the calibration.

After all PRE curves either converged or fell below a desired threshold, the calibration system provided an interactive interface (Figure 3.12(c)) for the user to terminate the process and output the final calibration result.

3.3 Results

Extensive tests were conducted to evaluate the accuracy, precision, and performance of the segmentation method and the overall calibration process.

3.3.1 Segmentation Accuracy, Precision and Robustness

Validation of Segmentation Accuracy

For a test of 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.

**Segmentation Precision: Automatic vs. Human**

We compared the precision of the automatic algorithm with that of human operators by evaluating the variance in segmentation outcomes. A set of distinct ultrasound images (acquired freehand from the Double-N phantom at various image depths and with varying image quality) were provided to seven volunteer subjects for manual extraction of the N-fiducials: every image was segmented 10 times by each subject to obtain the standard deviations (in pixels) in the results. The subjects were selected such that their level of expertise in segmenting ultrasound image varies from low to high: two were new students with zero or minimal expertise to the task, three were graduate students with moderate expertise, and one graduate student and one post-doctoral fellow had years of experience in ultrasound segmentation and did segmentation on a regular basis. We then performed the same procedure on the automatic algorithm. The results are shown in Table 3.1.

As expected, the automatic algorithm was deterministic and had zero variance in its output. The human segmentations, on the other hand, had large variances for all subjects. The standard deviations were larger in the lateral direction, an unsurprising result given the poorer (than axial) lateral resolution. Variances were also greater for the less experienced subjects, also unsurprising: the standard deviations of individual
Table 3.1: Standard deviation (in pixels) in the segmentation outcome: automatic system vs. human operators.

<table>
<thead>
<tr>
<th>Standard deviation (in pixels)</th>
<th>Automatic</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>max</td>
</tr>
<tr>
<td>Axial</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Lateral</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Segmentation are shown in Figure 3.13.

![Figure 3.13: Standard deviations in manual segmentation for 7 human subjects. Subjects 1 and 2 were new students; subjects 3 through 5 were senior graduate students with moderate expertise; subject 6 was a senior graduate student expert in the task; and subject 7 was a post-doctoral fellow expert in the task.](image)

**Segmentation Performance and Speed**

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.
3.3.2 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 ultrasound 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 ultrasound 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 PRE from the first 80 images in each group of data, computing the means and standard deviations per group.

- For calibration, images were taken from the other 120 images in each group of data and then fed to the calibration system to calculate the calibration parameters.

**Calibration Accuracy and Robustness**

Table 3.2 summarizes the validation results of our calibration system, including the mean and the standard deviation of PREs of the 50 independent random trials, as well as the best and worst cases among all trials. For comparison, all results were illustrated along the axial, lateral and elevation axes, respectively.

Observations on the calibration system’s accuracy and robustness include:
Table 3.2: PRE for 50 independent calibration trials.

<table>
<thead>
<tr>
<th>Number of Trials: 50</th>
<th>Point Registration Error (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>axial</td>
</tr>
<tr>
<td>Mean ($\mu$)</td>
<td>0.14</td>
</tr>
<tr>
<td>Standard Deviation ($\sigma$)</td>
<td>0.12</td>
</tr>
<tr>
<td>Minimum (best case)</td>
<td>0.05</td>
</tr>
<tr>
<td>Maximum (worst case)</td>
<td>0.29</td>
</tr>
</tbody>
</table>

- The PRE on average were $0.14 \pm 0.12$ mm, $0.19 \pm 0.22$ mm, and $0.62 \pm 0.64$ mm, along the axial, lateral and elevation axes, respectively. Overall, in 49 out of the 50 independent experiments, the PRE converged to sub-millimeter in all of the axial, lateral and elevation axes and 5 yielded a PRE in excess of 0.9 mm; there was only one trial had a maximum observed PRE slightly exceeding 1 mm (1.02 mm in the elevational direction). This indicates the calibration system was able to consistently achieve sub-millimeter accuracy in 98% of cases.

- The mean of 3D PRE (measured in Euclidian distance) of all 50 independent trials was 0.66 mm, which is consistent with accuracy reported by related work with N-wire phantoms [62, 83, 35].

- The elevation axis had a much larger PRE mean and standard deviation (more than three times larger) than that of the axial and lateral directions. On the other hand, the errors and variances were always the smallest along the axial direction.

The obvious differences of PRE distributions among the axial, lateral, and elevation axes were the direct results of how the ultrasound resolution varies axially, laterally, and elevationally:
As discussed above, the axial resolution (around 0.1 mm for the transducer at
the frequency we used) was much finer than the lateral resolution, resulting in
less uncertainty and error axially than laterally.

The ultrasound elevation resolution was the dominant influence on calibration
accuracy. Similar to the lateral direction, the ultrasound beam had a finite
beam width in the elevation axis, referred to as section thickness [29], which
determined the elevation resolution. Unlike the lateral axis, there were no mul-
tiple crystals in the out-of-plane direction to provide electronic focusing. The
beam was therefore only focused mechanically in the elevation axis, by either
curving the crystal or placing an acoustic lens in front [33]. This resulted in a
much larger elevation beam width, which contributed to the larger uncertainties
and errors among all axes.

Calibration System 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.

3.4 Conclusion

In this chapter, we presented an automated, freehand US calibration technology with
instant accuracy feedback to the user. We have conducted validation tests on a data
set of 10,000 images in 50 independent calibration trials to thoroughly investigate the
accuracy, robustness, and performance of the calibration system. On average, the
calibration accuracy (measured in 3D reconstruction error against a known ground
truth) of all 50 trials was 0.66 mm. In addition, the calibration errors converged to sub-millimeter in 98% of all trials within 12.5 seconds on average. Overall, the calibration system was able to consistently, efficiently and robustly achieve high calibration accuracy with real-time performance.
Chapter 4

USB-FW: A Beamwidth-Weighted Calibration Framework

The content of this chapter has been published in the Journal of Ultrasound in Medicine and Biology as: T. K. Chen, R. E. Ellis, and P. Abolmaesumi, “Improvement of freehand ultrasound calibration accuracy using the elevation beamwidth profile,” Ultrasound Med. Biol. 37(1), 1314-1326 (2011). Text has been edited to best suit the flow of this thesis.

Statement of originality: This work is the sole innovation and contribution of T. K. Chen. P. Abolmaesumi and R. E. Ellis were the principal investigators who funded and supervised the project.

4.1 Introduction

In Chapter 1 (Section 1.2.2), we have discussed the cause of localization errors and uncertainty in US imaging as a result of the US slice thickness (elevation beamwidth).
Here we present a new framework that employs the US elevation beamwidth into a filtered, weighted-least-square framework to improve the reconstruction accuracy of the real-time freehand calibration system we have discussed in the previous chapter.

The remainder of this chapter is organized as follows. First, the principles and design specifications of the slice-thickness calibrator are presented, along with an in-depth analysis of the quantitative US elevation beamwidth profile extracted using the device. Then, a brief overview of the calibration system components are given including the calibration phantom design, image acquisition and position tracking. Next, we proceed to explore the details of the USB-FW framework that was integrated into the calibration system to improve the calibration accuracy. Further, the experimental setup to evaluate the effectiveness and performance of the USB-FW is discussed. Finally, the extensive experimental results are presented, followed by a brief summary in the end.

4.2 Methods and Materials

4.2.1 Ultrasound Slice-Thickness Extraction

To quantitatively extract the US elevation beam width, we adopted a method based on the principle originally proposed by [28] and used by [73] and [69]. The basic idea is to take advantage of the slice-thickness artifacts generated when a sound beam interrogates a diffusive (non-specular) inclined plane.

Figure 4.1(a) shows a cross-sectional view of a diffusive inclined surface being interrogated by a linear-array transducer in the elevation plane. Because of the finite elevation beam width, the sound beam will intersect the inclined surface from A to
Figure 4.1: (a) Principles of slice-thickness measurement using an inclined plane. (b) The slice-thickness calibrator designed based on the inclined plane. (c) A combined view of the echo bands and their height measurements from multiple US images: the height of the echo band represents the elevation beamwidth at that axial depth.

$B$ 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 from the transducer. However, because the US machine algorithms always assume that all echoes received are from reflectors located on the central beam axis [33], 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 of the echo band) would actually be equal to the effective elevation beamwidth at that axial depth [28]. Note this type of measurement provides only an approximation of the elevation beamwidth for two reasons. First, the length of $CD$ would only be equal to the slice thickness if the edges of the elevation beam are parallel to each other which is clearly not the
case. However given the small dimensions of the beam profile, the measured thickness would provide a close estimate. Second, the segmentation of the echo band in the image depends on the subjective judgment by the operator.

Based on this principle, we designed a slice-thickness calibrator (Figure 4.1(b)). The device was constructed using two aluminum anglers to form a 45-degree inclined plane and had physical dimensions of 30 cm (L) × 12.6 cm (W) × 15 cm (H). The inclined surface was made of a 0.35-mm-thick natural rubber membrane (manufactured by Rubbermaid Inc., Ogdensburg, NY, USA). Rubber is an excellent choice for the inclined surface for two reasons: first, it has an acoustic impedance (1.81 Mrayls) close to that of water (1.48 Mrayls), making it a weak but adequate sound reflector in water; second, the membrane has a rough surface that serves as a diffusive reflector to backscatter sound signals. The transducer was positioned perpendicular to the horizon using an air-bubble balancer. US images were acquired from the slice-thickness calibrator in a clean-water bath. To approximate the speed of sound in tissue (1540 m/s), the water temperature was raised up to 37° Celsius, at which sound travels at approximately 1570 m/s [33].

By measuring the axial height of the echo band in the display, we were able to obtain the slice thickness at that axial depth. Consequently, we made a series of such measurements at various depths that covered the effective imaging area to estimate the relevant elevation beam profile. As an example, Figure 4.1(c) shows a combined view of the echo bands and their height measurements from multiple US scans of the slice-thickness calibrator: the height of the echo band represents the elevation beamwidth at that axial depth.
4.2.2 Double-N Calibration Phantom

The Double-N phantom consisted of a front plate and a back plate connected by two side walls, forming a simple open-ended box that was 105 mm × 70 mm × 50 mm in size. 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 the reassembling procedure, the plates were designed in plain rectangle-shape and made of 5 mm-thick aluminum. There were six holes on both the front and back plate to mount the upper and lower layers of N-wires, which were 10 mm apart. Each set of N-wires was 20 mm wide (measured as the distance between the two parallel wires), a width which best accommodates the size of the ultrasound transducer in use. The ends of nylon wires were affixed using a silicon glue to ensure sufficient tension of the wires which can also be quickly detached (when heated) for easy sterilization. The front plate had an extended arm that was used to hold a spatial localizer for tracking purposes. For more details of the design of the Double-N phantom, please refer to our preceding work [16].

4.2.3 US Image Acquisition and Tracking

US images were generated by a General Electric Voluson 730 Expert 3D/4D US machine and then fed to an ATI All-In-Wonder 7500 frame grabber.

The US probe used in our experiments was a 192-element GE SP10-16 Wide Band Linear Array transducer that operated at a central frequency of 12.5 MHz with an imaging depth of 3.4 cm. This unusually shallow depth was selected because the initial target applications of our US technology included relatively superficial bones.
in the wrist and shoulder, for which a longer focus would provide poorer images.

