NAVIGATING IN PATIENT SPACE USING CAMERA POSE ESTIMATION RELATIVE TO THE EXTERNAL ANATOMY

by

Jihang Wang

B.S, Northeastern University (CN), 2010

M.S, Carnegie Mellon University, 2012

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Jihang Wang

It was defended on

November 29, 2016

and approved by

Howard Aizenstein, M.D., Ph.D., Associate Professor, Department of Bioengineering,

University of Pittsburgh

Jelena Kovacevic, Ph.D., Hamerschlag University Professor and Head, Department of ECE |

Professor, Department of BME, Carnegie Mellon University

Dissertation Co-Director: George Stetten, MD, Ph.D., Professor, Department of

Bioengineering, University of Pittsburgh

Dissertation Co-Director: John Galeotti, Ph.D., Adjunct Assistant Professor, Department of

Bioengineering, University of Pittsburgh | Systems Scientist, Robotics Institute, Carnegie

Mellon University

NAVIGATING IN PATIENT SPACE USING CAMERA POSE ESTIMATION

RELATIVE TO THE EXTERNAL ANATOMY

Jihang Wang, B.S., M.S., Ph.D.

University of Pittsburgh, 2016

Ultrasound probe localization is essential for volumetric imaging with a 2D ultrasound probe, and for establishing a recorded anatomical context for ultrasound-guided surgery and for longitudinal studies. The existing techniques for probe localization, however, require external tracking devices, making them inconvenient for clinical use. In addition, the probe pose is typically measured with respect to a fixed coordinate system independent of the patient's anatomy, making it difficult to correlate ultrasound studies across time.

This dissertation concerns the development and evaluation of a novel self-contained ultrasound probe tracking system, which navigates the probe in patient space using camera pose estimation relative to the anatomical context. As the probe moves in patient space, a video camera on the probe is used to automatically identify natural skin features and subdermal cues, and match them with a pre-acquiring high-resolution 3D surface map that serves as an atlas of the anatomy. We have addressed the problem of distinguishing rotation from translation by including an inertial navigation system (INS) to accurately measure rotation. Experiments on both a phantom containing an image of human skin (palm) as well as actual human upper extremity (fingers, palm, and wrist) validate the effectiveness of our approach.

We have also developed a real-time 3D interactive visualization system that superimposes the ultrasound data within the anatomical context of the exterior of the patient, to permit accurate anatomic localization of ultrasound data. The combination of the proposed

tracking approach and the visualization system may have broad implications for ultrasound imaging, permitting the compilation of volumetric ultrasound data as the 2D probe is moved, as well as comparison of real-time ultrasound scans registered with previous scans from the same anatomical location. In a broader sense, tools that self-locate by viewing the patient's exterior could have broad beneficial impact on clinical medicine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0   INTRODUCTION

1.1   THESIS STATEMENT

This thesis describes the development and validation of a novel self-contained ultrasound probe tracking and visualization system, to address the problem of freehand ultrasound probe tracking without requiring an external tracking device, by mounting a video camera on the probe to identify location relative to the patient's external anatomy, making use of a high-resolution 3D surface map as an atlas of the anatomy.

1.2   OVERVIEW OF CONTRIBUTIONS

I.   We built a stand-alone system that allows estimation of the probe's pose by tracking natural skin features and subdermal cues (vein patterns).

II.   We implemented an approach to incorporate a prior high-resolution 3D surface map as an atlas of the anatomy to overcome error accumulation for ultrasound probe tracking, which exists in prior frame-by-frame probe tracking systems.

III.   We address the problem of distinguishing rotation from translation by including an inertial navigation system to accurately measure rotation for ultrasound probe tracking.

IV.   We systematically evaluated the system on the upper extremity of a human subject, and proposed an efficient solution to reduce error in challenging tracking regions.

V.   We developed a 3D interactive visualization system, enabling real-time display of the patient's anatomy with the ultrasound probe and corresponding ultrasound image superimposed in the correct pose.

1.3   THESIS ORGANIZATION

The next chapter describes relevant background and the significance of this project, including clinical motivation, previous and current work for freehand ultrasound probe tracking, and the proposed technique for probe tracking. Chapter 3.0 encompasses the hardware design of the system, including camera, inertial navigation system (INS) and probe housing, phantom development and light studio development. Chapter 4.0 summarizes methods and mathematical techniques we applied in the proposed tracking system, including the implementation of coordinate transforms, a description of a variety of image matching/optimization algorithms, and the application of image pre-processing routines. Chapters 5.0 and 6.0 describe application of the probe localization system to freehand 3D US in a phantom and in a real human subject, respectively. Chapter 7.0 presents the 3D interactive visualization system. Finally, Chapter 8.0 concludes the thesis by summarizing the contributions and limitations. Related projects in our laboratory and future work are discussed in this chapter as well.

2.0   BACKGROUND AND SIGNIFICANCE

In this chapter, the clinical significance of the project is discussed in section 2.1. The related work has been summarized in section 2.2, and the proposed system is summarized and compared with similar work in section 2.3.

2.1   CLINICAL SIGNIFICANCE

Composite tissue allotransplantation (CTA) is the transplantation of a composite tissue that may include skin, muscle, bone and nerve, from a donor to a recipient. These techniques have been applied to the hand, face, abdominal wall, larynx and other body parts, which has the potential to be useful for functional restoration of patients with severe tissue loss, as encountered with massive burns, traumatic injuries, congenital anomalies, and following tumor resection. Nerve regeneration poses a particular challenge for CTA, but the biggest hurdle for CTA is overcoming immune modulation. The necessity of immunosuppression and the potentially profound negative implications of immunosuppressants are well known and are a significant aspect of CTA. Immunosuppressants are toxic, and their complications include increased incidence of opportunistic infections, increased risk of malignancies, and chronic rejection. While there have been significant inroads and improvements in addressing toxicity, this remains a significant

hurdle to widespread application of the technology (Girlanda 2013; Lovasik et al. 2011; Klimczak and Siemionow 2007).

2.1.1   Nerve Regeneration

Unlike in solid organ transplants, successful nerve regeneration is critical to functional success after CTA (Yan et al. 2013). For example, a limb transplant could survive but be functionless or poorly functional due to lack of optimal nerve regeneration. Following a hand transplant, the distal donor nerve undergoes a degeneration process of the distal nerve stump referred to as Wallerian Degeneration (WD). This degeneration and the resultant death of the axons in the fascicles leaves only the nerve sheaths and tubes, which may serve as conduits for regeneration of the proximal host nerve ends, regenerating at a rate of 2-3 mm/day. Distal muscles that have lost their innervation have impaired or no function and will undergo permanent denervation atrophy if the nerves do not reach the motor end plates in time (Stoll and Müller 2006). Nerve regeneration is also critical to the outcomes after peripheral nerve injury (PNI), which occurs when a nerve is compressed, crushed or severed and proper communication between the distal peripheral and central nervous system is lost (Svennigsen and Dahlin 2013). The ability to objectively diagnose nerve injury or monitor nerve regeneration in patients undergoing regenerative, repair or transplant strategies (recipients of CTA such as limb transplant or after PNI treatments such as nerve repair) in a sequential, non-invasive manner is a significant step in the management of these patients. Better defining and monitoring the degree of injury and regeneration would help determine a timely and accurate treatment strategy. Novel neurotherapeutic agents (drugs such as tacrolimus, biologics such as chondroitinase, growth factors such as IGF1) that maximize nerve regeneration and enable motor/sensory recovery after

injury hold the promise of helping patients return to functional/employable status while improving performance, morale, mobility, agility and capability (Apel et al. 2010; Kostereva et al. 2016). Improved monitoring of regeneration is essential to such research, and could also facilitate prosthetic planning in amputees undergoing rehabilitation and improve consequent outcomes by delineating neuromas or heterotopic ossification in a dynamic manner.

Currently, the techniques for monitoring nerve regeneration in humans such as electromyography and nerve conduction studies have several limitations including poor localization, and variations of reading with the operator, ambient temperature and patient age. MRI is capable of imaging nerves with high resolution, but is expensive, non-portable, and not always accessible, making it non-ideal for longitudinal studies of nerve growth. Furthermore, MRI is not safe for patients with embedded metal shrapnel such as veterans with PNI due to blast trauma.

2.1.2 Chronic Rejection

Chronic rejection (CR) has long been the predominant cause of graft failure after organ transplantation, and it is now apparent that it can compromise long-term graft survival after CTA as well. While acute rejection develops between 3 and 6 months after transplantation, CR occurs slowly over months to years after transplantation and progresses throughout the life of the transplant. Therefore, CR is more difficult to detect compared with acute rejection. To date, 78 upper extremity transplants have been performed on 55 recipients around the world, with 7 at the University of Pittsburgh (UPMC), and 2 of the 55 patients have already been lost to sepsis from CR (Gorantla and Demetris 2011; Schneeberger et al. 2013). Currently, the key manifestation for detecting CR is increased thickness of intimal hyperplasia (IH) in the vasculature, on what is

called graft vascular disease (GVD). This progressive, insidious thickening of the lining of grafted blood vessels eventually results in the occlusion of the lumen, ischemia and graft loss (Valenzuela and Reed 2011).

Although skin biopsies are able to detect CR, CR changes are not manifested early in the small, near-surface blood vessels in the skin, but rather in the deeper, medium sized elastic arteries of the limb, where biopsies are not practical (Shulman et al. 1978). Other technologies for monitoring blood vessels include CT angiography and intravascular ultrasound. Both suffer from risks that could compromise the graft. Intravascular ultrasound requires an invasive catheter, and CT angiography is not suited for sequential studies due to accumulated radiation dose. These imaging methods are thus precluded, even though they might otherwise afford sensitive detection of subtle changes in intima or lumen of the vessels. The inability of conventional techniques to diagnose and monitor IH as the primary predictor of CR has been a limiting factor in long-term treatment after CTA.

2.1.3 Ultrasound Imaging

External (non intravascular) ultrasound (US) is a useful imaging modality for monitoring many anatomical and physiological properties in real time. It is low-cost, real-time, portable, and completely safe, with neither ionizing radiation nor strong magnetic fields (Duric et al. 2007). Therefore, it has become a standard of care for monitoring changes in anatomical structures, diagnosis of disease and guidance of interventional procedures.

External ultrasound scanners typically operate at frequencies between 1 to 20 megahertz (MHz), the spatial resolution, which defines the distance between two scatterers at which they are resolvable, is inversely proportional to frequency. Thus resolution improves with increasing

frequency. However, the attenuation of the ultrasound beam as it propagates through the tissue is also strongly dependent on frequency. The relationship between attenuation coefficient (in dB cm-1) and frequency (Hz) is approximately linear (Webb and Kagadis 2003). Therefore, the choice of transducer is always a trade-off between depth of penetration and resolution. The higher the frequency, the greater the resolution, but the less depth of tissue accessible (Wakefield 1999). This trade-off between resolution and penetration depth as a function of frequency is shown in Figure 2-1. For this reason, lower frequencies transducers (1 – 6 MHz) are used to study deep structures such as liver and kidney, while higher frequencies transducer (7 – 20 MHz), which provide better spatial resolution, are used to study more superficial structures such as muscles, tendons, nerves and breast.
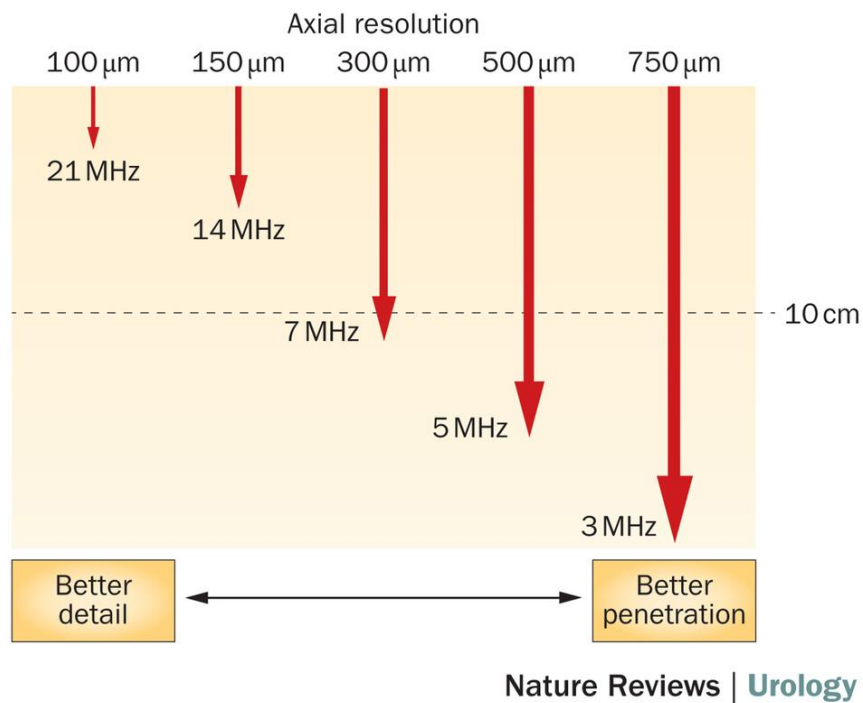
Figure 2-1: The trade-off between resolution and penetration depth as a function of frequency during ultrasound imaging (Fried and Burnett 2015).

8

2.1.4 High Resolution Ultrasound (HRUS)

The non-ionizing radiation of ultrasound makes it ideal for longitudinal studies where particular anatomy needs to be imaged repeatedly over a period of time. However, traditional ultrasound which operates at frequencies between 1 to 20 MHz is not able to provide sufficiently high resolution for the changes in arterial walls during IH. Solid-state array transducers have recently been developed capable of imaging at frequencies up to 70MHz at high frame rates. Such scanners can acquire ultrasound images at 30-micron resolution. HRUS has made possible the visualization of individual fascicles in nerves as well as IH in arterial walls associated with GVD, therefore providing a potentially effective manner to monitor nerve regeneration and CR after hand transplant surgery.

Figure 2-2 shows an example of a median nerve image in a human arm using state-of-the-art 50 MHz ultrasound. We can clearly see the fascicles (which contain small bundles of nerve fibers, not visible) in the HRUS scan.
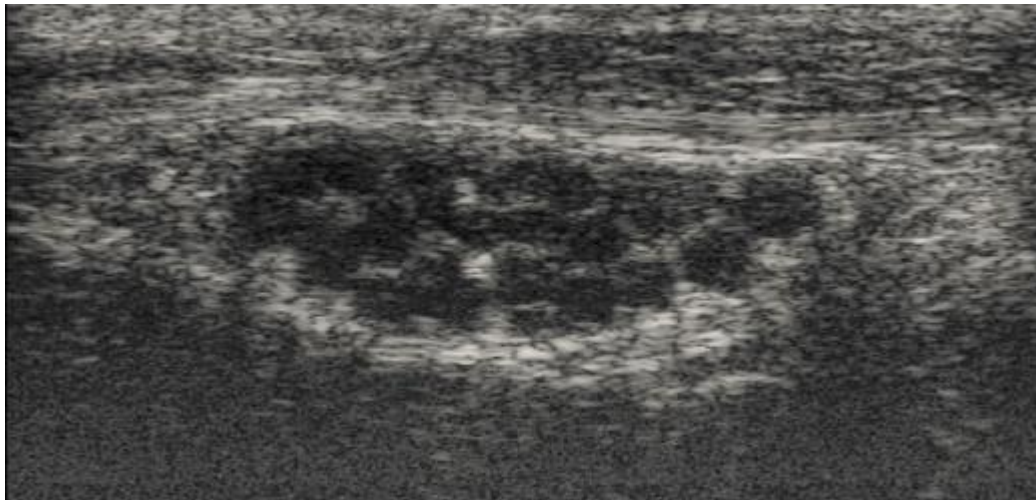
Figure 2-2: Nerve image with individual fascicles scanned using VisualSonics Vevo 2100 system at 50 MHz.