Four equally spaced focal points were used to maximize the clarity of bone-surface imaging in the wrist, shoulder and other superficial joints of interest in orthopedic applications. The frame rate for image acquisition was 24 Hz.

A Traxtal (Bellaire, TX, USA) VersaTrax Active Tracker (4-marker optical target) was mounted on the US probe and tracked in real time by a Northern Digital (Waterloo, ON, Canada) Optotrak Certus Optical Tracking System. Certus employs three cameras to track infrared signals emitted by the optical targets, and has a reported accuracy of 0.1 mm root mean square (RMS) error in the X and Y axes, and of 0.15 mm in the Z-axis, all measured at a stand-off distance of 2.25 m [25].

4.2.4 Calibration with USB-FW

Figure 4.2 shows our design of the real-time, freehand, N-wire-based US calibration system with USB-FW, which consists of an offline beamwidth-extraction stage and a real-time calibration stage.

In the beamwidth-extraction stage (step A1 to A2 in Figure 4.2):

A1. By scanning the slice-thickness calibrator in a water bath, we first extracted the elevation beamwidth of the designated US transducer (operating at the same imaging settings as for the calibration) at various axial depths within the valid imaging region.

A2. This discrete (sampled) elevation beamwidth pattern was then linearly interpolated to generate a smooth, continuous elevation beamwidth profile of the transducer to cover the entire effective imaging depth.
Figure 4.2: Design of the real-time, freehand, N-wire-based US calibration system with USB-FW in two stages: offline beamwidth-extraction \((A1 \text{ and } A2)\) and real-time calibration \((B1 \text{ to } B4)\).

In the real-time calibration stage (step \(B1 \text{ to } B4\) in Figure 4.2):

- **B1.** Serving as input, US images were continuously acquired from the Double-\(N\) calibration phantom. The position of the US probe was tracked by an optical tracking system.

- **B2.** The pixel locations of the cross section of N-wires in the US images (the dot appearance of N-wires in a US image is defined as the N-fiducials [16, 46, 62]) were then extracted in real time by a fully automatic segmentation algorithm.

- **B3.** The successfully segmented N-fiducials were then processed by USB-FW in two consecutive steps: first, a filtering stage that got rid of image data with a large elevation beamwidth, and second, a weighted-least-square estimation of the calibration parameters from the remaining data, all weighted by their respective elevation beamwidth.
B4. Measured in 3D point-reconstruction error against a known ground truth, the
accuracy of the current calibration result was fed back to the control loop to
determine whether or not it was satisfactory. The calibration accuracy was
updated, monitored and displayed in real time to the user. Once the error
converged or reached an acceptable level, an interactive interface was provided
to the user to terminate the process and save the calibration result.

Implementation-wise, the calibration system with USB-FW encompassed four es-
sential system components (as shown in Figure 4.2):

a. Acquisition of tracked US images: US images were acquired from the Double-N
calibration phantom and tracked in real time by a camera system;

b. Automatic US image segmentation: By taking advantage of the unique geome-
try of the Double-N phantom, we have developed a fully automated algorithm
to extract the N-fiducials from the US images in real time;

c. Closed-form calibration with USB-FW: The elevation beamwidth profile was
incorporated into the USB-FW framework to compute the calibration param-
eters;

d. Real-time calibration accuracy feedback: The current calibration accuracy (mea-
sured in PRE) was calculated and displayed instantly to the user.

The technical aspects of components a, b, and d can be found in our preceding
work [16]. Here we will present the detailed design and implementations of our closed-
form calibration algorithm with USB-FW.

The elevation beamwidth profile was essentially a distribution of errors in locating
an object with the tracked US transducer. As shown previously in Figure 1.2, the
range of these localization errors is directly proportional to the elevation beam width, which varies with the axial depth. Without prior knowledge of the slice thickness, conventional least-squares-based calibration technologies [16, 36, 46, 62, 83] treated all of the input data equally; with this knowledge, we can now choose to ignore or downplay the influence of the data with relatively larger elevation beamwidths because these data are more likely to be image artifacts and to have larger localization errors. This idea translated to the design of USB-FW in two phases: beamwidth-based data filtering and closed-form calibration with beamwidth-weighted least squares.

**Beamwidth Based Data Filtering**

This was a straightforward data-sifting process, in which we got rid of potentially unreliable or corrupted input data based on their elevation beamwidth. In our study, for the US transducer that we used, the measurable elevation beamwidth profile covered the axial imaging depth from 4.4 mm to 31 mm.

Any input data with an elevation beamwidth that is greater than a preselected threshold, $T_h$, would be removed and not used for calibration. In our experiments, $T_h$ was defined as twice of the slice thickness at the elevation focal point:

$$T_h = 2 \cdot E_{focal}, \quad (4.1)$$

where $E_{focal}$ is the minimum elevation beamwidth at the elevation focal point. Our selection of $T_h$ was based on the US physics [33]: the beamwidth of a focused sound beam (with one focal point) will expand to twice the minimum beamwidth at the focal point when the sound wave travels to a distance that is twice that of the near field, at which point the ultrasound resolution becomes very poor and imaging quality deteriorates rapidly.
In our experiments, this procedure effectively filtered out the image data with very large elevation beamwidths, which possessed great uncertainty and potentially large localization errors. The surviving data from the filtering stage were then processed by a beamwidth-weighted closed-form calibration method.

**Beamwidth Weighted Linear Least Squares**

We have previously developed a non-weighted closed-form least-square method to solve for the calibration parameters (Section 3.2.5, Equations (3.9), (3.10) and (3.11)). In USB-FW we implemented a weighted linear least-square method to control the data’s influence on the calibration outcome with respect to their elevation beamwidth.

In addition to simplicity and efficiency, weighted least-squares is best known for its ability to handle regression situations in which the data points are of varying quality [5]. This fits our problem well: in general, the larger the elevation beamwidth, the more uncertainty in the data position, hence the greater possibility for large localization errors; as such, the image data of different beamwidths are of different quality (possessing unequal possibilities of errors) and should therefore be treated unequally.

For an overdetermined system with \( m \) linear equations and \( n \) unknown coefficients \( \beta_1, \beta_2, \cdots, \beta_n \) \( (m > n) \),

\[
\sum_{j=1}^{n} X_{ij} \beta_j = y_i, \quad (i = 1, 2, \cdots, m),
\]

where \((X_{ij}, y_i)\) is a pair of data sample collected for the \( i \)th linear equation. We can write Equation (4.2) in a matrix form as:

\[
X \cdot \beta = y,
\]

\[ (4.3) \]
where

\[
X = \begin{pmatrix}
X_{11} & X_{12} & \cdots & X_{1n} \\
X_{21} & X_{22} & \cdots & X_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
X_{m1} & X_{m2} & \cdots & X_{mn}
\end{pmatrix}
\quad \beta = \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_n
\end{pmatrix}
\quad y = \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_m
\end{pmatrix}
\quad \text{(4.4)}
\]

If the errors in the data are uncorrelated, with a mean of zero, but have unequal variances, we can find a set of \( \beta \) to best fit the equations given the data samples (\( X \) and \( y \)), by minimizing a sum of the squared residuals in a weighted linear least-square model [5]:

\[
\arg \min_{\beta} \sum_{i=1}^{m} w_i (y_i - \sum_{j=1}^{n} X_{ij} \beta_j)^2 = \arg \min_{\beta} \| W^{1/2} (y - X \cdot \beta) \|,
\quad \text{(4.5)}
\]

where the operator

\[
\arg \min_{\beta} f(\beta)
\]

stands for the set of \( \beta \) values that minimizes the function \( f(\beta) \), and \( w_i > 0 \) is the weight of the \( i \)th pair of sample data, and \( W_{ii} = w_i \) is the diagonal matrix formed by the weights.

When the column vectors of \( X \) in Equation (4.4) are linearly independent, the coefficients \( \beta \) can be uniquely determined by solving the normal equation:

\[
(X'WX) \cdot \hat{\beta} = X'W \cdot y \quad \text{or} \quad \hat{\beta} = (X'WX)^{-1}X'W \cdot y,
\quad \text{(4.6)}
\]

where, based on a Generalized Gauss-Markov Theorem proposed by Aitken [64], \( \hat{\beta} \) is the best linear unbiased estimator (BLUE) of \( \beta \), if each weight is equal to the reciprocal of the variance of the errors in the data, or:

\[
W_{ii} = w_i = \frac{1}{\sigma_i^2},
\quad \text{(4.7)}
\]

where \( \sigma_i \) is the standard deviation of the errors in the \( i \)th pair of sample data.
Now, we employ Equations (4.5), (4.6), and (4.7) to solve for the $4 \times 4$ calibration transformation matrix $P_U T$ in Equation (3.9). To start, we first converted Equation (3.9) to the same standard form as that of Equation (4.3) by taking transposes on both sides of the equation:

$$P_U T \cdot U X = P_T T_H T \cdot H X \Rightarrow X \cdot T = Y,$$

(4.8)

where

$$X = (U X)', \quad T = (P U T)', \quad Y = (P_T T_H T \cdot H X)'.$$

(4.9)

Both $X$ and $Y$ in Equation (4.9) are $m \times 4$ matrices with $m$ being the number of N-fiducials collected for calibration (defined in Equation (3.11)), and $T$ is a $4 \times 4$ matrix. Note that $T$ and $Y$ can be expressed by their respective column vectors as:

$$T = [\beta_1 \beta_2 \beta_3 \beta_4], \quad Y = [y_1 \ y_2 \ y_3 \ y_4].$$

(4.10)

Hence, we could rewrite Equation (4.8) into four individual equations:

$$X \cdot \beta_k = y_k, \quad (k = 1, 2, 3, 4).$$

(4.11)

Further, because the column vectors of $X$ are homogeneous $x$, $y$ and $z$ Cartesian coordinates of the N-fiducials in the US image frame, they are linearly independent of each other. We could then uniquely find $\beta_k$ using Equation (4.6) as:

$$(X'WX) \cdot \hat{\beta}_k = X'W \cdot y_k \quad \text{or} \quad \hat{\beta}_k = (X'WX)^{-1}X'W \cdot y_k, \quad (k = 1, 2, 3, 4).$$

(4.12)

The most important question yet to answer is how to find the proper weights to make $\hat{\beta}_k$ BLUE in Equation (4.12). According to Equation (4.7), we would need the variance of the errors in the data which we do not know directly; however, it is possible to estimate it using the elevation beamwidth associated with the image data.

Statistically, if a random variable $x$ has a normal distribution of $N(\mu, \sigma)$, the
probability $P$ that covers 95% of the data could be expressed as:

$$P[-2\sigma < x - \mu < 2\sigma] = 0.95.$$  \hfill(4.13)

Hence, the range $R$ can be approximated as [34]:

$$R \approx 4\sigma \quad \text{or} \quad \sigma \approx \frac{R}{4}.$$  \hfill(4.14)

As discussed earlier, the effective elevation beamwidth is essentially the maximum possible range of the localization error caused by the US slice thickness. Hence, using Equation (4.14) and assuming that the localization errors follow the \textit{Gaussian} distribution, the standard deviation of the errors in each image data can be approximated as:

$$\sigma_i^2 \approx \left(\frac{E_i}{4}\right)^2 = \frac{E_i^2}{16},$$  \hfill(4.15)

where $E_i$ is the elevation beamwidth associated with the $i$th image data. Thus, the weights in Equation (4.7) can be rewritten as:

$$W_{ii} = w_i = \frac{1}{\sigma_i^2} \approx \frac{16}{E_i^2}.$$  \hfill(4.16)

As a result, $W_{ii}$ in Equation (4.16) yields $\hat{\beta}_k$, which is BLUE for Equation (4.12).