However, the fine detail within the nerve, which is visible for the first time with HRUS, comes with a price; depth of penetration at 50-70 MHz may be less than 1 cm, and so frequency (and resolution) must be traded off against depth of penetration, depending on the particular target, even as depth increases as one follows a particular nerve proximally into the body.

Chronic Rejection (CR) can also likewise be monitored using HRUS by measuring the thickness of the arterial wall and the patency of the lumen. We have shown such measurements to be possible using our Vivo 2100 ultrasound scanner at 70 MHz, as shown in Figure 2-3. The hand transplant program at UPMC is one of the two programs using HRUS to monitor CR changes in their patients (Kaufman et al. 2012). Early evidence suggests that both arteries and veins may be primary targets of CR in the hand.
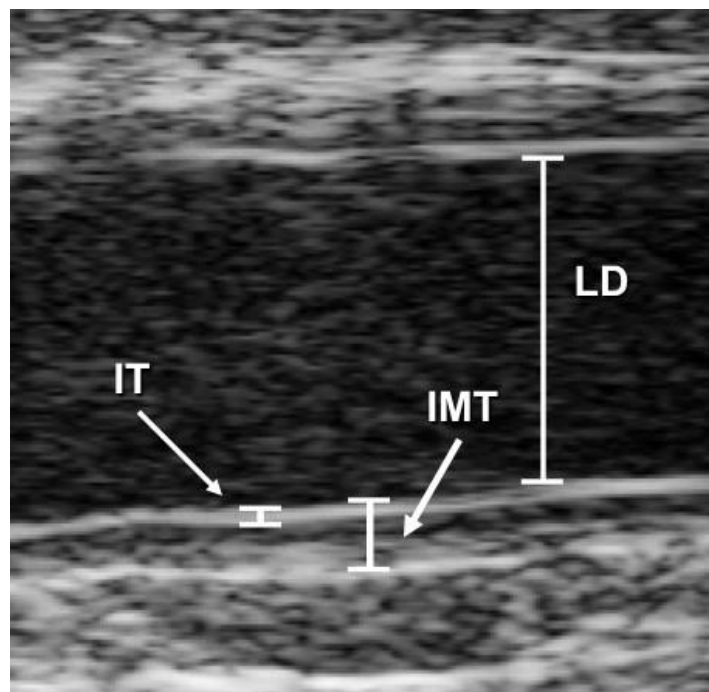
Figure 2-3: HRUS image (Vivo 2100) of artery showing measurement of Intimal Thickness (IT), Intima-Media Thickness (IMT), and Lumen Diameter (LD).

Although the UPMC experience confirms that HRUS is a useful tool to evaluate IH in vessels, a serious limitation remains in the local 2D nature of the current HRUS technology. During imaging, the operator's mind is continually aware of the location of the ultrasound images relative to the patient's external visible anatomy. However, after a free-hand scan is finished, even the most skilled clinicians find it challenging to effectively register the previously scanned 2D ultrasound images from a free-hand scan with the patient's anatomic 3D context, since the probe no longer provides a mechanical cue to image location. These challenges for US systems are intrinsic due to its lack of a stored location information, or even a stable coordinate system, compared with other image modalities such us CT or MRI, which store inherent coordinates.

2.2   RELATED WORK ON TRACKING FREEHAND ULTRASOUND

2.2.1   Early mechanical and acoustical tracking

To address this, some early work explicitly tracked coordinate values either acoustically or mechanically. For example, Dekker et al., developed a system that utilized an instrumented, 5-DOF mechanical arm to monitor the position of a single beam echocardiographic probe as Figure 2-4 shows (Dekker, Piziali, and Dong 1974). Mechanical arms are accurate, but can only track one object at a time, which can be a limitation during surgery. Moritz et al., used acoustic sensing for determining the probe's pose by measuring the transit times of spark-generated shock waves (Moritz and Shreve 1976; Moritz et al. 1983). Acoustic tracking technologies are

cumbersome and affected by variations in temperature, pressure and humidity, all of which affect

the propagation speed of sound in air (Mercier et al. 2005).
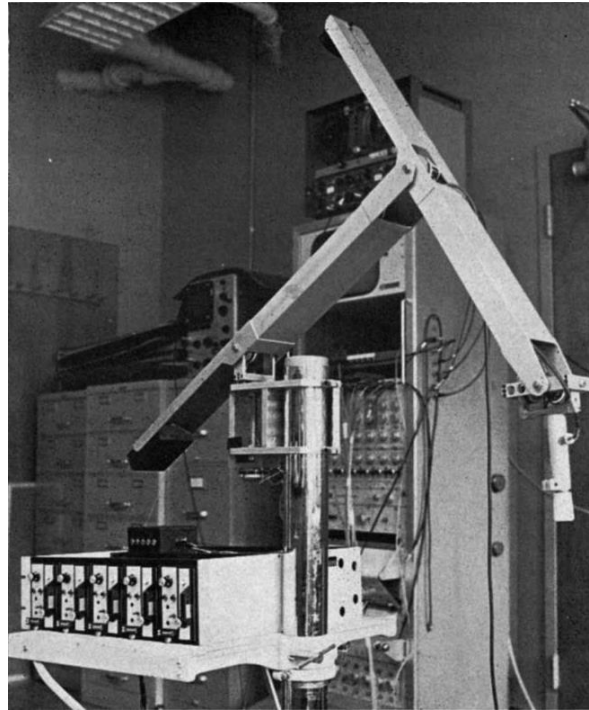
Figure 2-4: Photograph of the five degree of freedom arm including signal conditioners and stand. (Dekker, Piziali,

and Dong 1974)

2.2.2 Optical and electromagnetic tracking

Optical and electromagnetic (EM) tracking systems have been adopted in recent decades because

of their high accuracy. They generally consist of a physical tracking object (attached to an

ultrasound probe) and an external device that calculates the tracking object's position and

orientation at any point in time.

Two types of optical trackers exist, both used widely in clinical applications. For active optical tracking systems (Mills and Fuchs 1990; Gobbi, Comeau, and Peters 1999; Blackall et al. 2000; Han, Noble, and Burcher 2003), sequentially illuminated LEDs are mounted on the probe as markers, tracked by either two or three linearly mounted charged coupled device (CCD) camera units. The location of each LED is calculated by triangulation, based on the known geometric configuration of the CCD elements. A minimum of three non-collinear LEDs rigidly mounted as an array is necessary for determining 6-DOF pose information of the array. Since the LEDs must be powered, traditionally active systems were wired systems. For passive optical tracking systems, a pattern of the reflective markers in a unique pattern is attached to each item to be tracked so that independent determination of pose is feasible using an array of video cameras. One major advantage of these passive systems is that no wires are needed between the tracking system and the tracked probe (Cash et al. 2005; Flaccavento, Lawrence, and Rohling 2004; Poon and Rohling 2007). Although optical tracking technology is able to provide submillimeter accuracy in the clinical environment, the system suffers from several problems, including limited field of view (FOV), tracking inaccuracy due to small-size fiducials, and line of sight requirement between the cameras and the target.

Another method for probe pose tracking involves the use of an external EM tracker (Boctor et al. 2003; Hastenteufel et al. 2006; Huang et al. 2005; Hummel et al. 2008; Nakamoto et al. 2008). The idea behind the EM system is to have a receiver placed on a probe that measures the induced electrical currents from a magnetic field generated by either an alternating current (AC) or direct current (DC) transmitter. The EM tracker does not require line of sight between the transmitter and sensor, but ferromagnetic objects and electronic equipment near the tracker can severely affect tracking accuracy. Therefore, the EM tracking system is challenging

13

to use in an environment such as an operating room, where metallic objects and electronic equipment are likely to be present in the field (Birkfellner et al. 1998).

All the tracking devices discussed above work in a similar manner: the device tracks the position and orientation of specific devices mounted on the probe, which restricts motion of the probe during the procedure and requires separate tracking of patient location for determining the pose of the probe relative to the anatomy.

An alternate approach is to mount video cameras directly on the ultrasound probe, which has been used to estimate both relative needle pose and trajectory with respect to the probe (Sauer, F., Khamene 2003; Chan, Lam, and Rohling 2005). Following this approach, Rafii-Tari, et al., proposed a solution for determining the relative pose between the probe and patient anatomy, by attaching passive optical markers to the skin and mounting a camera on the ultrasound probe, permitting accumulation of 3D ultrasound data and navigation within that data (Rafii-Tari, Abolmaesumi, and Rohling 2011). Sun et al. have developed a similar system to estimate probe pose with respect to the body by tracking artificial skin markers (Sun, Gilbertson, and Anthony 2013). Requiring attaching artificial markers to be attached to the patient's skin can be problematic for longitudinal studies, since they need to be in place from one scan to the next. Surface markers can also block the passage of US into patient, and may degrade or be distorted when exposed to US gel. More recently, Sun et al. have developed a camera based system that tracks skin features directly, similar to our work described here, except that they still require an artificial marker to be affixed to the skin for scale calibration (Sun, Gilbertson, and Anthony 2014).

2.2.3 ProbeSight tracking

Along the lines of the research above, my master's thesis developed a novel method for localizing the probe's pose in real time by mounting stereo cameras on the US probe. The device is called ProbeSight, since it was, in effect, giving vision to the ultrasound transducer. By doing so, this approach allows the compilation of 3D data combining the surface and subsurface of the patient, without external tracking of the US probe.

The embodiment of this concept is as Figure 2-5(A) shows. Twin video cameras in a lightweight aluminum frame provided a view of the surface of a standard ultrasound phantom, upon which a sheet of tracing paper has been laid and saturated with US gel. A predetermined checkerboard pattern was printed on the tracing paper using an inkjet printer (a LaserJet printer would melt the wax within the paper) to provide rich features for stereo vision algorithms. The saturated tracing paper did not significantly interfere with the passage of ultrasound into the phantom. The position of the phantom relative to the cameras was computed in real time by standard stereo vision algorithms, wherein the separate viewpoints of the cameras allowed triangulation of the 3D coordinates of observed points. Each point thus identified on the phantom surface could be localized into the frame of reference of the cameras and, by extension, of the ultrasound probe itself. We demonstrated and validated this first-generation implementation of ProbeSight by demonstrating the real-time 3D graphical rendering of the phantom's surface with the ultrasound data superimposed at its correct relative location, as shown in Figure 2-5(B). Both the re-projection of the surface with the checkerboard pattern and the ultrasound data (containing a small tube in a cross-section) can be seen in the correct 3D location with respect to the probe and camera (Wang et al. 2012). The camera pose relative to

phantom's surface was correctly computed from the checker board point correspondences in the stereo video images.

As in previous research, the first-generation ProbeSight still required an artificially generated pattern to be affixed to the target.
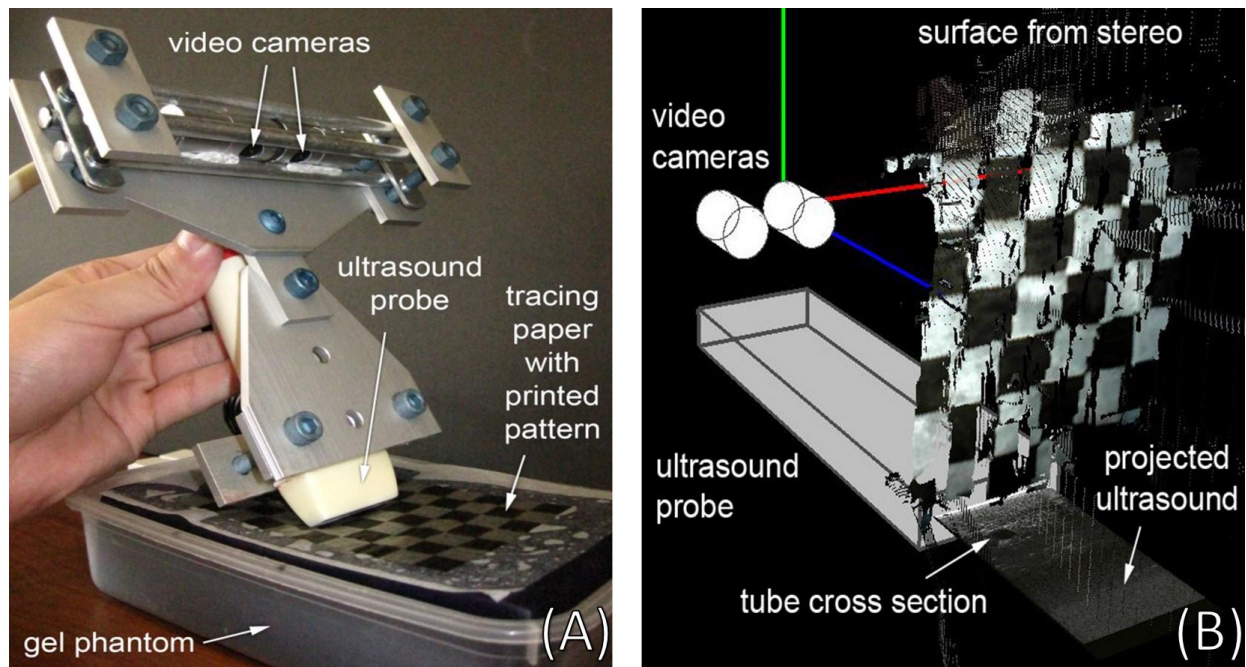
Figure 2-5: (A) Apparatus and test phantom with printed texture surface. **(B)** Real-time 3D simultaneous rendering of the gel phantom surface (from stereo), ultrasound, and probe/camera locations (Wang et al. 2012).

2.3   PROPOSED SYSTEM FOR PROBE LOCALIZATION

Tracking skin features directly is a way to inherently avoid the problems mentioned above with artificial markers.   In this thesis, a second-generation ProbeSight system is proposed and

developed for probe localization. This technique computes the pose of the ultrasound probe using the natural visual features in the patient's external anatomy. In particular, the algorithms operate by matching the probe-mounted camera's images of the patient's skin against an inverse rendered high-resolution surface map model. The surface map is acquired in advance using a multiple-camera imaging device designed for reconstructive surgery. With this prior information, we are able to match the surface map against images from the camera mounted on the ultrasound probe. The ultrasound camera is able to capture details such as skin folders and creases. We also mount a microelectromechanical systems (MEMS) inertial navigation system (INS) on the US probe, which we use in conjunction with the video algorithms to improve the accuracy of our estimates of probe pose.

The second-generation ProbeSight system differs from previous work summarized in section 2.2 by tracking natural skin features, by comparing these features with the prior scanned 3D high-resolution map, and by using an INS. To our knowledge, we are the first to develop each of these approaches. In addition to combining the benefits from both a pre-scan 3D surface map and INS, we also developed new methods and approaches that have made ProbeSight more reliable.

In our application, ProbeSight provides a capability of acquiring an anatomical context for the US data in terms of the exterior of the patient. This permits accurate anatomic localization of the US data, without using external tracking devices. Knowing the anatomic location from which images of nerves, arteries, ectopic ossifications, and other structures have been obtained, and being able to obtain follow-up images from the same locations, will facilitate understanding the condition of these tissues after injury and during healing, leading to improved strategies of therapy. Such localized data could be correlated across time and across patients, as

well as registered with other modalities such as MRI using diffusion tensor imaging of regenerating nerves in transplanted limbs. It could also enable guidance of the US probe in real-time to the same anatomical location as a previous scans. ProbeSight could facilitate collection of 3D ultrasound images in patients who are contraindicated for MRI (such as those with metal implants or shrapnel). In a broader sense, the underlying concept of ProbeSight could be extended to other clinical tools, such as scalpels or needles. Knowing where such tools are by having them, in effect, look at the patient's exterior could have an important beneficial impact on many aspects of clinical medicine.

3.0   HARDWARE DEVELOPMENT

This chapter presents ProbeSight's hardware and software design. The camera, INS and optical marker housing are first described, followed by phantom development. The use of a custom hand brace and lighting studio is then described. Finally, the software design of ProbeSight is discussed.

3.1   PROBE HOUSING WITH CAMERA, INS AND OPTICAL MARKER

For the proposed system, an Ethernet-powered machine vision camera (Prosilica GT1290C; Allied Vision Technologies, Canada) is rigidly mounted onto the ultrasound probe using a custom-machined aluminum probe bracket, designed to rigidly hold the camera onto the ultrasound probe. This probe bracket is composed of two parts: (1) a double-C shaped clamp that bolts together to precisely fit the circular slot at the tail of the ultrasound probe, preventing any relative motion between them, and (2) a prototype platform with half-inch hole spacing to mount the camera (Figure 3-1).
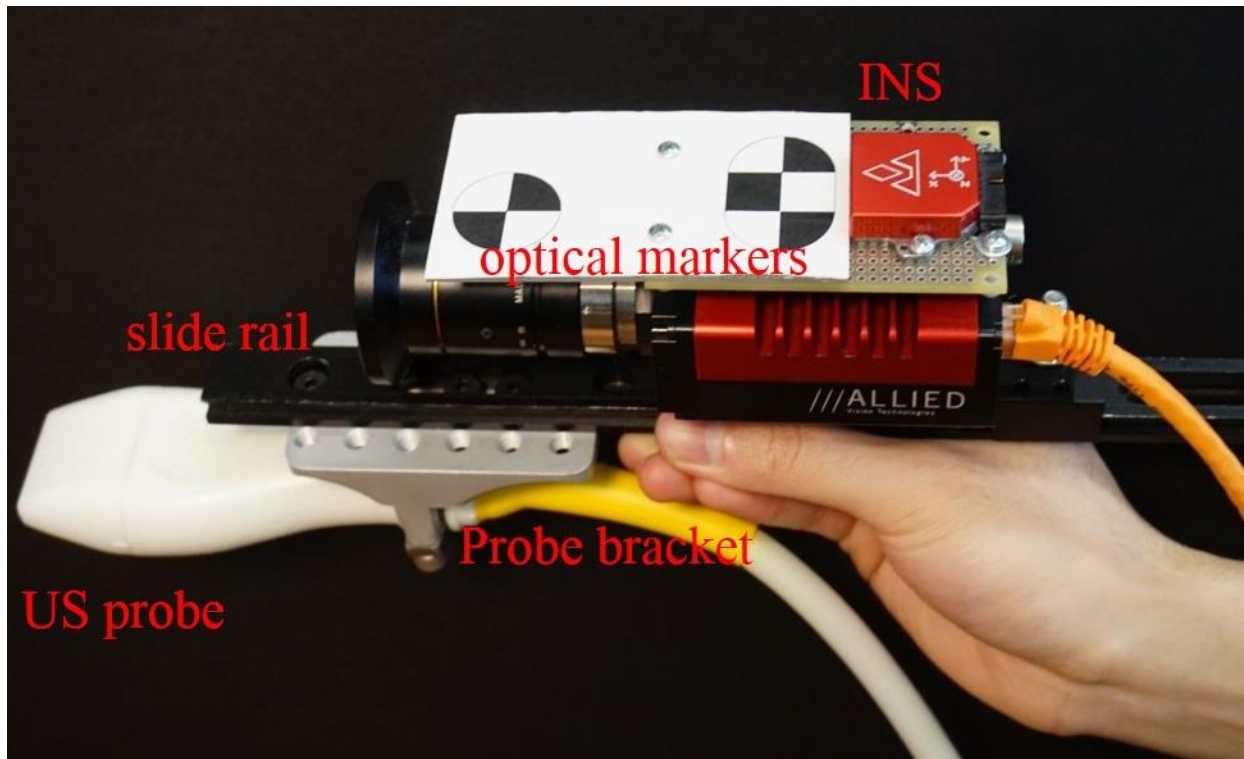
Figure 3-1: Ultrasound probe with mounting bracket and slide rail for the camera and INS.

To ensure high-resolution camera images, the machine vision camera is equipped with the then-best-in-class Sony ICX674 CCD 2/3" sensor, housed in the smallest and lightest case that provides sufficient heat dissipation to mitigate thermal noise (Dimension: 2.323" L x 1.811" W x 1.299" H, including connectors, w/o tripod and lens, mass: 106g).

In this experimental prototype, a slide rail allows for adjustment of the distance between the camera and skin surface, over a range limited by the camera's ability to focus to a minimum of 10 cm, particularly suitable for ProbeSight. When the camera image is cropped to 640 x 480 pixels from the camera's Bayer-pattern sensor (which has 1936 x 1456 pixels), the image maps to approximately a 6 cm x 4.5 cm field of view (FOV) on the patient's skin surface. The camera optical axis is parallel to the ultrasound plane and offset about 7 cm from the tip of the probe,

such that local skin surface deformation due to probe pressure should not cover the majority of the FOV (Figure 3-2). During scanning, camera frames are acquired synchronously at approximately 30 frames per second and archived as video files for later retrieval.
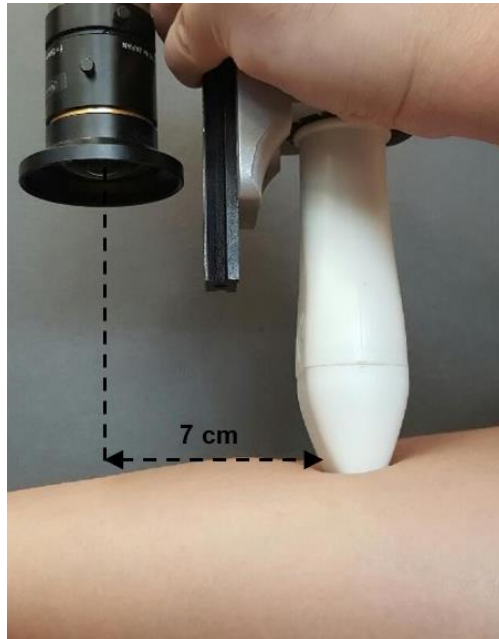
Figure 3-2: Picture of skin surface deformation due to probe compression.

When the US probe moves along the skin, its pose occupies a 6-DOF space as shown in Figure 3-3: translation along the x, y, and z axes, as well as the 3 rotations in pitch, roll, and yaw. ProbeSight operates by re-projecting the pre-acquired 3D surface map to simulate what would be seen by the real camera from the present point of view. Tracking is then achieved by estimating the probe's translation and rotation between the coordinates used to project the present simulated view and those required to match the next real camera image. When this work began, one of the biggest problems was robust estimation of the rotation between successive

21

camera frames. The effect of certain small translations and rotations on the displacement of the features in the image plane can be very similar, and this problem is made worse if the distribution of the features in the image plane is too homogeneous (Kneip, Siegwart, and Pollefeys 2012). To disambiguate between translation and rotation, we mounted a miniature, high performance INS on the probe (VN-100 Rugged; VectorNav Technologies, USA, Figure 3-4). The INS provides a real-time and drift-resistant 3D orientation measurement of the probe, which reduces the search space from 6 DOF to 3 DOF and resolves the ambiguity between rotations and translation by leaving the computer vision algorithms to determine only translation.
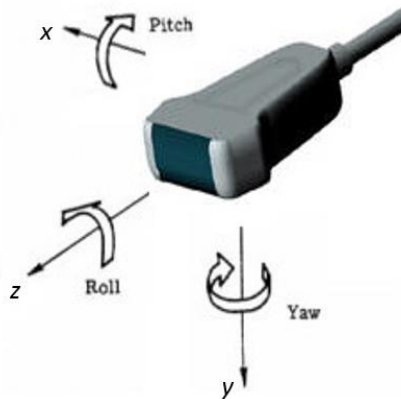
Figure 3-3: Schematic diagram of probe's rotation in space.

Figure 3-4: The VecterNav VN-100 Rugged IMU/AHRS.

The probe pose initialization and validation is assisted using a Micron Tracker optical tracking system (Sx60; Claron Technology, Canada), with optical markers attached to both the ProbeSight apparatus and to the phantoms. The Micron Tracker is a real-time optical pose tracking device, which uses ordinary visible room illumination to track passive markers attached to the object of interest Figure 3-5(A). Each marker consists of at least 3 black-white checkerboard intersections called *Xpoints*. The connection between two Xpoints is called a *Vector*. Figure 3-5(B) shows an L-shaped marker measuring 3.5 by 1.5 inches that contains 3 Xpoints and 2 vectors, providing an efficient use of marker area and making it an appropriate size for mounting on ProbeSight's video camera. The Micron Tracker processes images from its two internal camera sensors, to detect and correlate features from these small checkerboard patterns based on physically dimensioned marker templates in its database. To track a marker, each of the marker's Xpoints must be correctly recognized in each of the two images as Figure 3-5(C) shows.
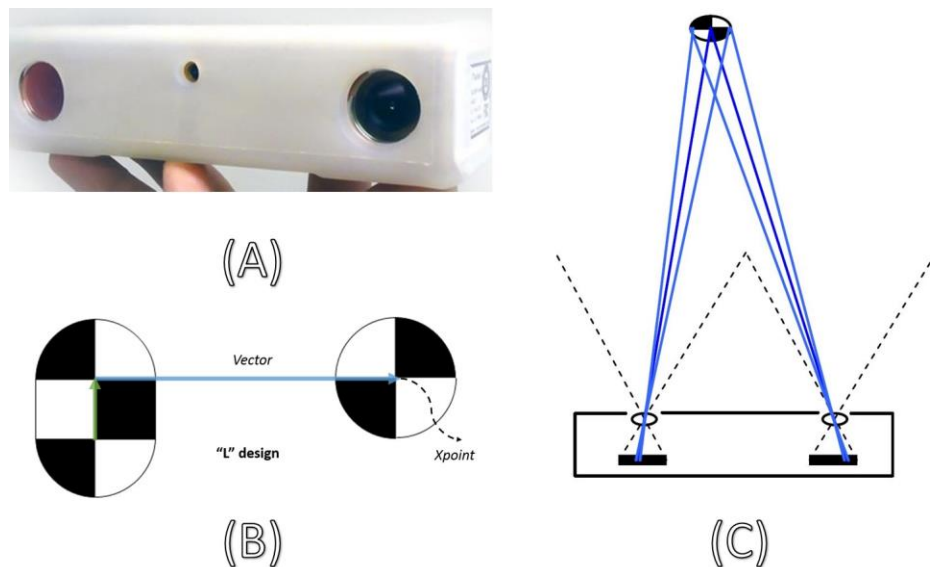
Figure 3-5: (A) The Micron Tracker and its coordinates. **(B)** An L-shaped marker consists of 3 Xpoints and 2 vectors. **(C)** Illustration of tracking process by using stereo camera.

23

3.2 PHANTOM DEVELOPMENT

3.2.1 Dinosaur model

A dinosaur model, roughly the same size as a human upper extremity, provided an ideal feature-rich environment that was used to initially test the inverse rendering process (which will be discussed in section 4.5.3) and the performance of the matching metric. As Figure 3-6 shows, the dinosaur model provided a wide variety of features, including color, surface curvature, edges and fine ridges on which image processing and computer vision algorithms can operate. Small white dots were painted on the surface of the dinosaur model as fiducials for coordinate transformation in the inverse rendering process, which will be discussed in section 4.3.

Figure 3-6: The dinosaur model for initial ProbeSight tracking accuracy test in each DOF.

3.2.2   Flat phantom

ProbeSight tracking was also tested on a flat acrylic phantom (Dimension: 6.920" L x 3.858" W x 1.575" H) on whose surface was attached a 2D printed high-resolution image of actual palm skin, acquired using a digital single lens reflex (dSLR) camera (Canon EOS 5D Mark I, Canon Inc., USA) with a macro-telephoto lens (Figure 3-7). A central rectangle was cropped from the palm area, and printed at 7" x 4". The cropped rectangle also served as the 3D surface model, as described further in section 4.3.

The flat phantom was developed as a transition from the feature-rich rigid dinosaur model to real skin on a human forearm. Multiple experiments were performed on the flat phantom including 1DOF, 2DOF, 3DOF translation only experiments and 6DOF freehand experiments.

There are several benefits to conducting experiments on a flat phantom before using a real hand.

First, the 3D surface model is created directly from the exact 2D image used to create the print on the surface of the phantom, and so no 3D scanning is required. Second, that image contains enlarged details of real palm features (approximately twice as large as their actual size), which facilitated testing of ProbeSight's tracking performance in these initial experiments before testing on an actual human. Third, compared with real human skin, the flat phantom is more stable and robust across multiple experiments. Once the relative pose of the phantom is fixed in the coordinates of the Micron Tracker, the coordinates of every pixel in that image can be computed. Any variation in lighting is largely due to the original conditions under which the photograph was taken. In summary, the flat phantom tests provided a convenient environment

similar in many important ways to the human forearm without some of the complexities described in the next section.
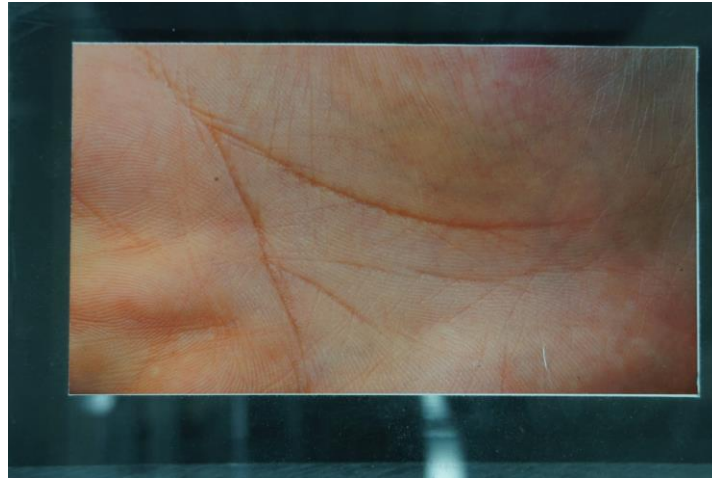
Figure 3-7: The flat phantom for ProbeSight tracking accuracy test in all 6-DOF simultaneously.