If, however, the errors in the data are not \textit{Gaussian} distributed, we can still estimate their standard deviation using \textit{Chebyshev’s} inequality theory [34]:

$$P(|x - \mu| < q\sigma) \geq 1 - \frac{1}{q^2}.$$  \hfill(4.17)

For example, when $q = 3$ (or 89% of the data), $R$ covers approximately $6\sigma$, or $\sigma \approx \frac{R}{6}$.

In our experiments, we assumed a \textit{Gaussian} distribution of the errors in the data and employed the weights defined in Equation (4.16).
4.2.5 Calibration Accuracy Evaluation

Using the same approach as discussed in details in Chapter 3, Section 3.2.6, we evaluated the calibration accuracy automatically by PRE (Equation (3.12)). We first scanned the Double-N calibration phantom and reconstructed the cross-sections of the N-Wires into the 3D world coordinate system using the calibration parameters, and then compared them, respectively, to their known physical locations (the gold standard) to compute the mean residual PRE. Since the same phantom geometry was utilized for both calibration and accuracy assessment, to avoid a biased assessment we separated the data used for PRE calculation from those used for calibration.

4.2.6 Experimental Setup

Extensive experiments were conducted on a 10,000-image dataset to thoroughly investigate the effectiveness, significance and performance of USB-FW in improving the calibration accuracy. We have performed two types of experiments: 54 exhaustive calibration trials and 20 real-time calibration trials. For strict comparison, each calibration trial was repeated twice: one with USB-FW and one without USB-FW, given the same sets of input data.

(a) 54 exhaustive calibration trials

This type of experiments was designed to thoroughly evaluate the effectiveness and significance of USB-FW on calibration accuracy. The primary goal of the set up was to reduce any possible negative impact on the calibration accuracy from sources other than the US slice thickness. This would include the errors in position tracking, corrupted input images from image acquisition, random errors in saving the data to medium. Hence, a significant number of input images were used in each calibration
The images were acquired from a wide range of angles and depths that the US probe could be physically placed within the Double-N calibration phantom. A total of 500 images were randomly selected for each calibration trial. In N-wire-based calibration technology, each set of N-wire structure produces one data point in the US image. For our Double-N calibration phantom in which two sets of N-wires were visible in one image, this yielded a total of 1,000 (500 × 2) data points per calibration trial to calculate the calibration parameters (N.B. this was almost 8 times more than what has been typically used (around 120 data points) in N-wire based calibration techniques [16, 36, 46, 62, 83]).

For validation, 1,500 images (different from those used for calibration) were randomly selected to evaluate the calibration accuracy. For our specific phantom design, this would yield a total of 3,000 point targets to compute 3,000 PREs per calibration trial. To alleviate potential outliers in the error calculation, we sorted these resultant 3,000 PREs in an ascending order and took only the top (smallest) 95% (or 2850) of them to calculate the mean and the standard deviation of the PREs for this calibration trial. Note to be fair and consistent in comparison, this outlier removal was applied to both experiments with and without USB-FW.

(b) 20 real-time calibration trials

These tests were used to evaluate the performance of USB-FW in a time-constraint calibration task where, typically, only a limited number of input images sufficient for the calibration to converge to a stable outcome were used. This was mainly to access how well USB-FW would perform in a clinical application for intraoperative use.

A stack of 120 images were acquired in real time for each calibration trial. To
converge, our calibration system always required fewer than 60 images with 2 data points per image, or 120 data points in total. (N.B. this matches the minimum reported number of data for high-accuracy calibration with N-wires [62, 46, 83, 36, 16].

For validation, another stack of 80 images (different from those used for calibration) were acquired in real time to evaluate the calibration accuracy.

4.3 Results

4.3.1 Quantitative US Elevation Beamwidth Profile

Using the slice-thickness calibrator, we measured the elevation beam width at 34 different axial depths for our broadband linear-array transducer operating at a central frequency of 12.5 MHz (Figure 4.3(a)), to cover imaging depths from 4.4 mm to 31 mm. The results were then linearly interpolated to produce a continuous elevation beamwidth profile across all axial depths within the effective imaging area (Figure 4.3(b)). To compare the elevation beamwidth profiles for different sound frequencies, we also measured the slice thickness at a central frequency of 16 MHz (Figure 4.3(a)) (N.B. the beamwidth profile at the 16 MHz frequency was not used for calibration, because our targeted clinical applications for orthopaedic surgeries only employed the 12.5 MHz central frequency).

We have made several key observations during the experiments:

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

- 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 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 [28], [73] and [69] and can be explained by the fact that the focusing of the US beam is performed mechanically [33]).
4.3.2 Results of Exhaustive Experiments

Table 4.1 summarizes our findings in the experiments. First, before USB-FW, the mean of PRE for the 54 exhaustive calibration trials was 0.78 mm, which is consistent with the accuracy reported for related work with N-wire phantoms [16, 36, 46, 62, 83].

Table 4.1: Reduction in 3D Point Reconstruction Errors (PRE) by USB-FW.

<table>
<thead>
<tr>
<th>Experiment Type</th>
<th>3D Point Reconstruction Error (mm)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Exhaustive</strong></td>
<td>mean</td>
<td>max</td>
<td>min</td>
<td>std</td>
<td>mean</td>
<td>max</td>
<td>min</td>
<td>std</td>
<td></td>
</tr>
<tr>
<td>before USB-FW</td>
<td></td>
<td>0.78</td>
<td>0.89</td>
<td>0.68</td>
<td>0.09</td>
<td>0.96</td>
<td>2.28</td>
<td>0.59</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>after USB-FW</td>
<td></td>
<td>0.72</td>
<td>0.62</td>
<td>0.66</td>
<td>0.08</td>
<td>0.75</td>
<td>0.74</td>
<td>0.57</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Reduction</td>
<td></td>
<td>7.3%</td>
<td>30.6%</td>
<td>2.2%</td>
<td>11.1%</td>
<td>21.9%</td>
<td>64.8%</td>
<td>2.7%</td>
<td>54.5%</td>
<td></td>
</tr>
<tr>
<td>Num. Trials</td>
<td></td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td></td>
<td>16 mins/trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13 secs/trial</td>
</tr>
</tbody>
</table>

Notes: (a) The maximum PRE reduction case in the experiments. (b) The minimum PRE reduction case in the experiments. (c) AMD Athlon X2 2.6GHz 2GB RAM Windows Vista 32bit.

Our key observations in the exhaustive tests include the following:

- The PRE was reduced for each of the 54 exhaustive calibration trials after USB-FW.

- As Table 4.1 shows, the mean of the PRE for the exhaustive experiments was reduced from 0.78 mm to 0.72 mm or by 7.3% after USB-FW.

- The standard deviation of the PREs for the exhaustive tests was reduced on average from 0.09 mm to 0.08 mm or by 11% after USB-FW.
• In the maximum-reduction case, the PRE dropped from 0.89 mm to 0.62 mm or by 30.6%, and the PRE dropped from 0.68 mm to 0.66 mm or by 2.2% as the minimum reduction.

• An exhaustive test took an average of 16 minutes to finish (on a Dell Optiplex 740 workstation with AMD Athlon 64 X2 2.6GHz and 2GB memory running Windows Vista 32-bit). This long processing time, however, was expected considering the significant number of data to be analyzed in this type of experiment.

Finally, a paired, two-tailed $t$-test was performed ($p = 0.001$), providing evidence that the reduction in the mean PRE after incorporating the beamwidth was statistically significant.

### 4.3.3 Results of Real-time Experiments

Key observations in the real-time experiments include the following:

• Like in the exhaustive tests, PRE was reduced for each of the 20 real-time calibration trials after USB-FW.

• With USB-FW, the mean PRE for real-time experiments was reduced from 0.96 mm to 0.75 mm or by 21.9% (Table 4.1).

• In the maximum-reduction case, the PRE was reduced even more significantly from 2.28 mm to 0.74 mm or by 64.8%. For the minimum-reduction case, the PRE dropped from 0.59 mm to 0.57 mm or by 2.7%.

• The standard deviation in the PREs was also reduced drastically for the real-time tests, from 0.44 mm to 0.20 mm or by 54.5% after USB-FW.
• Speed-wise, each of the 20 experiments converged in an average of 13 seconds, sufficiently fast for many applications in an operating room.

In addition, a paired, two-tailed t-test yielded \( p = 0.016 \) for the real-time experiments, indicating that the reduction in the mean PRE was statistically significant after the use of the beamwidth.

Furthermore, we have run an upper one-tailed F-test to see if the reduction in the variances of PRE was statistically significant. The F-test concluded that variance in PRE using USB-FW was smaller than those without USB-FW at 0.001 significance level \( (F = 4.8399, \text{Critical Value} = 4.4737, \alpha = 0.001) \).

The real-time calibration experiment consumed less than one-fourth the number of data used in an exhaustive experiment (120 vs. 500). Before USB-FW, the averaged reconstruction error was 23.1% higher in the real-time tests (0.96 mm) than in the exhaustive tests (0.78 mm). After USB-FW, however, we found the following (Table 4.1):

• PREs were reduced more significantly in the real-time experiments than in the exhaustive experiments: the averaged PRE reduction in the real-time tests (-0.21 mm, or -21.9%) was more than three times greater than that in the exhaustive experiments (-0.06 mm, or -7.3%). This result can be intuitively explained by the fact that the original PRE before USB-FW was already much smaller in the exhaustive tests \( (0.78 \pm 0.09 \text{ mm}) \) than in the real-time tests \( (0.96 \pm 0.44 \text{ mm}) \). The averaged PRE for the real-time tests was reduced to 0.75 mm, which was almost the same as the result obtained in the exhaustive tests (0.72 mm).
• More importantly, the statistically significant reduction in the standard deviation of PRE suggests that the variance in the errors was decreased by USB-FW. Based on the three-sigma (empirical) rule, assuming a normal distribution of the errors, nearly all (99.73% confidence interval) of PREs would fall within 3 standard deviations of the mean, i.e., ±0.60 mm after USB-FW, a more than 50% reduction from ±1.32 mm without using USB-FW. This result indicates a tightening of the entire error range by USB-FW in the real-time experiments that produces a less variable calibration outcome - a feature that is always desirable in intraoperative and clinical applications.

The findings in (a) and (b) are significant because they suggest that without increasing the number of input data, USB-FW can improve the accuracy of a real-time calibration to the same level as that of an exhaustive calibration (which consumes considerably more data/time and can only be performed offline).

The clinical implication here was that using USB-FW, we would be able to introduce a form of quality control on an intraoperative, time-constrained calibration task, by efficiently improving the calibration accuracy using only a small number of input data.

4.4 Conclusion

In this chapter, we presented a novel US calibration framework that embedded the prior knowledge of US slice-thickness into the calibration process, and have shown a significant improvement in the overall calibration accuracy.