3.3   HAND MOLD AND SPLINT

Unlike the flat phantom experiment, it is difficult to maintain constant pose and image feature positions when experimenting with a real human hand. The many joints of the hand move, as may the overall pose of the hand. Furthermore, a prior 3D surface map is now required. In order to keep overall hand pose unchanged between the acquisition of that map and real-time ProbeSight studies, a thermoplastic dorsal blocking splint (NC-33965-3, North Coast Medical, Inc, CA) was shaped specifically for each subject's forearm. A custom hand-imprint mold was also produced using 3D printing, and then attached to the end of the splint. The hand mold was

26

initially shaped using Model Magic modeling compound (Crayola 57-4400, Pennsylvania, USA) and 3D scanned to provide the specifications for the 3D printer. The combination of the hand mold and arm splint immobilize the test subject's hand both in terms of overall location and joint pose, but the subject's skin still remained accessible, pliable, and deformable. Both the splint and optical tracker were rigidly attached to an optical table to secure their relative pose during different experimental procedures.

3.4 LIGHT STUDIO DEVELOPMENT

During the experiments, skin features were used to match images from the real-time ProbeSight camera against the prior high resolution 3D model. The accuracy of the resulting pose estimation was directly related to both motion blur and the signal-to-noise ratio (SNR) of the camera image. Accuracy was also degraded by specular reflections, caused by the glossy photographic paper in the phantom experiments and by ultrasound gel in the human experiments.

Motion blur can be mitigated using faster shutter speeds, but less light exposure reduces SNR, either by decreasing signal (darker image) or by increasing noise (higher gain for proper exposure). We compensated by using bright external lighting, diffused by silk diffusion fabric so that the influence of specular reflections would also be small. The resulting light frame, shown in Figure 3-8, was attached to the optical table by four steel threaded rods, with light-weight 4" x 36" balsa wood supporting the 60" x 36" diffusion fabric. Four external fluorescent lighting sources (1100 lumens, 2700K, and 120° wide beam angle) were applied to provide bright diffused lighting for the ProbeSight experiments.

Figure 3-8: The illustration of light studio system.

Figure 3-9 compares camera images of the flat phantom and a real hand under different illumination conditions. In Figure 3-9(A), without special lighting and diffuse fabric, the image contrast is low and only coarse creases on the palm is clearly visible. In Figure 3-9(B), with unfiltered external lighting applied, the fine creases are visible. However, without diffuse fabric the specular reflection on the phantom's surface causes glare, which degrades the image contrast as well. Reduced reflection and better contrast are achieved with both external lighting and diffuse fabric, as shown in Figure 3-9(C). In the hand experiment, the influence of US gel specular reflections can be reduced in the same manner, as Figure 3-9(D)(E) and (F) show. Figure 3-9(F) has the most features visible, especially under the gel.
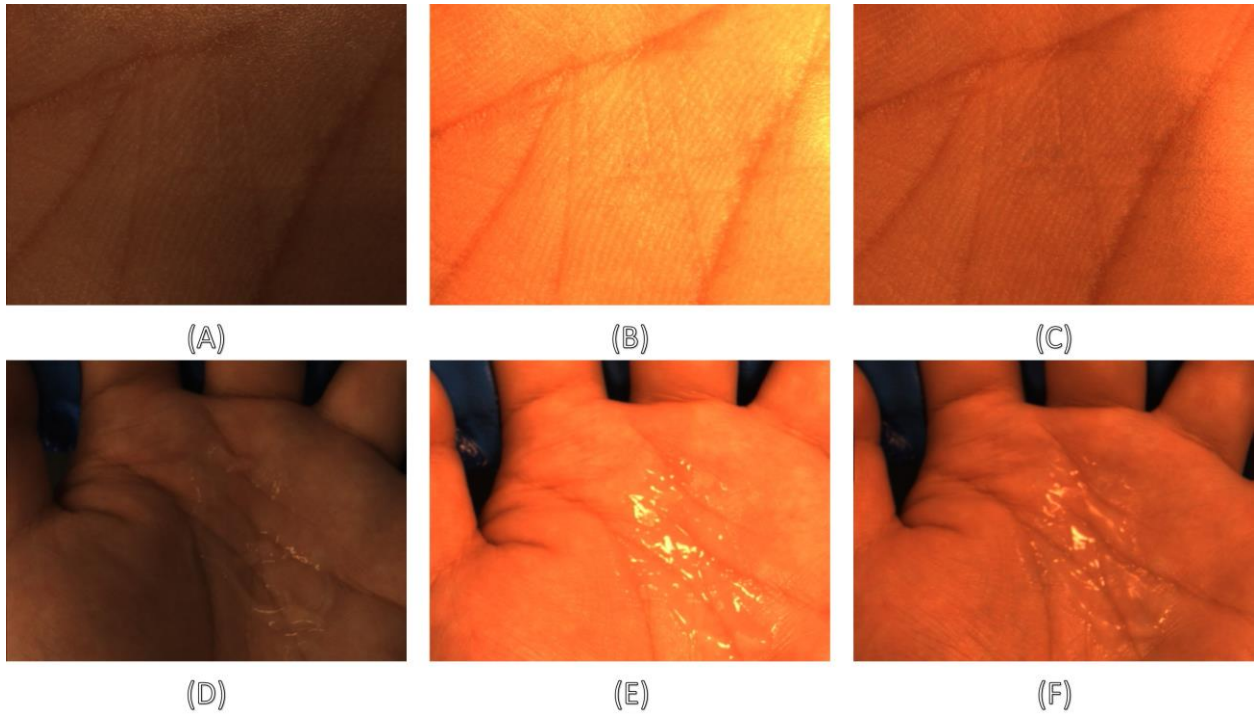
Figure 3-9: Upper row: A camera image of the phantom **(A)** with external lighting and diffusion fabric off, **(B)** with external lighting on and diffusion fabric off, and **(C)** with external lighting and diffusion fabric both on. **Lower row:** The camera image of real hand.

4.0   TECHNICAL APPROACHES

This chapter describes the underlying technical approaches employed in the experiments described later in Chapters 5.0 and 6.0 . We include these methods first, to lay the foundation for the actual experiments in which they are used.

4.1   SOFTWARE DEVELOPMENT OVERVIEW

Before discussing the underlying technical approaches, the infrastructure of software development is first described here. Figure 4-1 shows a flowchart of the overall ProbeSight system, which can be divided into three modules. When ProbeSight starts, the system enters the *data initialization module*, which captures input data from three different sources that "sample" the real world, including an initial camera image, current orientation of the probe reported by INS, and initial ground-truth pose tracking data, which is directly read from the Micron Tracker. The purple boxes in the upper right show the procedures of 3D model acquisition and camera calibration, which must be done before the system starts. By knowing the pre-acquired 3D model and pre-calibrated camera intrinsic parameters (camera matrix and distortion coefficients), we are able to re-project the 3D model and simulate what would be seen by the real ProbeSight camera at the initial pose. The probe is kept immobile during the whole initialization procedure.

When the probe moves to a new location and orientation, the *data updating module* captures the updated camera frame and new orientation data is read from the INS, which is used to compute the orientation change between the previous and current probe pose. Both the camera image and simulated re-projected image are then processed to reduce noise and improve image contrast for feature enhancement.

These data and images are then processed by the *data optimization module*. The core of this module is an OpenDR pipeline, which will be discussed in section 4.5.3. The pipeline is built to explicitly relate the updates of camera images with changes in the model parameters, allowing us to estimate motion of the camera and attached ultrasound probe. The estimated motion is iteratively applied to drive the 3D model to match the updated camera image, generating an updated simulated image. The estimated motion is also compared with ground truth data provided by the Micron Tracker for error analysis. The system then waits for the next camera frame to trigger the data updating module again, looping until the procedure is halted.

Both the data initialization and updating module were written in C++ (Visual Studio 2010), using several libraries, including the Open Graphics Library (OpenGL), Open Source Computer Vision (OpenCV) library, Micron Tracker Software Development Kit (SDK), and the VectorNav INS data export SDK. Inverse rendering and optimization were written in Python (Version 2.7.6), using both the OpenDR and Qt graphical user interface (GUI) framework.
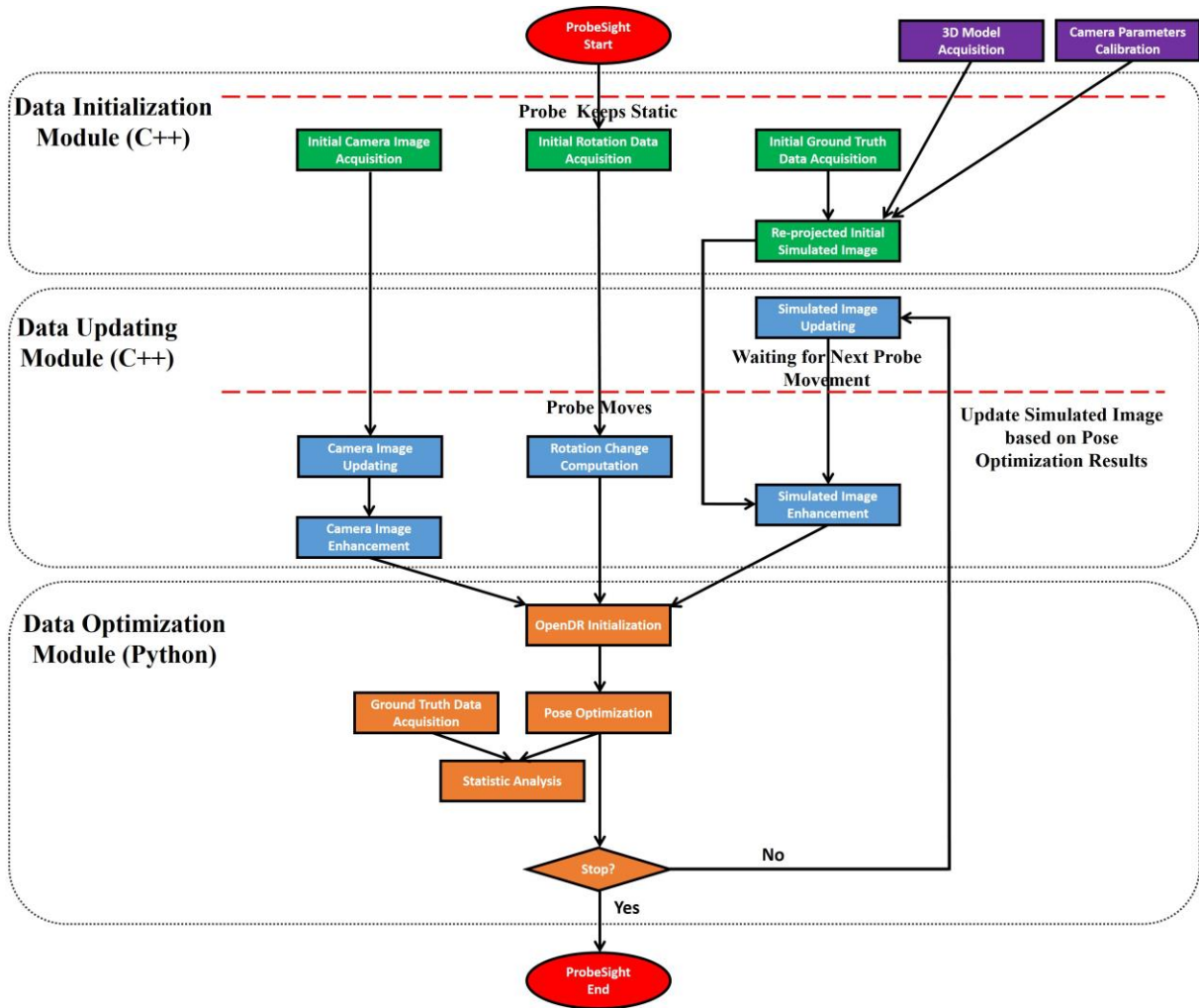
Figure 4-1: The flowchart of software development for ProbeSight system.

4.2   THE PRE-SCAN VECTRA M3 SYSTEM

ProbeSight requires an *a-priori* 3D surface map of the patient's anatomy in the region where the ultrasound will be performed. We currently acquire this 3D surface map using a clinical 3D imaging system (Vectra M3; Canfield Scientific Inc, US), whose 3 stereo pairs produce a composite 36-megapixel scan. The resulting surface map, computed using stereo matching, contains not just the surface location but the corresponding fine details of color and intensity, including skin creases, skinfolds and hairs, as shown in Figure 4-2. Reprojecting this surface map from a given location and orientation, with knowledge of the optical properties of the particular camera, yields a simulation of what the camera would see from that viewpoint. Comparing this simulated image with the actual camera image permits us to search for the correct viewpoint based on the current camera image.

Figure 4-2: The 3D reconstruction map of upper extremity acquired by Vectra M3 system, as viewed from up close to the palm.

4.3   AN OVERVIEW OF COORDINATE SYSTEMS AND TRANSFORMS

Coordinate transforms are central to the research of this dissertation.   Each component of ProbeSight has its own native coordinate system, requiring transforms between camera coordinates C_C and those of each of the other components.

Since the INS is mounted on the camera, its rotations are inherently in camera coordinates C_C. We can ignore the fixed translation between the INS and the camera, since we are only using the INS's rotation information.   Other important coordinate systems include anatomy (prior 3D scan) C_A, validation (Micron) tracker C_T, and the validation optical marker attached to ProbeSight C_M.   In order to provide simulated camera images of the 3D surface map to the OpenDR pipeline, all coordinates need to be transformed from anatomy coordinates C_A to camera coordinates C_C.   These coordinates and transforms are summarized in Table 4-1 and Table 4-2.

Table 4-1: List of coordinate systems

Coordinate Systems Abbrev.	Coordinate Systems Name	Coordinate Systems Unit
C_A	**A**natomical coordinates	mm
C_T	**T**racker coordinates	mm
C_M	**M**arker coordinates	mm
C_C	**C**amera coordinates	mm

Table 4-2: List of transforms

Transform Abbrev.	Transform Name
$T_{A \rightarrow T}$	Transform from **A**natomical to **T**racker coordinates
$T_{T \rightarrow M}$	Transform from **T**racker to **M**arker coordinates
$T_{M \rightarrow C}$	Transform from **M**arker to **C**amera coordinates

The transform flow for calibration and validation is illustrated using solid arrows in Figure 4-3(A). First, the 3D scan of the anatomy must be transformed into the common coordinate system of the Micron Tracker; this transform is denoted $T_{A \rightarrow T}$. During both calibration and validation, the Micron Tracker provides a real-time updated matrix M that transforms C_M to C_T, this transform being denoted $T_{M \rightarrow T}$. In practice, we need the inverse transform $M^{-1} = T_{T \rightarrow M}$. The final transform from C_M to C_C is a simple fixed-translation transform $T_{M \rightarrow C}$.

For the flat phantom experiment, the 3D model is created directly from a 2D printed image of skin attached to the flat top surface of an acrylic block. Since the underlying image is already known, no scanning is required. The 3D pose of the phantom's 2D image is measured by touching each corner of the printed image with a tracked tooltip, as shown in Figure 4-3(B). The 3D coordinates p_T of these corners are already in coordinate system C_T, and so no transform $T_{A \rightarrow T}$ is required for the phantom. To fully transform p_T into the camera's coordinates, we then use $p_C = T_{M \rightarrow C} T_{T \rightarrow M} p_T$.

For the dinosaur phantom or in-vivo human hand experiment, the 3D surface map from the Vectra preoperative 3D scanner serves as the model of the anatomy. As described in section

3.3, a custom 3D-printed, molded hand splint specific to the subject is used to keep the hand and finger pose constant between the prior scan and the experiment, as shown in Figure 4-3(A). The 3D coordinates of the exterior anatomy p_A are originally in the preoperative 3D scanner's coordinates C_A. For validation purposes, we perform a one-time transformation from C_A to C_T, denoted $T_{A \to T}$, which requires touching white-dot fiducials on the dinosaur model or hand splint with a tracked tooltip, analogous to the phantom experiment. To fully transform p_A into the camera's coordinates, we use $p_C = T_{M \to C} T_{T \to M} T_{A \to T} p_A$. The technical details of these transforms will be discussed in next section.

(A)

(B)
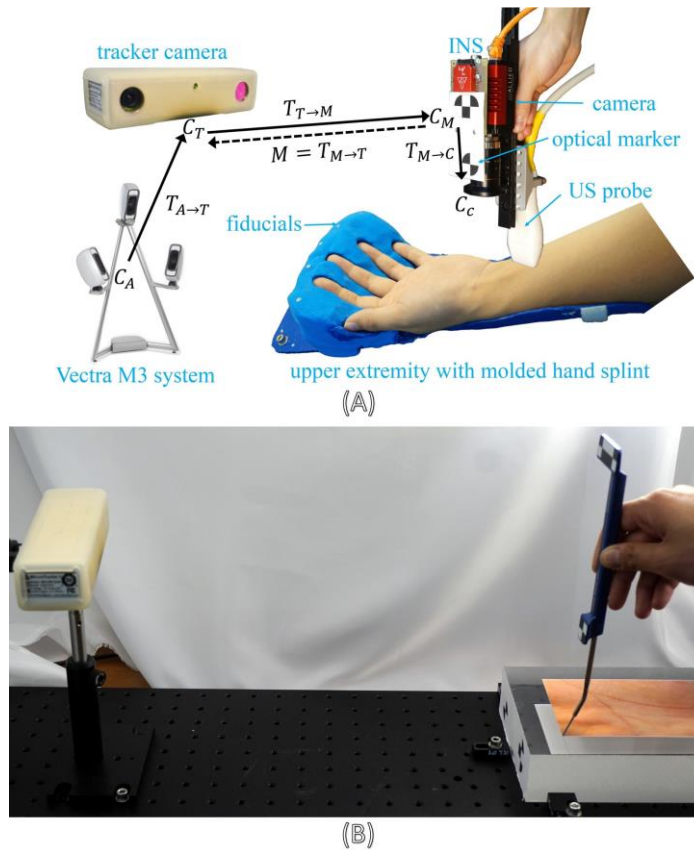
Figure 4-3: (A) Relationships among the coordinate systems of different devices. **(B)** The 3D pose of the phantom's 2D image is measured by touching each corner of the phantom's image with a tracked tooltip.

36

4.4 RENDERING THE PRIOR 3D SURFACE MAP

Starting with the pre-acquired 3D surface model discussed in Section 3.2.1, we are able to render projections from any viewpoint, searching for that viewpoint that matches the camera's actual view, and thus determine the camera's pose relative to the external anatomy. This procedure requires more effort than typical of graphical rendering for entertainment or visualization purposes. Because we are matching the image from an actual camera, the actual camera optics must be modeled accurately. The location of the camera's entrance pupil must be correctly determined and fixed geometric transforms for the camera need to be pre-computed.

As summarized in section 4.3, multiple transforms need to be performed to convert the model's 3D coordinates from its original preoperative 3D scanner's coordinates C_A to camera coordinates C_C. We discuss now in greater detail how these transforms are empirically determined.

4.4.1 Finding the transform between anatomical coordinates and tracker coordinates

To find the transform $T_{A \to T}$ between anatomical coordinate system C_A and tracker coordinate system C_T, we painted a set of white dot fiducials on the hand splint. After the preoperative 3D scan is done, we are able to manually identify the fiducials' 3D coordinates $p_{A(fid)}$ in anatomical coordinates C_A by importing data to MeshLab (open source software for processing and editing unstructured 3D triangular meshes).

37
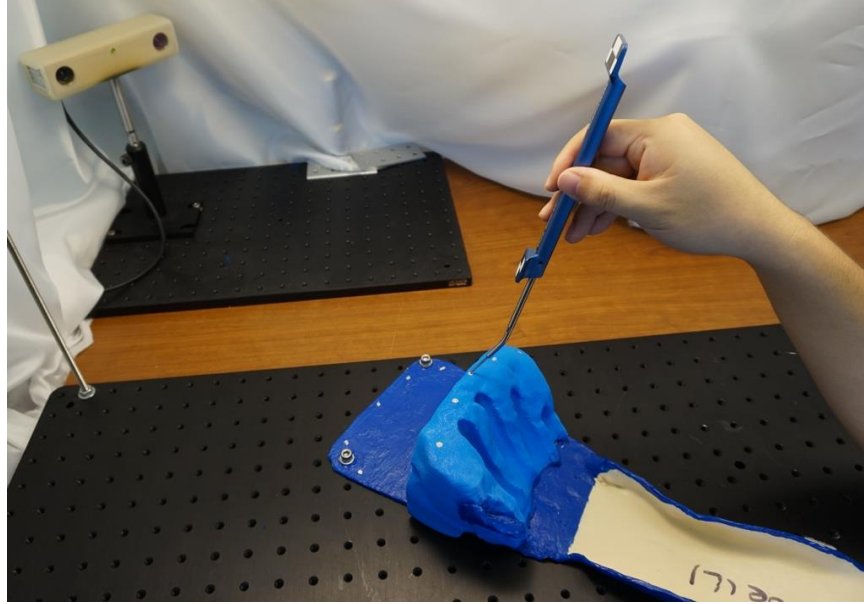
Figure 4-4: An example of fiducial identification in tracker coordinate using a tracked tool.

The optical tracker is then used to localize these fiducials in its own tracker coordinate system C_T by using the tip of the tracked tool (provided by ClaroNav with the Micron Tracker) to touch each of the fiducials successively. This yields the fiducial's coordinates in coordinate system C_T, which we denote $p_{T(fid)}$. The procedure for one of the fiducials in the in-vivo hand experiment is shown in Figure 4-4. $p_{A(fid)}$ and $p_{T(fid)}$ can be expressed as matrices:

$$p_{A(fid)} = \begin{bmatrix} x_{A1} & x_{A2} & x_{An} \\ y_{A1} & y_{A2} \ldots y_{An} \\ z_{A1} & z_{A2} & z_{An} \\ 1 & 1 & 1 \end{bmatrix} \qquad \textbf{Equation 4-1}$$

$$p_{T(fid)} = \begin{bmatrix} x_{T1} & x_{T2} & x_{Tn} \\ y_{T1} & y_{T2} \ldots y_{Tn} \\ z_{T1} & z_{T2} & z_{Tn} \\ 1 & 1 & 1 \end{bmatrix} \qquad \textbf{Equation 4-2}$$

where n represents the number of fiducials.

If the affine transform $T_{A\rightarrow T}$ is defined by a 4-by-4 matrix as,

$$T_{A\rightarrow T} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 4-3

we have, $p_{T(fid)} = T_{A\rightarrow T} p_{A(fid)}$, There are 12 unknowns in transform $T_{A\rightarrow T}$ of the equation so we need at least four fiducials to solve the equations. If more than four fiducials' coordinates are known, a random sample consensus (RANSAC) algorithm is applied to make the estimation result more robust (Fischler and Bolles 1981). The RANSAC algorithm returns with the estimated transform parameters that lead to the largest number of inliers, which estimates the affine transform with a high degree of accuracy. This computed transform is then applied to all points of the 3D model in anatomical coordinates p_A to transform them into tracker coordinates p_T, which can be written as, $p_T = T_{A\rightarrow T} p_A$.

Once a molded hand splint has been produced, the procedure of determining $T_{A\rightarrow T}$ needs to be performed only once for that particular splint's fiducial pattern, and it can be done without the patient present.

4.4.2 Find the transform between tracker coordinates and marker coordinates

To find the transform $T_{T\rightarrow M}$ between tracker coordinates C_T and marker coordinates C_M, we need the inverse transform of the onboard marker's pose M. This pose is tracked by the optical tracker at the probe's initial pose and can be expressed by a 4-by-4 matrix as,

$$M = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 4-4

Then, we apply $T_{T \rightarrow M} = M^{-1}$ to transform all coordinates p_T in the tracker's coordinates

into the marker's coordinates p_M, which can be written as, $p_M = M^{-1} p_T$.

For the present work, we are temporarily using the optical tracker to provide pose M to

quickly initialize ProbeSight's pose for the first frame only. In the long run, we will develop

automatic initialization technology and no longer need the optical tracker, making ProbeSight a

fully self-contained device. This will be further discussed in Chapter 8.0 .

4.4.3 Find the transform between marker coordinates and camera coordinates, part 1

The final transform from marker coordinates C_M to camera coordinates C_C is a fixed translation

$T_{M \rightarrow C}$, which can be computed prior to the experiment. Two steps are involved in this

procedure, including finding translation T_1 from the camera marker's center to the camera front

flange's center and translation T_2 from the flange's center to camera's entrance pupil. We

describe the first step in this section.

There is a rigid transform in space between the marker's center and the camera-lens front

flange's center. Figure 4-5 shows the method for accurately measuring this transform with the

assistance of the optical tracker. As Figure 4-5(A) shows, a circular marker which has the same

diameter as the camera's front flange was created and pasted on an aluminum board. The optical

tracker was mounted on a tripod, facing down to track the location of the circular marker's center

P_1, as

$$p_1 = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Equation 4-5
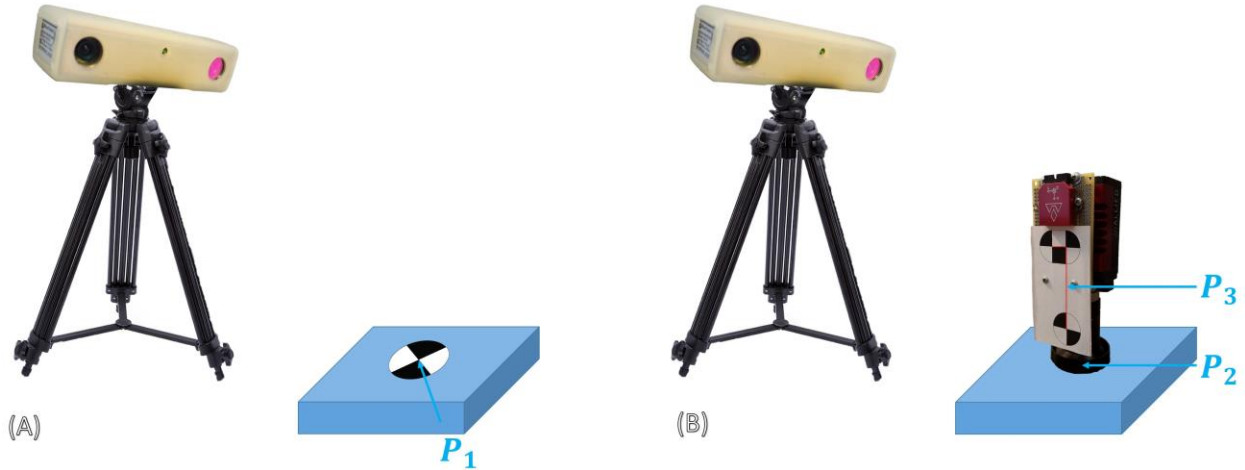
Figure 4-5: (A) Illustration of camera's front flange center detection using Micron optical tracker. **(B)** Illustration of camera marker center and camera's front flange center detection using Micron optical tracker. The marker's center is localized at the center of the long vector (discussed in section 3.1), represents in long red line on the figure.

The camera was then placed on an aluminum board so that its front flange just covered the circular marker, shown in Figure 4-5(B). If the center of camera's front flange is defined as P_2, we have $P_2 = P_1$. We slightly rotate the camera around the center of the front flange until its onboard marker can be tracked by the optical tracker. We define the location of onboard marker's center P_3 as,

$$p_3 = \begin{bmatrix} t_x' \\ t_y' \\ t_z' \end{bmatrix}$$

Equation 4-6

The translation T_1 from marker's center to the camera's front flange's center can be computed by subtracting point P_3 from P_2, denoted as:

$$T_1 = p_3 - p_2 = \begin{bmatrix} t_x' - t_x \\ t_y' - t_y \\ t_z' - t_z \end{bmatrix}$$

Equation 4-7

As opposed to measuring the transform by using a caliper, we don't need to manually identify the marker's center and front flange's center, which makes this method more accurate. Moreover, using a caliper might damage the lens, so our method is safer for the camera.

4.4.4   Find the transform to camera coordinates, part 2: entrance pupil

Having measured the transform T_1 from the marker's center to the camera's front flange center, we need an additional translation T_2 to subsequently transform to camera coordinates. Before discussing this transform, it is important to understand the concept of a camera entrance pupil and correctly determine it.

The entrance pupil is the point along the optical axis about which the camera can be rotated without changing the relative pixel alignment between the objects at different distances. For the simplest pinhole camera model, exactly one ray of light from a point in object space will pass through the camera's pinhole to form a corresponding point in the image plane. With a real lens however, light from a point in object space is collected from a solid angle of rays and projected through the lens onto the image plane, as illustrated in Figure 4-6. The extent of this solid angle of rays is limited by the lens elements and by the diameter of any diaphragms along the optical path. The limiting diaphragm is called the aperture stop of the lens. The entrance

pupil of the lens is the image of the aperture stop as it would be seen if viewed from an axial position in front of the lens (Willson et al. 2005).
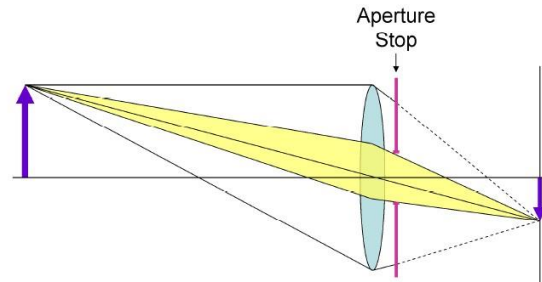
Figure 4-6: Image formation for a simple lens (Willson et al. 2005).