First, not assuming uniform slice thickness, we extracted a complete elevation beamwidth profile using an inclined-plane-based slice-thickness calibrator across all
effective imaging depth. Second, we developed a _US-beamwidth filtered, weighted-least-square framework_ (USB-FW) to compute the calibration parameters based the slice thickness associated with the input image data. Extensive experiments were conducted on a 10,000-image dataset to thoroughly investigate the impact of USB-FW on the calibration accuracy. The results showed that 3D reconstruction errors were significantly reduced in every experiment \( (p < 0.001) \). Real-time testing showed that the proposed method worked effectively with a small number of input images, suggesting great potential for intraoperative use where only a limited number of data may be available. This new framework can enable efficient quality control of calibration accuracy in real-time operating-room use.
Chapter 5

Intraoperative Calibration for Prostate Brachytherapy

The content of this chapter has been published in the Journal of Medical Physics as:
T. K. Chen, T. Heffter, A. Lasso, C. Pinter, P. Abolmaesumi, E. C. Burdette, and
G. Fichtinger, "Automated intraoperative calibration for prostate cancer brachyther-
apy," Med. Phys. 38(11), 6285-6299 (2011). Text has been edited to best suit the
flow of this thesis.

Statement of originality: In this work, the high-level iCAL system design, the
iCAL calibration phantom design (the basic idea and CAD drawing/blueprints in
Solid Edge), the real-time closed-form calibration algorithm, the automated calibra-
tion accuracy evaluation method, and the design of all scientific validation experi-
ments with different ultrasound imaging systems and brachytherapy stepper systems
are the sole innovations and contributions of T. K. Chen. T. Heffter and C. Pinter
contributed to the software development of ultrasound image acquisition and position
tracking with the brachytherapy steppers. A. Lasso was the project manager of the
iCAL team. G. Fichtinger and E. C. Burdette were the principal investigators who funded and supervised the project.

5.1 Introduction

In previous chapters we have gone through the design, development and validation of an automated US calibration framework incorporating the slice-thickness information. In this chapter, we are concerned with a clinical application that would take advantage of such technology. In particular, we have developed a fast, automated, pure-computation based calibration solution (iCAL) for prostate cancer brachytherapy.

The remainder of this chapter is organized as follows. First, we give a brief introduction to the current brachytherapy practice and the existing calibration technologies in use. Then, a high-level overview of the proposed iCAL system is presented, followed by a CAD design and specifications of the calibration phantom. Next, TRUS image acquisition and position tracking are discussed, together with the synchronization requirement (temporal calibration) between these two processes. Further, the automatic segmentation algorithm is explored step by step. We then reveal the details of our closed-form calibration solution, and illustrate iCAL’s real-time accuracy evaluation mechanism. Finally, we present the experimental setup and validation results obtained with multiple, commercially available TRUS scanners and brachytherapy stepper systems, followed by a brief summary in the end.
5.1.1 Clinical Background and Significance

Prostate cancer is the second most frequently diagnosed cancer and the sixth leading cause of cancer death in men [39]. Brachytherapy has emerged as a definitive treatment for early stage prostate cancer. The procedure entails permanent implantation of small radioactive isotope capsules (seeds) into the prostate to kill the cancer with radiation [81].

![Figure 5.1: TRUS-guided prostate-cancer brachytherapy: needles are inserted into the prostate through a template into a patient in the lithotomy position.](image)

Prostate brachytherapy is delivered with real-time transrectal ultrasound (TRUS) image guidance (Figure 5.1). Typically, the probe is translated and rotated by a mechanical stepper in the rectum with its displacement and rotation angle tracked by encoders on the stepper. Individual TRUS images of prostate contours are then compounded into a volume based on which an implant plan can be created and radiation dose calculated. Finally, under real-time, intraoperative TRUS image guidance, the actual implants are delivered transperineally by needles inserted through a template that contains a rectilinear grid of guide holes. Success of this treatment depends on
an accurate plan of radiation dosimetry and a precise delivery of the implant.

The intrinsic accuracy of a brachytherapy system is solely determined by a unique procedure called “calibration”, where a spatial registration between the coordinate systems of the TRUS and the template must be established prior to the implant procedure. Inaccurate system calibration causes faulty needle and radiation source placement, which may directly contribute to dosimetry errors, toxicity and treatment morbidity [51, 52, 79, 81].

5.1.2 Current Brachytherapy Calibration in Practice

In current practice, brachytherapy system calibration is a laborious, three-stage process.

- **Stage 1**: An operator (typically a medical physicist) ascertains whether the TRUS image truthfully represents the size and shape of scanned objects and whether a series of individual images can be correctly stacked in space to reconstruct an accurate TRUS volume. For these purposes, artificial objects (phantoms) are employed with known geometry suspended in tissue-mimicking gel (to match the speed of sound in tissue) [82]. Phantoms are made commercially for these tasks; e.g., the industry-standard Brachytherapy Phantom CIRS 045 manufactured by Supertech, Elkhart, IN. (US PATENT #5196343). The operator scans the phantom, measures the distance, size, shape and volume of the visible 2D and 3D features in the TRUS images, and then compares them to the known geometric specifications provided by the phantom manufacturer. Such measurements are conducted manually using rulers and calipers, either on the display of the ultrasound scanner, or on the printed TRUS images.
• **Stage 2:** The operator calculates the relative spatial transformation between the coordinate frame of the TRUS images and the coordinate frame of the template [30]. In the usual workflow, the operator mounts the template and the TRUS probe on a stand, dips the probe in a water tank, inserts needles through the template into the tank under TRUS imaging, marks the needle tips in the images, and calculates the transformation between the TRUS and template coordinate frames. For three needles a simple mathematical formula is available.

• **Stage 3:** For some TRUS scanners that offer the ability to superimpose a square grid of coordinates on the real-time image, the overlaid grid lines must be aligned with the grids on the template. This is typically done by using eyesight and manually adjusting the scanner’s setup. The user dips needles through the template into a water tank and then turns the knobs on the TRUS scanner until the grid lines appear to coincide with the artifacts created by the needles [56].

There are a number of technical elements in the calibration workflow that can lead to substantial bias and error in the final result:

• The needles may be bent, therefore the segmented tip positions do not truthfully correspond to the physical locations of the template holes, which leads to erroneous template-TRUS registration;

• The needle tip may be inaccurately segmented, especially when beveled implant needles are used;

• The coordinates of the needle holes may be erroneously recorded;
• The depth of the needle may be erroneously measured and recorded;

• The number of needles used may be inappropriate; typically, too few needles are used;

• The distribution of needle positions may be inappropriate, introducing bias if needle tips do not properly surround the location of the prostate;

• The speed of sound in water is different from the speed of sound in human tissue, which can result in significant distance measurement errors in the TRUS image.

Overall, the procedure is laborious, more qualitative than quantitative, and involves a great deal of eyeballing and subjective judgments by the operator.

Furthermore, the calibration is performed only periodically (primarily due to the inefficiency of the procedure), mostly outside the operating room, with the assumption that calibration parameters remain valid over time. In reality, however, calibration parameters may change during storage, transportation and setup of the equipment.

Perhaps most critically, the system calibration errors are difficult to detect during the procedure so the brachytherapist has no assurance whether the system is functioning correctly in the operating room. There is no validation mechanism in the current procedure to verify and ascertain the calibration accuracy in the operating room.

Finally, brachytherapy calibration, with its current practice, is a major recurring cost for care facilities, consuming manpower, time and money. One must book a calibration room, de-commit the TRUS unit from clinical use, transport the equipment, prepare supplies (needles, water tank, etc.), set up the system, collect and process
data, and log, analyze and document the results, dispose all used supplies, pack away
the brachytherapy system, and return the TRUS scanner to the clinic. This workflow
needs to be repeated from time to time.

5.1.3 Our Contributions

In this work, we aim to remove the aforementioned problems by performing the con-
ventional brachytherapy template calibration tasks (Stage 2 and Stage 3) at once.
We have developed a fast, automated, pure-computation based, intraoperative cal-
ibration (iCAL) technology for prostate cancer brachytherapy, intended to be used
in the operating room when the patient is being prepared for surgery. Our method
eradicates the current practice of pre-operatively performed, labor-intensive and sub-
jective calibration processes. The new calibration technology may simultaneously
reduce treatment costs, increase safety and improve on the accuracy of all prostate
cancer brachytherapy systems.

5.2 Materials and Methods

5.2.1 iCAL System Design and Workflow

The design of iCAL entails a new device and an automated, computational method
to calibrate the brachytherapy systems, intended to be performed in the operating
room. The essence of this invention is a mechanical coupling of a precision-made
calibration phantom and a geometric replica of a standard brachytherapy template,
which effectively combines the iCAL phantom with the brachytherapy stepper system
as one member (Figure 5.2). This unibody design shares some similarity with the
design of a phantom developed by Ng et al. for registration of TRUS and cone-beam CT [60].

Figure 5.2: Design of the iCAL system: calibration phantom mounted on a standard brachytherapy stepper.

The fixture is mounted over the TRUS probe on the stepper, using the standard mounting posts and holes provided for the template. The TRUS probe makes contact with a rubber-membrane window on the posterior (bottom) side of the phantom, where the probe can be translated and/or rotated to acquire TRUS images from the interior of the phantom. The details of the phantom design are given in the next section.

Figure 5.3 shows the workflow of iCAL consisting of five consecutive stages.

1. Serving as input, TRUS images are continuously acquired from the iCAL phantom (Step 1A). The motion (translation and rotation) of the TRUS probe is tracked by the brachytherapy stepper (Step 1B).
2. A temporal calibration process is performed to synchronize the individual TRUS image frames with their respective positions (Step 2).

3. The TRUS images of the iCAL phantom are automatically segmented to extract the pixel locations of the phantom geometry (Step 3).

4. The pixel locations of the segmented phantom features, together with their corresponding physical coordinates collected in the phantom space, are fed to a closed-form formula to calculate the calibration parameters (Step 4).

5. Measured by a reconstruction error against a known ground truth, the accuracy of the calibration result is fed back to the control loop to determine whether or not it is satisfactory. The reconstruction accuracy is updated, monitored and displayed in real time. Once the process converges or the reconstruction error reaches an acceptable level, the procedure is terminated and the final calibration outcomes exported (Step 5).

Compared to the conventional, manual brachytherapy calibration (Stage 2, 3), iCAL accomplishes all the required tasks in one automated loop. First, the calibration outcome from iCAL contains the homogenous spatial transformation parameters that...
register the TRUS image plane to the template. This accomplishes Stage 2 of the conventional calibration.

Second, as a byproduct of the calibration results, iCAL overlays the location of the template grid onto the transverse TRUS image, and in real time updates and displays a virtual grid through an interactive graphics interface to the user whenever the probe is being translated and/or rotated. This accomplishes Stage 3 of the conventional calibration.

From a software-architecture point of view, iCAL was designed and developed using a multiple-component-based object-oriented methodology [9]. In addition to the unique calibration phantom design, iCAL encompasses five essential system components: TRUS image acquisition and tracking, temporal calibration, automatic segmentation, closed-form calibration, and real-time reconstruction accuracy feedback. A number of open-source software libraries were employed, including the Visualization Toolkit (VTK), QT and Vision Numerics Library (VNL).

5.2.2 iCAL Phantom Design

The iCAL calibration phantom mechanically combines a CAD-designed N-wire (Z-fiducial) calibration phantom and an exact, exterior replica of the brachytherapy template into one member, as part of the phantom geometry.

For geometric precision and structural integrity, the iCAL phantom was first designed in the Solid Edge CAD software (Figure 5.4(a)). The CAD model was then exported to a Dimension 1200es 3D Printer (Stratasys, Inc., Eden Prairie, MN, USA) to manufacture the parts using a production-grade, high-density, high-strength, and liquid-proof thermoplastic. There are three individual parts designed to complete the...
phantom assembly (Figure 5.4(b)): a container box, an inner N-wire mount, and a sealing cover.

The container box embeds the exterior replica of the template and was measured to be 156 mm × 96 mm × 92 mm in size (Figure 5.4(b)). In our prototype design, to approximate the speed of sound in tissue (1540 m/s), the container box was filled with distilled water heated to a temperature of around 37 °C, at which sound travels at approximately 1570 m/s [33]. In the final product, the box will be sealed with acoustic coupling medium to exactly match the speed of sound in tissue.

The inner N-wire mount is replaceable and fits snugly in the container box, with dimensions of 146 mm × 80 mm × 62 mm (Figure 5.4(c)). It consists of a front and a
back plate connected by two side walls, forming a simple open-ended box. There are holes on both the front and back plate to mount the N-wires. The N-wires are made of nylon line at 0.4 mm in diameter, a size comparable to the TRUS wavelength used for prostate imaging (ranging typically from 0.2 mm to 0.5 mm) which would optimize the image appearance of the wires [17]. The location of the N-wires encompasses the targeted area of the prostate in a clinical setup, thus maximizing the calibration accuracy.

The sealing cover contains a rubber window for TRUS imaging (Figure 5.4(d)). A 0.8 mm thick, natural rubber membrane (manufactured by McMaster-Carr Inc., Robbinsville, NJ, USA) forms the imaging window on the posterior (bottom) side of the phantom. Natural rubber has an acoustic impedance (1.81 Mrayls) that is similar to that of water (1.48 Mrayls), allowing sound transmission in and out of the container with only small attenuation.

Finally, the embedded replica is made from the brachytherapy template model D1-1784RA (Burdette Medical Systems, Inc., Champaign, IL, USA) that measures 81.0 mm × 71.1 mm × 19.1 mm in dimensions and has a matrix of 13 × 13 holes at 5-mm spacing vertically and horizontally.

There are two immediate benefits of this unibody design to mechanically couple the iCAL phantom to the template replica. First, it preempts the sterilization issues of the phantom, because it would be otherwise impractical to attach the phantom to the template which needs to be sterilized in the operating room. Second, the precision-engineering design and prototyping ensure a very high accuracy and precision in localizing the phantom geometry during the calibration process.
5.2.3 TRUS Image Acquisition, Tracking and Temporal Calibration

TRUS Image Acquisition

iCAL is equipped with two common types of TRUS image acquisition interfaces: an analog and a digital data acquisition.

**Analog Data Acquisition.** The TRUS images are transferred from the analog data output (e.g., S-Video or Composite port) of the TRUS scanner to an ImageSource DFG/USB2-LT USB framegrabber (Imaging Source, LLC., Charlotte, NC, USA) installed in the host computer at 30 frames per second (fps). The major benefit of using the analog data output lies in the fact that it is the most common interface available on a standard TRUS machine, therefore provides the best hardware compatibility for iCAL to work with virtually any commercially available TRUS scanners currently in the market. The major downside however, is the relatively lower image quality compared to the digital format. Because the ultrasound machine processes and stores all scan-converted image data digitally in its internal memory, a digital-to-analog (D/A) conversion of the data needs to be performed when outputting the signals to the analog video port, which is then converted back to digital by the USB video-capturing device on the host computer. This double conversion results in a degradation in image quality of the original, digital image [33]. We have tested the analog data acquisition with iCAL on five commercially available TRUS scanners:

- Leopard 2001 (BK-Medical Systems, Inc., Peabody, MA, USA),
- Sonix MDP 4.0 Analog Output (Ultrasonix Medical Corp., Burnaby, BC, Canada),
• Sonix TOUCH Analog Output (Ultrasonix Medical Corp., Burnaby, BC, Canada),

• VLCUS (Carolina medical Systems, Inc., USA),

• Terason 2000 (Teratech Corp., Burlington, MA, USA).

**Digital Data Acquisition.** For better imaging quality, higher data acquisition speed and research purposes, some TRUS scanners on the market now offer a digital interface to acquire images directly from the internal image memory of the ultrasound machine. In iCAL, we have also developed a digital data acquisition based on the OPENSonix platform from Ultrasonix Medical Corp., Burnaby, BC, Canada. We have tested the digital data acquisition with iCAL on Ultrasonix SonixMDP and SonixTOUCH scanners.

**Stepper Position Tracking**

While the TRUS images are being continuously acquired from the iCAL phantom, the motion of the TRUS probe is simultaneously tracked by the brachytherapy stepper. Typically, there are two separate, optical encoders to independently track the motion of the TRUS probe in real time: a translation encoder that reads the displacement of the probe along the Z-axis, and a rotation encoder that reads the rotation of the probe transversely. This position data is then retrieved back to the host computer via a serial-port connection to be associated with the image data. We have tested iCAL with two types of brachytherapy stepper systems that are currently commercially available on the market:

• Target Guide Stepper (Burdette Medical Systems, Inc., Champaign, IL, USA),
• Accuseed DS300 Stepper (Computerized Medical Systems, Inc., Saint Louis, MO, USA).

Temporal Calibration

Since the images and the tracked probe positions are generated by separate hardware (the TRUS scanner and the stepper tracking system), proper synchronization between the two must be established to correctly associate each acquired TRUS image with its corresponding positional data. This process is commonly referred to as the “temporal calibration” [50].

When a TRUS image is acquired and its corresponding stepper position recorded, both data can be time-stamped. However, this time-stamping process itself introduces some delay, because of the difference in processing speed between the ultrasound machine and the stepper system, as well as the necessary time required for data transfer (from the ultrasound machine and the stepper system to the host computer). The goal of a temporal calibration is to determine this delay (“latency”). Temporal calibration is typically conducted by introducing some form of abrupt change in the motion of the ultrasound transducer that would also result in a traceable difference in the ultrasound image [26, 32, 49, 58, 66, 71, 76]. The basic idea is to identify and match this difference in both the positional data and in the ultrasound image, based on which a latency between the tracking and image acquisition is then computed.

For iCAL, we have developed a fast and automated temporal-calibration technique by repeatedly and rapidly pausing and translating the TRUS probe on the stepper every two seconds while imaging the iCAL phantom. Figure 5.5 shows an example of the results of the temporal calibration in iCAL. The abrupt and repeated motion...
caused positional changes in both the TRUS image contents (red line) and the stepper readings (green line), which were automatically detected and registered together to compute the temporal latency.

![Graph](image)

Figure 5.5: An example of the automated temporal calibration in iCAL.

### 5.2.4 Automated Segmentation

Demanding no human interference at all, we developed an algorithm to automatically segment images acquired from the iCAL calibration phantom. This automated segmentation algorithm was discussed in length in Chapter 3, Section 3.2.4.

### 5.2.5 The Closed-Form Calibration Method for iCAL

iCAL employed a similar closed-form least-square method as we discussed in Section 3.2.5 (Equations (3.9), (3.10) and (3.11)), with necessary modifications to accommodate the stepper tracking mechanism that is different than a freehand, optical
tracking system. Figure 5.6 shows the N-wire geometry of the iCAL calibration phantom.

![Image of TRUS image with N-wire geometry]

Figure 5.6: The unique N-wire geometry of the iCAL calibration phantom.

The objective is to determine $P_U$, the transformation that brings a position from the TRUS image frame ($U$) to the TRUS probe frame ($P$) through a series of 3D spatial frame transformations as depicted by Figure 5.7.

To start, we first acquire a set of TRUS images from the iCAL calibration phantom. The intersection point of a wire and the TRUS image plane could be expressed in the TRUS image frame ($U$) and in the phantom frame ($H$) as $U_{X}$ and $H_{X}$, respectively (Figure 5.6):

$$U_{X} = (P_{U})' \cdot P_{S} \cdot P_{H} \cdot H_{X}. \quad (5.1)$$

On the left side of Equation (5.1), $U_{X}$ could be measured as the N-wire positions in the TRUS image frame. On the right side, $P_{S} \cdot P_{H}$, the transformation from the stepper frame to the TRUS probe frame, was known from the stepper’s position tracking of the probe. $P_{H} \cdot H_{X}$, the transformation from the phantom frame to the stepper frame,
was known by the iCAL phantom design that mechanically couples the phantom geometry with an exact replica of the template affixed onto the stepper. $^H X$ is the corresponding physical position of $^U X$ in the phantom frame and can be calculated using the similar-triangle geometry of the N-wires [16]. Finally, $^P_T$ is the unknown calibration parameter for which can be solved using a straightforward implementation of linear least squares.

5.2.6 Real-time Evaluation of Calibration Accuracy

Similar to the freehand calibration system that we discussed in Section 3.2.6, we employed the iCAL calibration phantom to evaluate the calibration accuracy. We first scanned the iCAL phantom and reconstructed the cross-sections of four sets of parallel wires (#1, #3, #4, and #6 in Figure 5.6) into the 3D world coordinate system using the computed calibration parameters. We then compared them, respectively, to
the wires’ known physical locations (the gold standard) to compute a mean residual reconstruction error, defined as the line reconstruction error (LRE):

\[ \| \text{LRE} \| = \| H^X - \frac{H}{S}T \cdot \frac{S}{P}T \cdot \frac{P}{U}T \cdot \frac{U}{X} \|, \]  

(5.2)

where \( \frac{P}{U}T \) is the calibration outcome to evaluate, \( \frac{S}{P}T \) is the transformation from the TRUS probe frame to the stepper frame given by the stepper’s position readings, \( \frac{H}{S}T \) is the transformation from the stepper frame to the iCAL phantom frame that is known by the iCAL phantom geometry mechanically coupled to the stepper system, \( \frac{U}{X} \) is the identified position of the wire in a TRUS image, and \( H^X \) is the corresponding wire location known by the iCAL phantom design (as the ground truth).

Note the LRE is a point-line distance as opposed to the PRE which is a point-point distance (Equation (3.12)), because we choose to reconstruct the two parallel wires in the N-wire geometry instead of the slant wire in the middle (from which the wire-image intersection can be deducted using similar triangles, as explained in Section 3.2.5 and Equation (3.4)). This selection of validation criteria was justified by the following two reasons. First, iCAL phantom was designed to have the four outlier (parallel) wires encompass the typical location and size of a prostate gland in a TRUS image during the brachytherapy procedure, in such that the value of LRE would give a good estimate of needle-insertion accuracy on average across the effective treatment region (which covers the entire prostate). Second, what LRE does not capture is the accuracy in the depth axis (Z-axis or along the needle insertion direction) which is not the target/concern of a typical brachytherapy calibration because the needle insertion depth is only manually controlled by the physician and independent of a brachytherapy system. The accuracy along the needle insertion path only becomes relevant when the needle (tip) is also tracked and calibrated as part of the
brachytherapy system, which is not the case in the present brachytherapy practice.

Other than this difference, LRE measurement shares some similar features with PRE. First, LRE is an absolute Euclidean distance between a reconstructed point and the respective wire in space, so it remains invariant to frame transformations, and has units of millimeters.

Second, iCAL measures LRE in real time by automatically extracting the wire positions and reconstructing them in the physical phantom space using the computed calibration parameters.

Further, because the same phantom geometry is utilized for both calibration and accuracy evaluation, we separated the data used for LRE calculation from those used for calibration to avoid a systematic bias in the error evaluation toward the calibration results.