Our strategy to find the entrance pupil is based on the first concept described above. We point the camera toward two objects so that they perfectly overlap when the camera is facing directly toward them. The camera is then rotated in the horizontal plane, as the left column of Figure 4-7 shows. If the entrance pupil is not the same as the rotation point, the two objects will no longer overlap when the camera is rotated (Middle column of Figure 4-7). We then iteratively move the camera on a slider until we find the location about which, when the camera is rotated, the two objects always overlap (Right column of Figure 4-7).
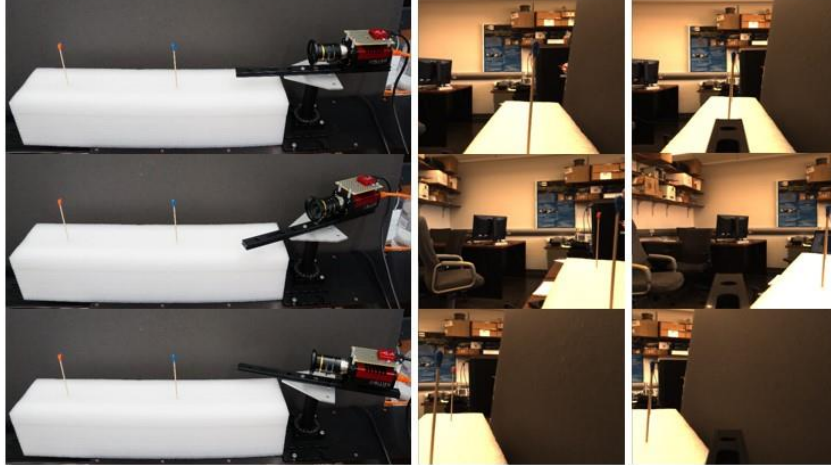
Figure 4-7: Left column: Experimental setup to locate entrance pupil. **Middle column:** Example of misalignment between pivot point and entrance pupil; both sticks are visible in 2nd and 3rd rows. **Right column:** Entrance pupil is aligned, and so the front stick always occludes the back stick.

As Figure 4-8 shows, in the aligned case, the entrance pupil of camera P_4 is aligned vertically with the rotating point P_5. Although it is difficult to measure the distance d from camera's front flange P_2 to the entrance pupil P_4, we can easily measure the distance which equals to d on the slider rail from point P_6 which aligned vertically with P_2 to the rotating point P_5 using calipers.

Both P_2 and P_4 lie on the optical axis (red line in Figure 4-8), which overlaps with the y axis in C_M. The translation T_2 from the center of the camera's front flange to the camera's entrance pupil can be expressed as,

$$T_2 = p_4 - p_2 = p_6 - p_5 = \begin{bmatrix} 0 \\ d \\ 0 \end{bmatrix} \qquad \textbf{Equation 4-8}$$

The combination of translation T_1 (computed in the previous section) with T_2 provides the complete transformation from marker coordinates C_M to camera coordinates C_C. Since these are both pure translation matrices, the total translation of $T_{M \to C}$ can be represented as
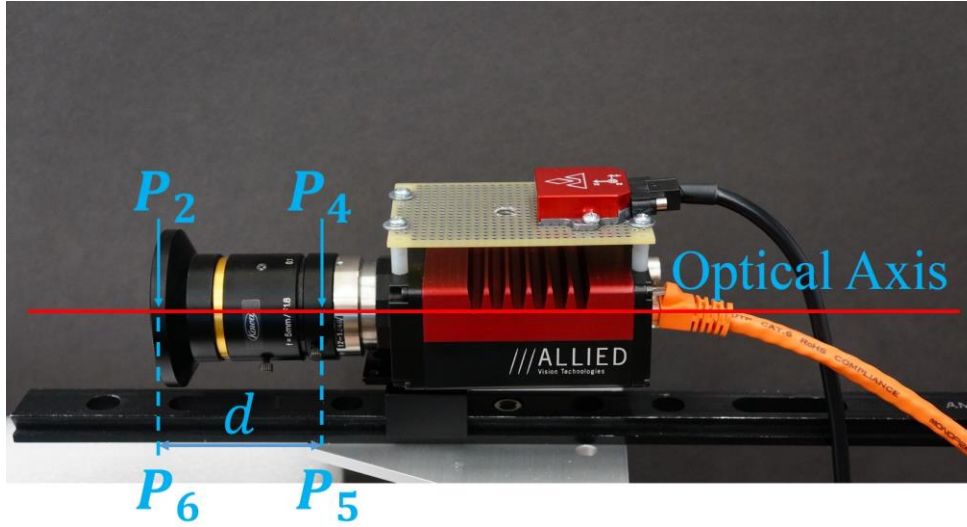
Figure 4-8: Illustration of the measurement from camera-lens front flange to the entrance pupil.

$$T_1 + T_2 = \begin{bmatrix} t_x^{'} - t_x \\ t_y^{'} - t_y + \mathrm{d} \\ t_z^{'} - t_z \end{bmatrix}$$

Equation 4-9

representing this translation equation as a 4 x 4 matrix using homogenous coordinates, we get,

$$T_{M \to C} = \begin{bmatrix} 1 & 0 & 0 & t_x^{'} - t_x \\ 0 & 1 & 0 & t_y^{'} - t_y + d \\ 0 & 0 & 1 & t_z^{'} - t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 4-10

Once we calculate all transforms including $T_{A\rightarrow T}$, $T_{T\rightarrow M}$ and $T_{M\rightarrow C}$, the 3D points P_A in the initial anatomic coordinates can be transformed to the final camera coordinates as P_C using the equation,

$$P_C = T_{M\rightarrow C} T_{T\rightarrow M} T_{A\rightarrow T} P_A \qquad \textbf{Equation 4-11}$$

substituting all pre-computed matrices in, we have,

$$
\begin{bmatrix} x_C \\ y_C \\ z_C \\ 1 \end{bmatrix} =
\begin{bmatrix} 1 & 0 & 0 & t_x' - t_x \\ 0 & 1 & 0 & t_y' - t_y + d \\ 0 & 0 & 1 & t_z' - t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} x_A \\ y_A \\ z_A \\ 1 \end{bmatrix}
\qquad \textbf{Equation 4-12}
$$

Since these matrices are all precomputed, they can be combined into a single precomputed 4 x 4 matrix, including rotation R and translation T,

$$
\begin{bmatrix} x_C \\ y_C \\ z_C \\ 1 \end{bmatrix} =
\begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} x_A \\ y_A \\ z_A \\ 1 \end{bmatrix}
\qquad \textbf{Equation 4-13}
$$

4.4.5 Camera models and calibration

In the previous sections, we discussed the procedure to find R and T, which denote the coordinate system transformations from 3D world coordinates (anatomical) to 3D camera coordinates. In computer vision, R and T are defined as extrinsic parameters. In order to further project the 3D camera coordinates to camera's 2D image plane, we need the intrinsic parameters which describe the properties of the camera. This can be done by applying standard methods of camera calibration to obtain the intrinsic parameters in two parts, which are a model of the camera's geometry and a distortion model of the lens. The camera's geometry, including the

46

focal lengths (f_x and f_y) and principal points (c_x, and c_y), determine the homogeneous projection matrix K for the perspective-projection camera model:

$$K = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Equation 4-14

parameter γ represents the skew coefficient between x and y axes, and for the present work it is assumed to be zero (empirically, assuming $\gamma = 0$ seems to work well).

Distortions are inherent in any real lens design, and we model them as separate polynomial expansions in the radial and tangential directions.

Radial distortion arises because the lens behaves differently at the center of the image than at the periphery, resulting in "barrel" or "pincushion" distortion, as Figure 4-9 shows. We characterize this by a Taylor series expansion along lens radius r, which represents the distance from the image center. For our current high-quality optical lenses that were designed for computer vision, we generally require only the first 2 terms in our Taylor expansion, which are conventionally termed k_1 and k_2. For highly distorted optics such as fish-eye lenses it may be necessary to use a third radial distortion term k_3 (J.G. Fryer; D.C. Brown 1986). Because of assumed symmetry across the optical axis, only the even powers in the expansion are required. Any location (x, y) on the image sensor can thus be corrected according to the following equations,

$$x_{corrected} = x\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right), \quad y_{corrected} = y\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right)$$

Equation 4-15
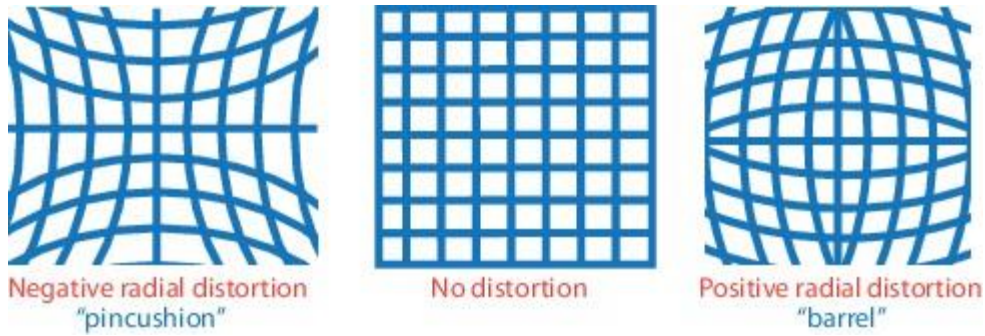
Figure 4-9: Illustration of radial distortions.

Tangential geometric distortion arises from imprecision during the manufacture of the camera resulting in the lens not being exactly parallel to the imaging plane as Figure 4-10 shows. This can be minimally characterized by two parameters, p1 and p2, as follows (Brown 1966).

$$x_{corrected} = x + \left[2p_1 xy + p_2 \left(r^2 + 2x^2 \right) \right], \; y_{corrected} = y + \left[2p_2 xy + p_1 \left(r^2 + 2y^2 \right) \right] \qquad \textbf{Equation 4-16}$$
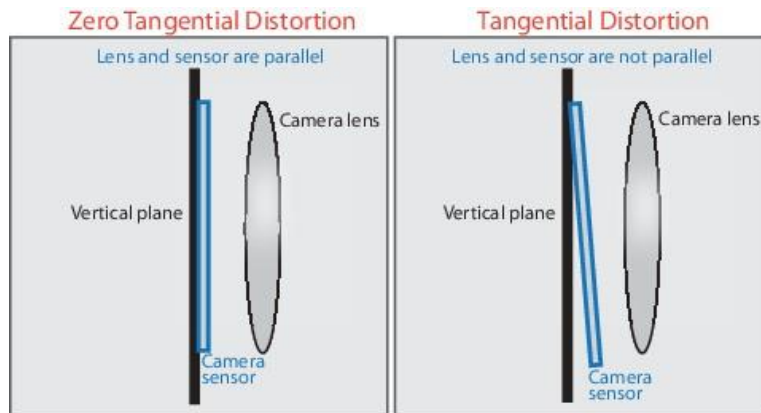
Figure 4-10: Illustration of tangential distortion.

For ProbeSight, it is essential to account for these distortions, since "up-close" (near focused) imaging accentuates them.

We used Matlab (The Mathworks Inc, Natic, MA, USA) to estimate all these parameters using standard calibration methods, which were originally derived from Bouguet (Bouguet 1999). During the calibration process, the camera was targeted on a known structure that had individual and identifiable points, namely, the corner points of squares in a black-and-white chessboard pattern. The algorithm computes relative location and orientation of the camera for each image as well as the intrinsic parameters of the camera while viewing a chessboard pattern from a variety of angles. The chessboard is repeatedly rotated and translated in order to provide multiple viewpoints in the process of calibration. 20 camera images were captured at varying camera views in this calibration, to provide sufficient constraints for the solution.

Once the intrinsic parameters of the camera are correctly computed, the perspective projection of the model from 3D camera coordinates to the 2D image plane (used for image rendering) can be obtained as:

$$
\begin{bmatrix} \hat{x}_c \\ \hat{y}_c \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}
\qquad \textbf{Equation 4-17}
$$

where (x_c, y_c, z_c) is the corrected location, which is first calibrated by radial coefficients, then calibrated by tangential coefficients in Matlab implementation.

4.5 A REVIEW OF THE METHODS FOR MATCHING A CAMERA IMAGE TO THE 3D MODEL

With the establishment of the coordinate transforms and camera calibration discussed in the previous section, ProbeSight is able to re-project the 3D surface map to simulate what would be seen at the initial camera's pose. When the US probe moves in space, its on-board camera image (observed image) is continually updated. If we can match the (simulated camera image of the) 3D surface map with the updated camera image, then calculating the camera pose change between two consecutive frames will give us the probe's motion. Matching two images requires the definition of a metric by which the match between them may be quantified, and thus optimized. In this section, different image matching algorithms are reviewed. All of these algorithms are applied in different phases of our ProbeSight experiments.

4.5.1 Mutual information based image matching

Mutual information (MI) is a popular metric used for image matching. It is based on the information-theory concept of entropy, and can be considered a measure of the statistical dependency between the images (Viola and Wells 1995). This metric requires the computation of joint histograms, with a 2D domain of bins, as shown in Figure 4-11. The value n_{ij} is the number of pixels with the intensity value i in the first image and intensity j in the second image. This indicates the frequency that we see this pair of intensities at the same corresponding locations. For example, if the joint histogram count is $n_{ij} = 3$, where $i = 6$, and $j = 5$, this means that, at 3 separate locations in both images, when we encountered the intensity of 6 at that

location in the first image, we encountered the intensity of 5 at the same location in the second image. All entries of the joint histogram are actually normalized by the total number of elements in either image (should be the same size), which gives the joint probability distribution between both images denoted as P_d.
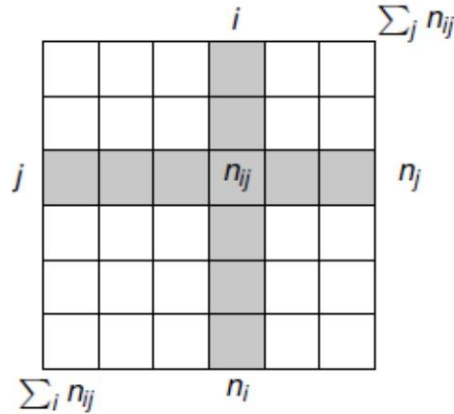
Figure 4-11: Illustration of the joint histogram between two images (Ghannam and Abbott 2013).

To define the MI between two images, we regard them as random variables X and Y and their intensity values at certain coordinates in the images as the joint outcome of an unknown underlying random process. The measurement of matching information is based on Shannon Wiener entropy theory, which defines MI in terms of entropy as follows:

$$MI(X,Y) = H_e(Y) - H(Y \mid X) = H_e(X) + H_e(Y) - H_e(X,Y) \qquad \textbf{Equation 4-18}$$

where $H_e(X) = -\sum_X P_d(X) \log(P_d(X))$ represents the entropy of random variable X and $P_d(X)$ is the probability density, as estimated by the histogram of the single image. $H_e(X,Y) = -\sum_X \sum_Y P_d(X,Y) \log(P_d(X,Y))$ represents the joint entropy between the two

random variables X and Y and $P_d(X,Y)$ is the probability density in the joint histogram (Ghannam and Abbott 2013).