Finally, automatic error retrieval enabled us to quickly test iCAL through 50 independent calibration trials with the error computed in real time, which would not be possible for a conventional, manual brachytherapy calibration procedure.

### 5.2.7 Real-time Overlay of the Template Grid on the TRUS Images

Once the calibration parameters are computed, iCAL automatically registers (overlays) the location of the template grid onto the transverse TRUS image, and displays it to the users via an interactive 3D graphical user interface (Figure 5.8). The user can rotate and enlarge the 3D scene to visually examine the spatial relationship between the TRUS image plane and the template grid. Note the matrix of green dots shown in Figure 5.8 are the needle guiding holes on the front surface of the template.
Figure 5.8: The interactive graphics interface of iCAL showing the 3D overlay of the template grid to the TRUS image in real time.

iCAL also updates in real time the position change of the TRUS image and the template when the probe is being translated and/or rotated and when the template is being displaced by the user, respectively.

On newer TRUS scanners, such as the Sonix MDP or Sonix TOUCH (Ultrasonics, Inc., Burnaby, BC, Canada), the calibration parameters can also be set by iCAL through a manufacturer-provided application programming interface (API) to directly update the superimposed template grid on the display of the TRUS machine.
5.2.8 Experimental Setup

To investigate the accuracy, precision, robustness and speed of iCAL, we performed four types of independent tests:

- Experiments with 50 real-time independent calibration trials,
- Experiments with multiple different TRUS scanners,
- Experiments with multiple different brachytherapy stepper systems,
- Experiments with needle insertion to validate template-TRUS calibration.

Experiments with 50 Real-Time Independent Trials

This type of experiment was designed to validate the accuracy, reliability and speed of iCAL. In the experiment, the TRUS scanner used was a Sonix MDP 4.0 (Ultrasonics, Inc., Burnaby, BC, Canada) that operated at a central frequency of 6 MHz with an imaging depth of 7 cm. The TRUS image data was acquired digitally using the Ulterius SDK provided by the manufacturer. The stepper system used was a Target Guide Stepper (Burdette Medical Systems, Inc., Champaign, IL, USA).

A total of 50 independent, real-time calibration trials were individually performed by iCAL, with the transducer inserted and attached to the stepper at all time. In each calibration experiment:

- A total of 300 live TRUS images were acquired from the iCAL phantom to compute the calibration parameters. For N-wire-based calibration technologies, a minimum reported number of 120 data points are necessary for high-accuracy calibration [62, 46, 83, 36, 16]. In the iCAL phantom design (Figure 5.6) where
two sets of N-wires are visible in a single TRUS image, the 300 images used for calibration would yield 600 data points, which is more than sufficient for the process to converge.

- Another 100 live TRUS images acquired from the iCAL phantom were used to compute the LRE (Equation (5.2)) in order to evaluate the calibration accuracy. On each validation image, the positions of the wires #1, #3, #4 and #6 (Figure 5.6) were automatically extracted and reconstructed in the 3D space using the calibration parameters. This yields 400 LRE measurements per experiment, from which a mean and standard deviation of the errors are calculated.

- This setup purposely prevents the same data from being used in both calibration and accuracy evaluation, which would create a biased validation.

Finally, the LREs from all 50 independent experiments were statistically analyzed.

**Experiments with Multiple Different TRUS Scanners**

These experiments were designed to evaluate the robustness and compatibility of iCAL working with ultrasound machines of different data interfaces, types/sizes, ages, and manufacturers:

- Interfaces: analog (S-video) and digital (proprietary API) data acquisition;

- Type/Size: full-size, portable and laptop size.

- Ages: some old and latest generations of TRUS machines were tested.

- Manufacturer: TRUS scanners are from four different manufacturers.
Table 5.1: Experimental setup with multiple TRUS scanners.*

<table>
<thead>
<tr>
<th>TRUS Scanner</th>
<th>Manufacturer</th>
<th>Data Interface</th>
<th>Type</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonix MDP 4.0</td>
<td>Ultrasonix Medical Corp., Burnaby, BC, Canada</td>
<td>S-video</td>
<td>Full</td>
<td>2000s</td>
</tr>
<tr>
<td>Sonix TOUCH (analog)</td>
<td>Ultrasonix Medical Corp., Burnaby, BC, Canada</td>
<td>S-video</td>
<td>Full</td>
<td>2010s</td>
</tr>
<tr>
<td>Sonix TOUCH (digital)</td>
<td>Ultrasonix Medical Corp., Burnaby, BC, Canada</td>
<td>Ulterius</td>
<td>Full</td>
<td>2010s</td>
</tr>
<tr>
<td>VLCUS</td>
<td>Carolina Medical Systems, Inc., NC, USA</td>
<td>S-video</td>
<td>Portable</td>
<td>1990s</td>
</tr>
</tbody>
</table>

*Stepper used: Target Guide Stepper (Burdette Medical Systems).

Table 5.1 lists the detailed setup for the experiments. In each experiment, 300 live TRUS images were acquired from the iCAL phantom to compute the calibration parameters, and then another 100 live TRUS images to compute the LRE. To limit the testing variable to the TRUS machines only, the same Target Guide Stepper (Burdette Medical Systems, Inc., Champaign, IL, USA) was used through all the experiments.

**Experiments with Multiple Different Brachytherapy Stepper Systems**

These experiments were designed to examine the robustness and compatibility of iCAL working with different brachytherapy stepper systems, offering varying position-tracking quality. The experiments included four different Target Guide Steppers and
an Accuseed DS300 Stepper. In each experiment, 300 live TRUS images were acquired
from the iCAL phantom to compute the calibration parameters, and then another
100 live TRUS images to compute the LRE. Table 5.2 gives the details of the tested
stepper systems. To limit the testing variable to the stepper systems only, we used
the digital data acquisition from Sonix TOUCH (Ultrasonix Medical Corp., Burnaby,
BC, Canada) for all the experiments.

Table 5.2: Experimental setup with multiple brachytherapy stepper systems.

<table>
<thead>
<tr>
<th>Brachytherapy Stepper</th>
<th>Manufacturer</th>
<th>Number of Units Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Guide Stepper</td>
<td>Burdette Medical Systems, Inc., Champaign, IL, USA</td>
<td>4</td>
</tr>
<tr>
<td>Accuseed DS300 Stepper</td>
<td>Computerized Medical Systems, Inc., Saint Louis, MO, USA</td>
<td>1</td>
</tr>
</tbody>
</table>

Experiments with Needle Insertion to Validate Template-TRUS Calibration

Finally, we validated the template-to-TRUS calibration/registration outcome by mea-
suring a target registration error (TRE) with the water tank method [30, 56]. Seven
brachytherapy needles were inserted to the same depth through the template holes C3,
C5, D5, E3, E5, b4, and e4 and scanned by TRUS in a water tank. The brachytherapy
needles used in the tests are 18-gauge Mick TP Prostate Seeding Needles (Mick Radio-
Nuclear Instruments, Inc., Bronx, NY, USA), having a 1.270 ± 0.013 mm outer diam-
eter. The TRUS image has a size of 640 × 480 pixels and a resolution of 0.2 mm/pixel
(or 5 pixels per millimeter). The position of each needle tip in the TRUS image was
manually segmented by an experienced human operator and then compared to the
computed location by iCAL to obtain the respective TRE value.

5.3 Results

5.3.1 Results of Automated Segmentation

We tested the automated segmentation algorithm with over 10,000 TRUS images acquired from the iCAL phantom. These images were taken at various displacements and rotation angles to cover every possible view of the N-wires. We then compared the segmentation outcomes with the ground-truth results identified by an experienced human operator. In all the images tested, there was no failure in identifying the N-wires and a careful inspection could not identify visible segmentation errors. Figure 5.9 shows an example of segmentation.

Figure 5.9: Automatically segmented N-Wires from the TRUS image of the iCAL phantom.
5.3.2 Results of Experiments with 50 Real-Time Independent Trials

Table 5.3 shows the validation results for the 50 independent, real-time calibration trials performed by iCAL. LRE was computed as the “line-reconstruction error” or the “point-line distance”, as defined in Equation (5.2). The TRUS image has a size of $640 \times 480$ pixels and a resolution of 0.2 mm/pixel (or 5 pixels per millimeter).

Table 5.3: Validation results of iCAL for 50 independent, real-time calibration trials.

<table>
<thead>
<tr>
<th>Number of Trials: (^a)</th>
<th>Line Reconstruction Error (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 (independent)</td>
<td>Mean ((\mu)): 0.57 mm</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation ((\sigma)) : 0.13 mm</td>
</tr>
<tr>
<td></td>
<td>Minimum (best case) : 0.26 mm</td>
</tr>
<tr>
<td></td>
<td>Maximum (worst case) : 0.91 mm</td>
</tr>
<tr>
<td></td>
<td>Runtime (^b) : 20 secs/trial</td>
</tr>
</tbody>
</table>

\(^a\)With a Sonix MDP scanner and a Target Guide Stepper.
\(^b\)Intel Q6600 2.4GHz 4GB-RAM Windows Server 2008 64-bit.

Key findings include the following. First, all 50 calibration trials reached a sub-millimeter accuracy: the average LRE for all trials was 0.57 mm and the maximum error (the worst case scenario) was 0.91 mm. The clinical translation of this result into a brachytherapy procedure is that assuming there is no needle bending, the accuracy of needle insertion and seed placement based on the template-TRUS registration provided by iCAL would be 0.6 mm on average. In addition, this result is also consistent with the reconstruction accuracy reported for related ultrasound calibration literature using N-wires [16, 36, 46, 62, 83].

Second, because the calculation of LRE is based on the ground-truth position of
the wires #1, #3, #4 and #6 (Figure 5.6) which, by the design of iCAL phantom, encompass the targeted area of the prostate in a clinical setup, the scope of the LRE provides a sound estimate of the accuracy in localizing anatomical targets in the prostate during a brachytherapy procedure.

Further, the standard deviation of the LRE was 0.13 mm, which suggests that iCAL also has an excellent precision in producing a consistent, and repeatable calibration. The very low variability in the calibration outcome is desirable for use in the operating room.

Finally, each of the 50 calibration trials converged in an average of 20 seconds (on an Intel Core 2 Duo Q6600 workstation at 2.4GHz with 4 GB memory running Windows Server 2008R2 64-bit), sufficiently fast for use in the operating room.

### 5.3.3 Results of Experiments with Different TRUS Scanners

Table 5.4 shows the validation results of iCAL tested with different TRUS scanners. All experiments were conducted using the same Target Guide Stepper (Burdette Medical Systems, Inc., Champaign, IL, USA).

Our key observations and findings include the following. First, iCAL was able to consistently achieve a sub-millimeter, high calibration accuracy and precision across all the tested TRUS imaging platforms: the average LRE is 0.37 mm with a standard deviation of 0.25 mm.

There was no difference in iCAL’s accuracy level between the digital data acquisition (Sonix TOUCH Ulterius API) and the analog data acquisition from SVideon (the rest of the tested scanners), even though the digital platform offers better image quality (less noise) and finer per-pixel resolution than the analog units. In our tests, the
Table 5.4: Validation results of iCAL tested with different TRUS scanners.*

<table>
<thead>
<tr>
<th>TRUS Scanner</th>
<th>Frequency (MHz)</th>
<th>Depth (cm)</th>
<th>Data Type</th>
<th>Type</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leopard 2001</td>
<td>6.5</td>
<td>9.0</td>
<td>S-video</td>
<td>Full</td>
<td>1990s</td>
</tr>
<tr>
<td>Sonix MDP 4.0</td>
<td>6.0</td>
<td>7.0</td>
<td>S-video</td>
<td>Full</td>
<td>2000s</td>
</tr>
<tr>
<td>Sonix TOUCH</td>
<td>6.0</td>
<td>7.0</td>
<td>S-video</td>
<td>Full</td>
<td>2010s</td>
</tr>
<tr>
<td>(analog)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sonix TOUCH</td>
<td>6.0</td>
<td>7.0</td>
<td>Ulterius</td>
<td>Full</td>
<td>2010s</td>
</tr>
<tr>
<td>(digital)</td>
<td></td>
<td></td>
<td>Digital</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terason 2000</td>
<td>Norm</td>
<td>8.0</td>
<td>S-video</td>
<td>Laptop</td>
<td>2000s</td>
</tr>
<tr>
<td>VLCUS</td>
<td>Norm</td>
<td>7.0</td>
<td>S-video</td>
<td>Portable</td>
<td>1990s</td>
</tr>
<tr>
<td><strong>average</strong></td>
<td><strong>0.37</strong></td>
<td><strong>0.25</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*All tests used the same Target Guide Stepper (Burdette Medical Systems).

digital TRUS image has 0.2 mm/pixel resolution (or 5 pixels per millimeter) while the analog image only has 0.26 mm/pixel resolution (or 3.8 pixels per millimeter). This suggests that iCAL does not demand high image quality and resolution to perform accurately, and would be compatible with the majority of TRUS machines currently on the market which are equipped with a standard analog data output.

There was also no difference in iCAL’s accuracy level between the latest TRUS machines (manufactured after 2000) and some of the older ones (manufactured in the 1990s). The newer TRUS scanners typically offer more polished hardware design and better signal transmission, retrieving and processing quality, which in turn results in overall better TRUS imaging quality than the older technologies. This result
confirms the robustness of iCAL in dealing with varying TRUS hardware and imaging conditions.

Finally, iCAL achieved the same level of accuracy in different types of TRUS machines including standard full-size, portable and even laptop-size TRUS scanners. This suggests that iCAL does not require high-processing power from a typical full-scale TRUS system to function properly, which provides more flexibility and mobility in an intraoperative brachytherapy situation to work with small, portable TRUS devices if desired.

5.3.4 Results of Experiments with Different Brachytherapy Steppers

Table 5.5 shows the validation results of iCAL tested with different brachytherapy stepper systems. In all experiments, the TRUS images were acquired from the Sonix TOUCH (Ultrasonix Medical Corp., Burnaby, BC, Canada) via the analog SVideo data output using the USB framegrabber.

Key observations and findings include the following. iCAL was able to consistently achieve a sub-millimeter, high calibration accuracy and precision across all the tested brachytherapy stepper systems: LREs of all tests were below 0.5 mm with an average of 0.29 mm and a standard deviation of 0.16 mm. This suggests that iCAL is robust in working with different stepper systems of varying mechanical condition and/or tracking quality.

For the four Target Guide Steppers (Burdette Medical Systems, Inc., Champaign, IL, USA) we have tested, the mean of LRE was 0.32 mm with a standard deviation of 0.18 mm. This was consistent with the results of the previous tests with multiple
Table 5.5: Validation results of iCAL tested with different brachytherapy steppers.*

<table>
<thead>
<tr>
<th>Brachytherapy Stepper</th>
<th>LRE: mm</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std</td>
</tr>
<tr>
<td>Target Guide #1</td>
<td>0.25</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target Guide #2</td>
<td>0.34</td>
<td>0.21</td>
</tr>
<tr>
<td>Target Guide #3</td>
<td>0.33</td>
<td>0.25</td>
</tr>
<tr>
<td>Target Guide #4</td>
<td>0.36</td>
<td>0.10</td>
</tr>
<tr>
<td>Target Guide Stepper</td>
<td>0.32</td>
<td>0.18</td>
</tr>
<tr>
<td>(average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuseed DS 300</td>
<td>0.16</td>
<td>0.10</td>
</tr>
</tbody>
</table>

*TRUS images acquired by Sonix TOUCH Analog (Ultrasonix Corp.).

TRUS scanners using a Target-Guide Stepper (Table 5.4).

More importantly, we found that Accuseed DS 300 achieved a significantly higher calibration accuracy and precision than Target Guide steppers: the LRE mean of the Accuseed was one half that of the Target Guides (0.16 mm versus 0.32 mm), with a much smaller standard deviation as well (0.10 mm versus 0.18 mm). This result is due to the fact that the Accuseed DS 300 stepper provides higher position-tracking accuracy than the Target Guide stepper and is also mechanically more stable and precise in design.

Finally, there are two significant clinical implications of these findings. First, any improvement in the stepper tracking accuracy and precision, and the mechanical stability may significantly improve the brachytherapy calibration. Second, iCAL is capable of providing a means of real-time quality assurance of the brachytherapy stepper systems in the operating room, by monitoring and reporting any unexpected change in the calibration accuracy and precision in an intraoperative brachytherapy
procedure.

5.3.5 Results of Needle Insertion to Validate Template-TRUS Calibration

Table 5.6 shows the TRE results of the template-TRUS calibration accuracy of iCAL. The brachytherapy needles used in the tests are 18-gauge Mick TP Prostate Seeding Needles (Mick Radio-Nuclear Instruments, Inc., Bronx, NY, USA), having a 1.270 ± 0.013 mm outer diameter. The TRUS image has a size of 640 × 480 pixels and a resolution of 0.2 mm/pixel (or 5 pixels per millimeter).

Table 5.6: TRE results of needle insertion to validate template-TRUS calibration accuracy.

<table>
<thead>
<tr>
<th>Template Grid</th>
<th>C3</th>
<th>C5</th>
<th>D5</th>
<th>E3</th>
<th>E5</th>
<th>b4</th>
<th>e4</th>
<th>mean</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRE (mm)</td>
<td>0.31</td>
<td>0.25</td>
<td>0.91</td>
<td>0.56</td>
<td>0.88</td>
<td>0.35</td>
<td>0.69</td>
<td>0.56</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Our key observations and findings include the following. The TREs were all below 1 mm in the seven needle-insertion experiments, with an average of 0.56 mm and a standard deviation of 0.27 mm. These results suggest an excellent template-TRUS calibration accuracy and precision of iCAL.

Further, the TRE measurements are consistent with the sub-millimeter LRE results reported in Tables 5.3, 5.4 and 5.5.

Finally, the tested locations where the needles were inserted (C3, C5, D5, E3, E5, b4, and e4) encompassed the targeted area of the prostate in a clinical setup. Therefore, the TRE measured in this setup provides a good approximation of the template-TRUS calibration accuracy using iCAL in the operating room.
Figure 5.10: Needle insertion to validate template-TRUS calibration accuracy (TRE).

Figure 5.10 shows a visual confirmation of the results: the template grid (in green) was registered to the TRUS images by iCAL and displayed to the user in real time while the needles were being inserted through the grid holes in the water tank. As can be clearly observed, the needle artifacts in the TRUS image, which were the corresponding positions of the needle tip inserted through grid holes C5, D5, E5, b4, e4, C3, and E3, all perfectly coincided with the template grid positions computed by iCAL, showing excellent TRE.

5.3.6 Results of 3D Overlay of Template to TRUS

Figure 5.11 displays the 3D overlay of the template grids on the real-time TRUS images, after the calibration parameters were computed by iCAL. The green grid matrix in the view is the front, painted face of the template and the rear gray-scale
image is the real-time TRUS images (in this particular case showing the N-wires of the iCAL phantom). The user can interact with the 3D display to translate, rotate and zoom in/out the 3D overlay (four arbitrary viewing angles and positions are shown in Figure 5.11).

Figure 5.11: Results of the final, real-time 3D overlay of the template to TRUS.

Because the positions of both the TRUS probe and the template are tracked in real-time by the stepper (and monitored by iCAL), iCAL will accurately update and display any position change of the TRUS image and the template grids while the user translates or rotates the TRUS transducer during the scan, or displaces the template
back or forth.

5.4 Conclusion

In this chapter, we presented a new device and a fast, automatic method to calibrate the brachytherapy system in the operating room, with instant error feedback. The device was CAD-designed and precision-machined, which mechanically couples a calibration phantom with an exact replica of the standard brachytherapy template. From real-time TRUS images acquired from the calibration device and processed by the calibration system, the coordinate transformation between the brachytherapy template and the TRUS images was computed automatically. The system instantly generated a report of the target reconstruction accuracy based on the current calibration outcome.

Four types of validation tests were conducted. First, 50 independent, real-time calibration trials yielded an average of $0.57 \pm 0.13$ mm line reconstruction error (LRE) relative to ground truth. Second, the averaged LRE was $0.37 \pm 0.25$ mm relative to ground truth in tests with six different commercial TRUS scanners operating at similar imaging settings. Furthermore, testing with five different commercial stepper systems yielded an average of $0.29 \pm 0.16$ mm LRE relative to ground truth. Finally, the system achieved an average of $0.56 \pm 0.27$ mm target registration error (TRE) relative to ground truth in needle insertion tests through the template in a water tank. In all, the proposed automatic, intraoperative calibration system for prostate cancer brachytherapy has achieved high accuracy, precision and robustness.
Chapter 6

Conclusions and Future Work

6.1 Summary of Novelty and Contributions

In this thesis, we have presented the research and development of three major technological advancements in the tracked US calibration process:

1. An automated, freehand US calibration system with real-time accuracy feedback to determine the calibration quality in the operating room,

2. Characterization and incorporation of US slice-thickness profiles in a beamwidth-weighted US calibration framework that accounts for localization errors and uncertainties associated with slice thickness data,

3. The first automated, intraoperative TRUS calibration solution for prostate cancer brachytherapy. The solution has high accuracy, precision and robustness and was thoroughly tested with multiple commercially available TRUS scanners and brachytherapy systems.
6.1.1 Automated, Freehand Calibration With Accuracy Feedback

We have presented a real-time, freehand ultrasound calibration system with automatic accuracy control. The system includes several features that are crucial for intraoperative use, namely:

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 automated segmentation and calibration algorithm that requires no human interaction beyond acquiring the ultrasound images; and

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

Extensive validations were conducted on a set of 10,000 images from 50 independent calibration trials, enabling a thorough investigation of the calibration system’s accuracy, robustness, and performance. The average calibration accuracy (measured in terms of 3D reconstruction error against a known ground truth) of all 50 trials was 0.66 mm. In addition, the calibration errors converged to the sub-millimeter range in 98% of trials in an average of 12.5 seconds. Overall, the calibration system was able to consistently, efficiently and robustly achieve high calibration accuracy during real-time performance.
6.1.2 USB-FW: A Beamwidth-Weighted Calibration Framework

We have presented a novel method that incorporates the US elevation beamwidth profile into a filtered, weighted-least-square framework, or USB-FW, to improve the reconstruction accuracy of a real-time, freehand US calibration system. A complete US slice-thickness profile was manually extracted by the slice-thickness calibrator across all effective imaging depths.