In information theory, entropy measures the uncertainty inherent in the distribution of a random variable. Joint entropy extends that measure to the uncertainty in the joint distribution of a pair of random variables. When the two images are best matched, there is the smallest uncertainty in their joint distribution, which makes the smallest $H_e(X,Y)$. Since the entropy of each image $H_e(X)$ and $H_e(Y)$ is fixed, that indicates a maximum mutual information value. In general, the mutual information value is a measurement of the amount of information that one random variable contains about another random variable. The larger the value, the less uncertainty exists between the two variables, which implies greater similarity between the underlying information in two images, even if the mappings between that information and intensity value differs between the two imaging modalities.

In medical imaging, the use of mutual information in image registration for different modalities has yielded excellent results (e.g. CT to MRI, MRI to PET etc.) (Maes et al. 1997; Studholme, Hill, and Hawkes 1997; Wells 1996). The aim of our purpose is to match the 2D real-time camera image with a given projection of the 3D surface, which is, in essence, a form of image registration. We found mutual information to be more reliable compared with other metrics such as Sum of Absolute Differences (SAD) and Normalized Cross Correlation (NCC), since mutual information can accommodate the differences in color, background and illumination, inherent to the different conditions during the pre-scan and the real-time operation of ProbeSight. Therefore, the mutual information metric was used to validate tracking accuracy in each degree of freedom on the dinosaur model experiment.

4.5.2   Ridge matching based on voting and maximal correlation in transform space

In our experiments using a real human upper extremity, we explored a different method of matching images that does not require iterative optimization. My lab colleague Mingzhi Tian and I implemented and tested this approach to find the optimal 2D transform between two images (2 translations and 1 rotation), although in theory the technique could be extended to higher dimension transforms. The goal was find robust features whose numbers were few enough to find an optimum transform between the underlying images in a single step. The step consists of voting in transform space for every possible pair of features between the images. The features we chose are points along the fine ridges in the skin, which are stable anatomical structures fairly resistant to changes in illumination, in the sense that they produce corresponding ridges in intensity in the image. The ridge points each have a location and orientation, so matching a pair between two images implies a 2D transform between the images. We employed three steps to match these ridge points in space, described in the following three sections.

4.5.2.1 Ridge point detection and representation

The ridge detection problem has been intensively studied by Pizer and his colleagues (Stephen M. Pizer et al. 1994). Gauch and Pizer define ridges from topographical watersheds computed in a scale-space representation of the image data (Gauch and Pizer 1993). Morse computes "core" descriptors in a multi-scale fashion by propagating a measure of edge strength from each edge point and then detecting peaks in the measure of "medialness" so obtained (Morse, Pizer, and Liu 1994). A more extensive discussion of different types of ridge detectors is presented by Eberly, including extensions to higher dimensions (Eberly et al. 1994). Lindeberg put forward a

novel concept of a ridgeline detection algorithm, defined as a connected set of points. We adopted Lindeberg's approach, as described next:

To identify ridge points, a convenient framework is to define the feature descriptors in terms of local directional derivatives relative to a preferred local coordinate system. A natural way to construct such a system suitable for ridge detection is as follows: For any image point (x_0, y_0) in Cartesian coordinates, a local (p,q) system aligned to the principal curvature directions of the brightness function is introduced. To express directional derivatives in these coordinates, which are characterized by the mixed second-order derivative being zero, $L_{pq} = 0$, we rotate the coordinate system by an angle β defined by,

$$\cos \beta \Big|_{(x_0,y_0)} = \sqrt{\frac{1}{2}(1 + \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}})} \Big|_{x_0,y_0}$$

$$\sin \beta \Big|_{(x_0,y_0)} = \text{sgn}(L_{xy}) \sqrt{\frac{1}{2}(1 - \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}})} \Big|_{x_0,y_0} \qquad \textbf{Equation 4-19}$$

Directional derivatives in this local (p,q) system are related to partial derivatives in the Cartesian coordinate system by

$$\partial_p = \sin \beta \partial_x - \cos \beta \partial_y , \ \partial_q = \cos \beta \partial_x + \sin \beta \partial_y \qquad \textbf{Equation 4-20}$$

The bright (or dark) ridge can then be defined in terms of the local (p,q) system as a connected set of points for which the intensity assumes a local maximum (minimum) in the direction of the main principal curvature. This requirement for the point to be a bright ridge point can be written as

$$L_p = 0$$
$$L_{pp} < 0 \qquad \textbf{Equation 4-21}$$
$$|L_{pp}| > |L_{qq}|$$

where L_{pp} and L_{qq} are the eigenvalues of the second-order derivative matrix.

By doing this, the original image is reduced to a set of ridge points represented as the matrix S below, in which each ridge point contains its Cartesian coordinates (x_0, y_0) and its orientation along the ridge line, defined as β (Tian et al. 2016).

$$S = \begin{bmatrix} x_1 & y_1 & \beta_1 \\ x_2 & y_2 & \beta_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & \beta_n \end{bmatrix}$$

Equation 4-22

The resulting matrix S is used as input to the ridge matching algorithm, described in the next section.

4.5.2.2 Transformation representation using voting based method

Our lab used the matrix S from each image containing ridge points to find the best transformation between the two images, using a voting method described next. The transformation T_{ij} between each ridge point i represented as (x_i, y_i, β_i) in image I_1 and each ridge point j represented as (x_j, y_j, β_j) in image I_2 can be computed as the equation shows below,

$$T_{i,j} = \begin{bmatrix} \Delta x_{ij} \\ \Delta y_{ij} \\ \Delta \theta_{ij} \end{bmatrix} = \begin{bmatrix} x_j \\ y_j \\ \theta_j \end{bmatrix} - \begin{bmatrix} \cos \Delta \theta & -\sin \Delta \theta & 0 \\ \sin \Delta \theta & \cos \Delta \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ \theta_i \end{bmatrix}$$

Equation 4-23

Assuming there are a total of n_1 and n_2 ridge points correspondingly in the two images, there will be a total of $n_1 \times n_2$ pairs, each implying a particular rigid 2D transform. Thus each pair can be represented as a location in a 3D $x \times y \times \beta$ transform space K. When a large number of ridge point pairs are mapped to approximately the same location, this particular location

represents the transformation that is most likely to be true. In other words, compared with other possible transforms, applying this transform will cause the most ridge points in one image to match the corresponding ridge points in another image.

4.5.2.3 Ridge matching based on maximal correlation

In practice, ridge point pairs will not be mapped at exactly the same location in transform space due to image noise and other variations. However, when a good match exists between two images, the point cloud will tend to form a dense cluster near an optimum location in the transform space K. Therefore, a measurement of point cloud density at each location in K reflects how the two images correlate with each other under the corresponding transform. Our lab compute a density function by convolving the point in transform space K with a blurring kernel f, whose value is one in the cuboidal region of size $1{\times}1{\times}0.2$ centered at (0, 0, 0) and whose value is 0 elsewhere. The optimal transformation between the two images is determined to be the location in transform space K of the global maximum of this convolution (Tian et al. 2016).

Figure 4-12 shows the ridge line detection results on an enhanced image of the palm (Figure 4-12(A)). The scale space pyramid concept is applied here by convoluting the original image with different scales Gaussian kernels, which will be discussed further in section 4.5.3.3. The ridge detection results at various scales are shown from Figure 4-12(B) to (D), with the ridge line maps labeled using white overlaid on the original image. Notably, different types of image structures give rise to different ridges curves. At smaller scales, such as shown in Figure 4-12(B), numerous ridges reflect the details of fine creases, which could benefit the image matching procedure, but which suffer from greater noise and longer computation time. Conversely, at larger scales, such as shown in Figure 4-12(D) fewer ridges are evident, and are

relatively coarse, representing general shapes. These may not be as robust for precise image matching, but require less computation time. For our purposes, we tended to choose a midlevel scale, as shown in in Figure 4-12(C), which contains ample ridges for image matching, but can achieve efficiency for real-time computation as well.
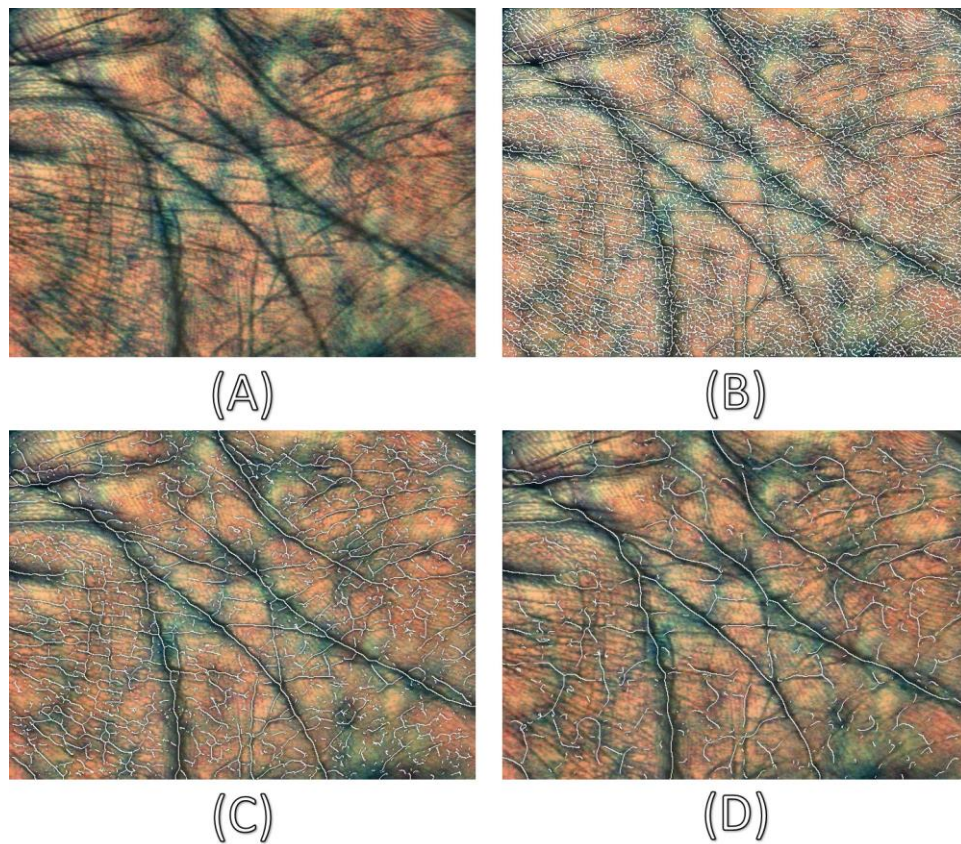
Figure 4-12: (A) The enhanced camera image of the palm. **(B)-(D)** The enhanced camera image of the palm with both palm and subdermal ridges computed at scale level 1, 3 and 6.

Transform space ridge matching provides a non-iterative way to find the transform between two images with strong ridge features. Although the work has only been implemented in 2D, it could potentially be extended to include scale changes between two images using voting

in a higher dimensional transform space. However, implementing the voting algorithm in higher dimensions would entail a heavy computational cost, which makes it unlikely to be practical in real-time.

4.5.3   OpenDR pipeline for 3D pose localization

We have discussed using mutual information for image matching in each degree of freedom probe's movement separately. However, in a real application of ProbeSight, the probe's pose needs to be tracked in 6 DOF simultaneously, which is a much larger search space. We have applied INS to reduce the search space to just 3-DOF, purely in translation. We now apply the OpenDR search these 3 DOFs simultaneously. The OpenDR pipeline, described below, provides a framework to express the forward graphics model, automatically obtaining derivatives with respect to the model parameters to optimize over them. With built-in local optimization and matching algorithms, it further narrows the search space and therefore greatly reduces computation time.

4.5.3.1 Defining the OpenDR Forward Process for ProbeSight

OpenDR is an open-source software system for inverse graphics that optimizes efficiently using a differentiating renderer. It was initially introduced by Matthew M. Loper and Michael J. Black for estimating human body shape from Kinect depth and RGB data (Loper and Black 2014). Inverse graphics rendering is the reverse process from computer graphics. Whereas computer graphics is primarily focused on creating realistic images from known geometry, surface reflectance, and illumination, inverse graphics attempts to take sensor data and infer 3D geometry, illumination, materials, and motions by trying to make rendered graphics realistically

reproduce the observed scene. The core idea of ProbeSight is to infer the probe's motion by matching the camera image with the prior model, which can be categorized as inverse rendering to solve for motion. Accordingly, OpenDR provides an ideal framework for estimating the probe's pose. As part of this process, OpenDR is used both to render the model and to optimize the motion/pose (extrinsic parameters).

The OpenDR pipeline implemented in ProbeSight can be explained as follows. When ProbeSight moves in space, its on-board camera image is continually updated. The OpenDR pipeline explicitly relates these updates in the observed image with changes in the model parameters θ, allowing us to estimate motion of the US probe. We define the rendered image as a function of the particular parameters we are trying to solve, $f(\theta)$. Minimizing an error metric between the observed image I and rendered 3D model (f) solves for the parameters. The standard simple error metric used by OpenDR is,

$$E(\theta) = \left\| f(\theta) - I \right\|$$

<div align="right">**Equation 4-24**</div>

OpenDR further separates the model parameters θ into 3D vertex locations V of the surface (approximately 40,000 vertices) in camera coordinates, per-vertex brightness values A that depend on the illumination model, and pre-calibrated camera parameters C, so that $\theta = \{V, A, C\}$ (Loper and Black 2014). OpenDR solves for changes in (some or all of) these parameters by optimizing them for the minimum matching error, E. To enable rapid gradient-descent optimization, a key contribution of OpenDR is that it automatically computes appropriate partial derivatives during the forward rendering process, instantaneously providing the requisite gradient, $\nabla E(\theta)$. We will further discuss this process in next section.

4.5.3.2 Differentiating Forward Rendering

For efficiency, OpenDR performs as much computation as possible in the 2D image plane, and additionally utilizes approximate heuristics. To describe the partial derivatives of the forward rendering process, the intermediate variable U represents 2D projected vertex coordinate positions from the 3D prior model. Differentiation follows the chain rule, illustrated in Figure 4-13. The differentiation process can be categorized into two groups. One is the change of appearance ($\partial f / \partial A$ and $\partial A / \partial V$) related about illumination optimization, which for ProbeSight may only be useful as a means of better measuring motion and geometry. Therefore, the illumination is currently set to be constant and is not optimized in the current framework of ProbeSight, in order to reduce computational time. The other group contains the changes in projected coordinates ($\partial U / \partial C$ and $\partial U / \partial V$) and the corresponding effects on image plane coordinates as $\partial f / \partial U$. We will only discuss the measurement of changes in projected coordinates (geometric change) in this section.