Extensive experiments were conducted on a 10,000-image dataset. Our results showed that the 3D reconstruction error was reduced in all of the experiments after incorporating the US beamwidth. The reduction ranged from a minimum of 2.2% to a maximum of 64.8%. This improvement was statistically significant ($p < 0.001$). This result confirmed that it is possible to improve the calibration accuracy beyond limits imposed by the US beamwidth. Moreover, the calibration accuracy can be significantly improved if the beam pattern can be properly obtained and incorporated into the process.

Furthermore, our real-time experiments have shown that USB-FW is effective with a small number of input images. This result suggests great potential for intraoperative use or other time-constrained clinical applications in which, typically, only a limited number of data can be accessed. With USB-FW, we can introduce an efficient form of quality assurance on calibration outcomes that are performed in the operating room.

6.1.3 Intraoperative Calibration for Prostate Brachytherapy

We are the first to propose a new device and an automated, computationally based TRUS calibration solution for brachytherapy systems. This system is designed to be
used in the operating room while the patient is being prepared for surgery. Four types of validation tests were conducted. In all of the experiments, the proposed method consistently achieved sub-millimeter accuracy and precision, high robustness, and good compatibility with various commercial ultrasound machines and brachytherapy steppers.

In total, 50 independent, real-time calibration trials yielded an average of $0.57 \pm 0.13$ mm line reconstruction error (LRE) relative to ground truth. The average LRE was $0.37 \pm 0.25$ mm relative to ground truth in tests with six different commercial TRUS scanners operating at similar imaging settings. Furthermore, testing with five different commercial stepper systems yielded an average of $0.29 \pm 0.16$ mm LRE relative to ground truth. Finally, the system achieved an average of $0.56 \pm 0.27$ mm target registration error (TRE) relative to ground truth in needle insertion tests conducted with a brachytherapy template in a water tank.

These results demonstrate that the proposed method is capable of providing an accurate, robust, and efficient means of quality assurance in the operating room.

6.2 Discussion and Future Work

6.2.1 Automated, Freehand Calibration With Accuracy Feedback

The Double-N phantom design produced fast, accurate calibrations with minimal operator interaction. The system required more images than other related systems, and the phantom design could be improved. One notable limitation of the phantom design is its lack of temporal calibration.
Our system required less than 60 images with an average of 2 data points (2 sets of N-wires) each to converge. In comparison, other systems required 30 images with an average of 4 N-wire sets per image [62, 83, 12] or 6 images with an average of 19 N-wire sets per image [35]. In this work and others, we have observed that approximately 120 data points is the minimum value required for a high-accuracy calibration. Other phantom designs have provided more data per image, but at the cost of complexity and a requirement for human interaction during the image-segmentation stage. We have made a trade-off between the number of images required (more), the level of human interaction (zero), and the phantom’s complexity (minimum).

The calibration phantom described here is a preliminary design and can be improved in future work. The small size of the phantom limited our freehand motion. The phantom design was optimized for high-frequency ultrasound probes and is not suitable for low-frequency probes.

The phantom was inspected by a local hospital’s biomedical engineering group and was deemed suitable for intra-operative use. The phantom, including its dynamic reference, was successfully sterilized using a low-temperature hydrogen peroxide plasma. In one test, some of the nylon wires failed due to melting. Future versions of the phantom may use a different wire material or may be designed for easy assembly in the operating room. A simple solution, for example, would be to use surgical clips instead of a silicon glue to affix its wires and adjust its wire tension.

The automated segmentation algorithm relies on the collinearity of points and parallelism of the lines for detection. This is an effective and practical approach, however the fundamental idea we proposed here is to introduce a machine-recognizable pattern into the phantom geometry that can be employed as a prior knowledge to
defeat false-positive detections in feature extraction. Many other useful patterns (e.g., triangle or circle) can therefore be adopted and there are many mature computer-science/machine learning algorithms available to detect these versatile patterns more reliably and robustly. One possible and more flexible solution, for example, is to use GPU-based Hough Transform detection which can work with different image patterns and run sufficiently fast by a parallelly structured graphics engine.

A minor technical limitation of our system is that we did not perform temporal calibration, i.e., we did not tightly synchronize tracking and image capture [50]. While the effect of temporal calibration is minimal when the probe moves relatively slowly, it would be relatively straightforward to include this potential improvement in a future generation of our design.

6.2.2 USB-FW: A Beamwidth-Weighted Calibration Framework

The design of the slice-thickness calibrator facilitated the quantitative extraction of US elevation beamwidth profiles. However, this measurement was manual, time-consuming and prone to human error, making it impractical for clinical use. Maintaining a 45-degree alignment between the inclined slope and the US scan plane also posed a challenge in the current design. We are working on a new design with an advanced software system that fully automates the procedure to achieve more accurate and repeatable beamwidth measurements.

The main advantages of weighted least squares are its efficiency in dealing with small datasets (crucial for a real-time application) and its capability to yield the most accurate parameter estimates possible for data points of varying quality [5]. Its main
weakness, however, lies in the basic assumption that the weights are known exactly, which is usually not the case in real-world applications. Estimated weights must be used as a workaround. This is precisely what we did to estimate the weights based on the US elevation beamwidth. Nevertheless, such estimations do not need to be very accurate (another benefit of weighted least squares), because empirical experience has established that small variations in the estimated weights do not typically influence the regression results and the parameter estimates [5].

In this work, we did not conduct specific tests to show if the USB-FW would improve the calibration precision (point reconstruction precision). However, both our exhaustive and real-time experiments demonstrated a reduction in the standard deviation of the PREs (especially in real-time tests, where the error of the standard deviation was reduced by half, on average, after USB-FW). These results may indirectly imply that USB-FW is capable of improving the calibration precision as well.

Furthermore, we designed USB-FW to combine the effects of both data filtering and weighted-least-squares estimation, and it remains a research interest to see how these two techniques would perform independently in terms of their calibration accuracy. The filtering and the weighted-least-squares estimations operated on different input data; the former eliminated data at the bottom of the US images (where the elevation beamwidth was the largest), and the latter processed the rest of the data. Therefore, their influence on calibration was of the result of two independent effects. Hence, we hypothesize that each effect alone would not be able to outperform the combination of both effects together. In our future work, we will conduct individual experiments to verify this hypothesis.


6.2.3 Intraoperative Calibration for Prostate Brachytherapy

To the best of our knowledge, this is the first automated method that has been proposed for template calibration in brachytherapy systems. The method is intended for use in the operating room while a patient is prepared for surgery. The method is fast and is computationally based to avoid subjectivity and human error. The elimination of conventional calibration sessions, which are lengthy, laborious, and periodic, will simultaneously achieve two main goals: it will make the brachytherapy system more accurate and consistent, as well as make it much less expensive.

In general, the novelty of the proposed iCAL system pertains to ultrasound-guided needle insertion procedures, where surgical end-effectors (such as a template, needle guide, needle holder or robotic needle driver) and TRUS images are spatially coregistered. The basic idea is to mechanically couple a precision-made calibration device with the surgical end-effector. There are many possible ways to accomplish this goal. However, for manufacturing simplicity and to preempt any sterilization issues in the operating room (i.e., the template needs to be sterilized during each treatment, thus making it impractical to directly assemble a calibration device to the template), we chose to combine the calibration phantom and an exterior replica of the surgical end-effector (template) in a unibody construct.

This principle can be equally applied to other forms of transperineal prostate interventions, including, but not limited to, localized therapies (thermal ablation, cryoablation, injections, etc.) and prostate biopsies.

We encountered a number of challenges in the design and development of iCAL. First, we incorporated a large opening on the container box to provide easy access to the iCAL phantom’s inner wires (Figure 5.4). This created difficulties in sealing
water inside the phantom. The water weight (about 900 grams) increased the pressure within the container box, causing a slow water drip along the edges of the rubber seal. In final production, the phantom will be sealed with a tissue-mimicking gel that is less prone to leakage problems. We are also experimenting with optimal container box designs (e.g., downsizing the container to reduce water weight) and better sealing methods to keep the phantom completely dry.

Another challenge was ensuring sufficient contact between the TRUS probe and the iCAL phantom’s rubber window for proper imaging. Naturally, ultrasound gel was used to provide acoustic coupling between the transducer crystals and the rubber window. However, because the iCAL phantom was mounted directly on the brachytherapy template holder (located on top of the transverse transducer) (Figure 5.2), the coupling gel easily fell out (due to gravity) when the probe was displaced and/or rotated underneath the rubber window, causing a loss in visibility of the phantom during imaging. We are exploring more efficient and reliable coupling methods to resolve this issue.

The weight of the phantom has caused an additional problem. With the water sealed inside, the complete phantom assembly weighs approximately 1.8 kilograms. Because the entire phantom is affixed to the brachytherapy template holder, the weight caused the phantom to sag, introducing errors into the calibration. This problem was easily remedied by lifting the sagging end of the phantom with a supporting stand. In the next release, we will optimize the phantom’s design by providing appropriate structural support based on its weight and size.

All of the aforementioned challenges are related to production engineering, where work is currently underway to make iCAL ready for clinical trials.
In this work, iCAL was only used to calibrate the transverse image of the TRUS probe to a brachytherapy template. Our next step is to extend the automated calibration method to the sagittal plane so that the TRUS probe’s transverse and sagittal images can both be accurately coregistered. Combining precisely calibrated transverse and sagittal images together will produce a more accurate reconstruction of the TRUS volume.

Finally, it is important to note that iCAL determines the intrinsic quality of a brachytherapy system (i.e., the accuracy of the brachytherapy device itself), but it cannot prevent extrinsic errors caused by ultrasound signal refraction, needle bending/deflection, tissue deformation or organ motion. These errors are the result of random, unforeseeable factors that cannot typically be detected during the system-calibration stage.

6.3 Potential Use of Beamwidth in Other US-Guided Interventions

In our research we have developed a new calibration strategy for tracked US imaging: to account for the higher localization uncertainty associated with regions of large slice thickness in a weighted least-squares fitting technology, as opposed to the conventional approaches that treat all the input US data equally in their influence in the calibration process.

In general, localization errors and image artifacts that are caused by US slice thickness are inherent problems for US imaging. The novelty of beamwidth weighting not only pertains to tracked ultrasound calibration procedures, in which a surgical
instrument and US images are spatially coregistered but also to typical registration problems. In the latter case, tracked US data need to be aligned with various imaging modalities, such as CT, MR or US itself. The basic principle is that US data should be treated non-uniformly with respect to localization uncertainties associated with its US beamwidth. An adaptive approach may lead to more accurate and consistent spatial registration, just as we observed in the case of probe calibration. We are currently exploring the potential of applying a similar approach to US image registration by adaptively registering image pixels based on their beamwidth information.

During our phantom validation, we observed that a relatively simple modification to the basic least-squares algorithm led to significant improvements in our calibration results. Furthermore, we believe that compensating for slice thickness may make a difference in 3D US reconstruction and elastography. Considering the non-uniform elevation beamwidth may offer significant advantages when computing the partial overlap (speckle correlation) between US image slices in the elevation direction [43].

Our experiments also revealed that, without beamwidth weighting, the calibration accuracy was roughly proportional to the average slice thickness of the imaging system. This suggests that localization errors and uncertainties are proportional to the slice thickness at various depths within the effective imaging range of a given US system. More importantly, it also suggests that US beamwidth profiling can be used as a quality assurance tool to quantitatively determine the US imaging resolution.

Work is currently underway with PARTEQ Innovations (Queen’s University technology transfer office) and Precision Therapeutic, Inc. (Ontario, Canada) to commercialize this technology.
Bibliography