The pixel values in image plane coordinates relate to 3D coordinates and camera calibration parameters via 2D coordinates of the image, U. The corresponding partial differentiation equations can be written as:

$$\frac{\partial f}{\partial V} = \frac{\partial f}{\partial U} \frac{\partial U}{\partial V} \, , \, \frac{\partial f}{\partial C} = \frac{\partial f}{\partial U} \frac{\partial U}{\partial C}$$

Equation 4-25

$$E = \arg\min \| f(R, t) - I \|$$
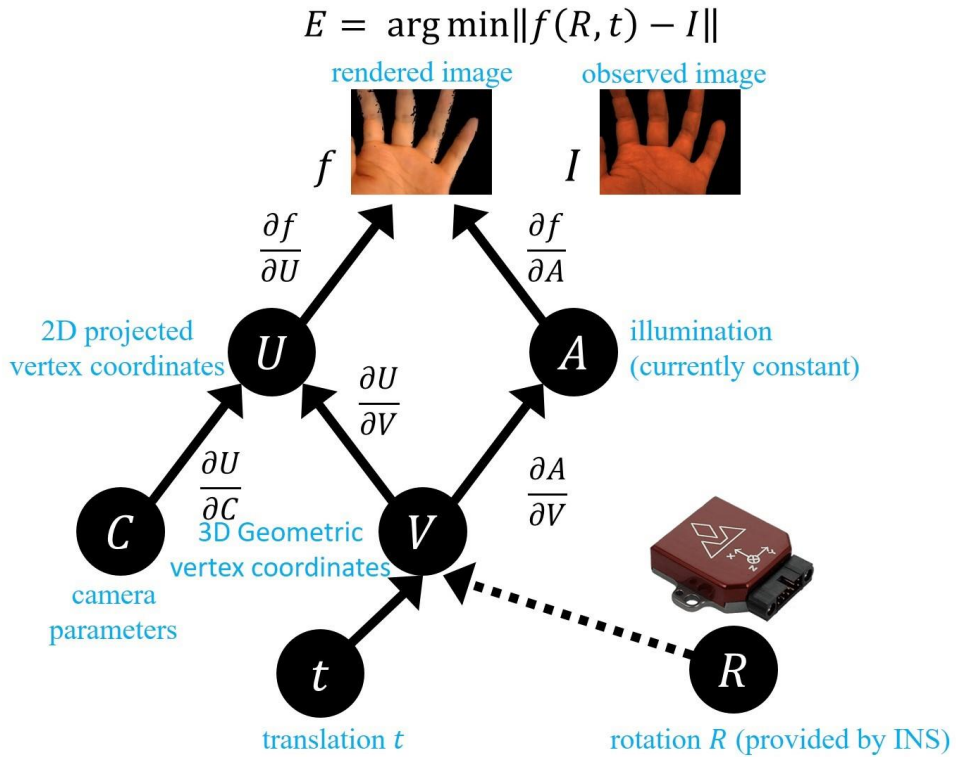
Figure 4-13: Partial derivative structure of the OpenDR renderer. (Figure is adapted from(Loper and Black 2014))

Since the projection procedure (based on the extrinsic and intrinsic parameters) is well-known and has been discussed in section 4.4, the partials $\partial U / \partial V$ and $\partial U / \partial C$ are straightforward.

In order to compute $\partial f / \partial U$, OpenDR first segments the pixels into occlusion boundary pixels and interior pixels, as described by (de La Gorce, Paragios, and Fleet 2008). The change at boundary pixels is primarily due to the replacement of one surface by another, whereas the change at interior pixels is related to the image-space projected translation of the rendered surface patch.

For *interior boundary pixels*, consider a patch translating to the right in the image plane by one pixel: each pixel in the new scene becomes replaced by its left-hand neighbor. Specifically, for the pixels not neighboring an occlusion boundary, the kernel $\frac{1}{2}[-1,0,1]$ is performed for the horizontal filtering. As Figure 4-14 shows, for the pixel neighboring an occlusion boundary on the left (red square), we use a kernel that does not include boundary pixels $[0,-1,1]$ for horizontal filtering, and for the pixel neighboring an occlusion boundary on the right (blue square), we use $[-1,1,0]$. With occlusion boundaries on both sides (yellow square), we make the rough approximation of assuming derivatives are zero. For vertical filtering, we use the same kernels transposed.

When a pixel is intersected by one occlusion boundary (green square), which is defined as an interior/boundary pixel, a kernel like $\frac{1}{2}[-1,0,1]^T$ is used. This approximates one difference (that between the foreground square boundary and the triangle surface behind it) with another (that between the foreground square boundary and a pixel neighboring the triangle surface behind it). Instead of "peeking" behind the triangle occluding boundary, we are using a neighboring pixel as a surrogate and assume that the difference is not too great. In practical terms, the boundary gradient is almost always much larger than the gradient of the occluded background surface patch, and therefore dominates the direction taken during optimization.

When more than one occlusion boundary is present in a pixel, it is defined as a *many-boundary pixel*. While object space methods provide exact derivatives for such pixels by modeling all the geometry, we treat this as an interior/boundary case. This crude heuristic is justified because very few pixels are affected by this scenario and because the exact object-space computation would be prohibitively expensive.

Figure 4-14: The illustration of differentiating Intensity with respect to 2D image coordinates. Three objects are rendered with the order from front to back as: oval surface, square surface and triangle surface. Four different pixel rendering cases are marked using different colors.

To summarize, OpenDR uses approximate differentiation at boundary pixels, where it approximates one difference (nearby pixel minus occluded pixel) with another (nearby pixel minus almost-occluded pixel). This results in efficiency in practice, but it is important to recognize that better approximations are also possible (de La Gorce, Paragios, and Fleet 2008).

OpenDR performs one render pass when a raw rendered image is requested, and an additional three passes (for boundary identification, triangle identification, and barycentric coordinates) when derivatives are requested. Although much of the computation for each pass is accomplished on the graphical processor unit (GPU), each pass also requires reading back data from the GPU.

By applying these steps discussed above, OpenDR is able to efficiently perform gradient descent optimization by deriving derivatives of the rendered image with respect to the model's

3D geometry V. We reparametrize V into translation t, as optimized by OpenDR, and rotation R, which is instead provided by the INS.

4.5.3.3 Image optimization and matching

In order to optimize the match between the reprojected 3D model and the observed image, we first read the observed camera image, then render the re-projected simulated image from the model using the forward process discussed in last section. The simulated image can be rendered at different scales into the layers of a Gaussian pyramid. As Figure 4-15 shows, in each subsequently higher layer in the Gaussian pyramid, images are scaled to lower resolution by convolution with a Gaussian blur kernel, each pixel containing a weighted average of pixels in a corresponding neighborhood on a lower level of the pyramid (Adelson et al. 1984). Images in the higher levels contains fewer features, and features that are relatively coarse, representing general shapes, while images in the lower levels contain more features that are relatively fine, representing details.
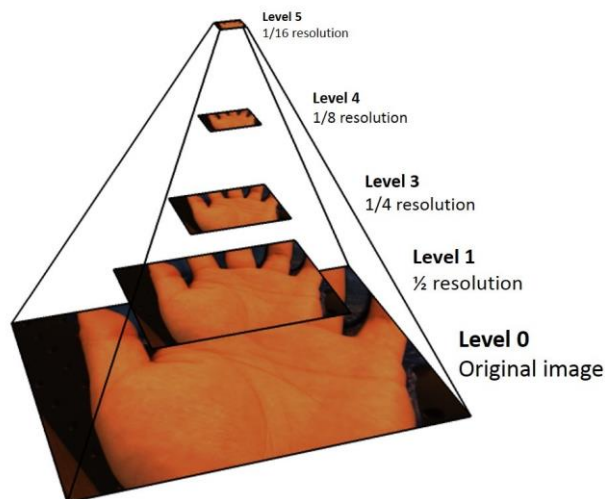
Figure 4-15: One example of a Gaussian pyramid with 5 levels of the observed image.

64

OpenDR allows image optimization across different levels of the pyramid. The higher the pyramid level we choose, the fewer features are extracted for optimization and the quicker the algorithm converges. However, this increases the matching error, since matching is mainly based on coarse features and outline information. For our ProbeSight experiments, in order to reduce computation time but still achieve a relatively high matching accuracy, we empirically chose the level of pyramid labeled 3 in Figure 4-15 (1/4 resolution). With probe motion, given the known rotation R reported by INS, we could then find a local minimum of the error with simultaneous optimization of translation t. We perform this nonlinear optimization using Powell's dogleg method (a combination of gradient descent and Newton's method) (Lourakis and Argyros 2005) which was implemented in Chumpy (Loper, n.d.) to iteratively match the prior model with the observed image.

4.6   IMAGE PRE-PROCESSING

4.6.1   Image feature enhancement

ProbeSight operates by re-projecting the 3D surface map to simulate what would be seen by the real camera from any possible point of view. An image of the re-projected surface map and the corresponding observed camera image are shown in Figure 4-16(A) and (B), respectively. For ProbeSight to accurately determine the real camera's point of view, there must be adequate detail visible in both the surface map and the camera images.
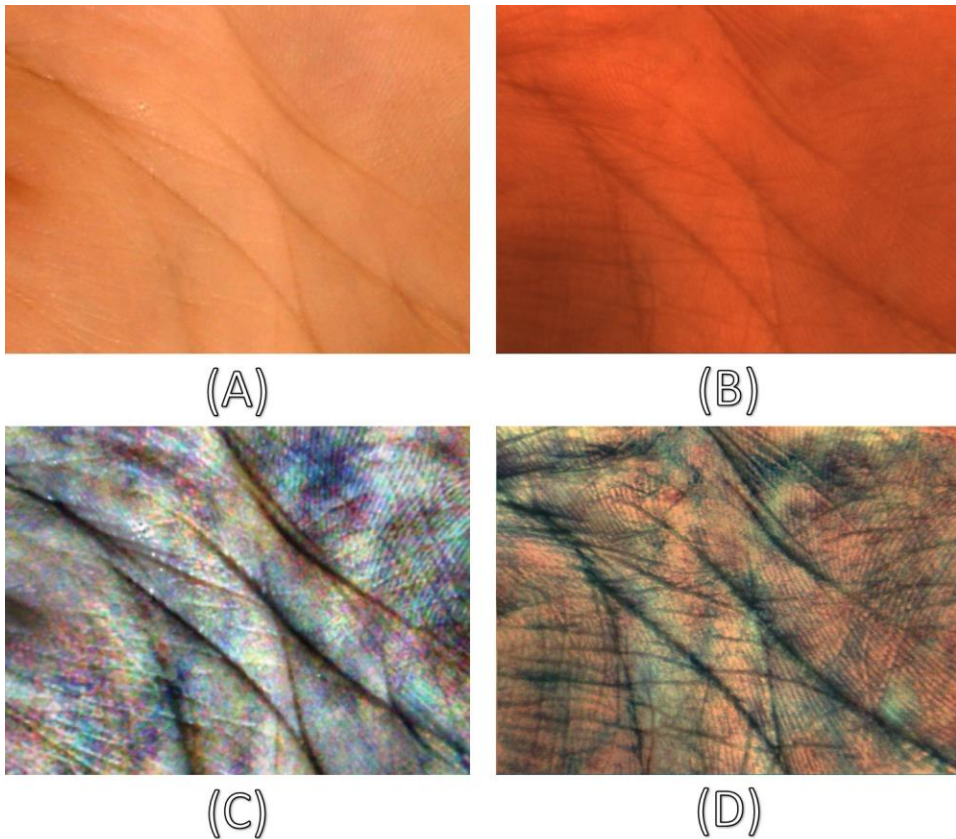
Figure 4-16: (A) and (C) Camera image before and after CLAHE. **(B) and (D)** Rendered image before and after CLAHE.

Various skin features such as skin textures, tones, and folds can be observed when the camera is focused sufficiently close to the target. However, these features can still be attenuated due to lighting, shading and other factors. To robustly track and match a sufficient number of skin features, we enhance image contrast by applying Contrast Limited Adaptive Histogram Equalization (CLAHE) to both the surface map and to the real camera images. CLAHE operates locally on small regions in the image, called tiles, rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the histogram specified by the 'Distribution' parameter. Neighboring tiles are then combined using

bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image (S.M. Pizer et al. 1990; Stephen M. Pizer et al. 1987). Figure 4-16(C) and (D) show the re-projected image and the observed camera image after applying CLAHE. The color in these images is the result of the independent processing of each of the color channels. Coarse creases are more readily apparent, and more fine creases provide additional cues for tracking. We also see some subdermal veins, which provide additional cues for tracking. Having all these features improves the results of subsequent optimization using OpenDR.

4.6.2   Image foreground-background segmentation

Background information in either the 3D surface map or the camera image that does not correspond to the other image can interfere with image matching, reducing OpenDR's accuracy. Therefore, during preprocessing our lab segments the foreground (anatomy) pixels from the background, for both the original 3D surface map and the raw camera images. In the case of the 3D surface map, this segmentation needs only be performed once, and is simplified by knowing the color of the foreground vs. the background beforehand. For the raw camera images, because the hand splint is blue, a typical "blue screen" segmentation can be performed using color-thresholding in HSV (Hue, Saturation, Value) space. The resultant foreground masks are then applied to the CLAHE-enhanced image.

Figure 4-17 shows an example of foreground-background segmentation. The original camera image with hand, splint and other background information are shown in Figure 4-17(A) and the enhanced segmented image containing only the hand is shown in Figure 4-17(B). All background pixels are set to zero to make the background completely black.
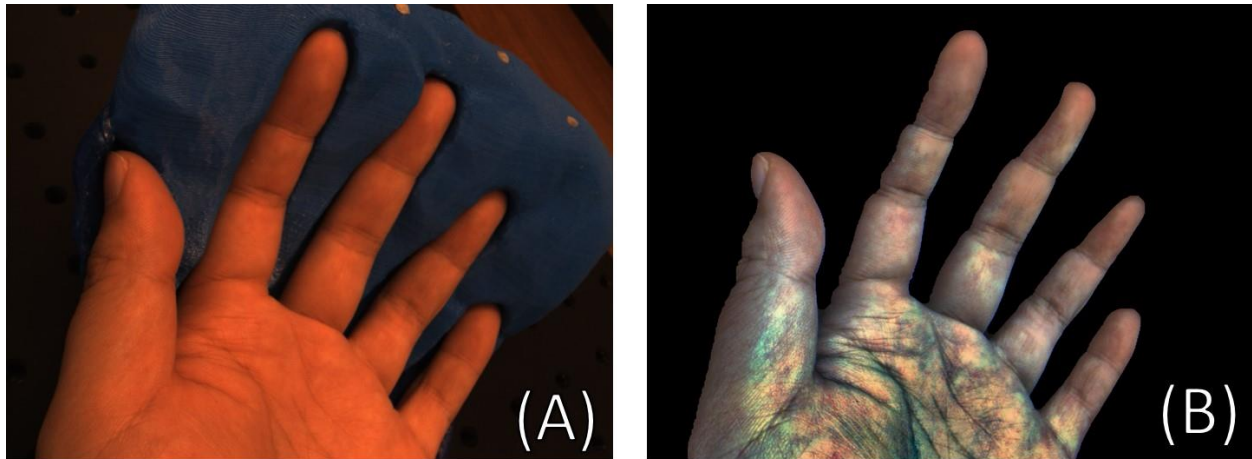
Figure 4-17: (A) The original camera image of the hand with background information. **(B)** The segmented feature enhanced image of the hand.

4.6.3 Image Cropping and Downsampling

After each 1936×1456 pixel camera image is acquired, a 640×480 central pixel region (RGB) is cropped from the camera's sensor. Since the underlying camera's raw data has a lower color resolution, inherent to all Bayer-pattern sensors, we then downsample the cropped region to 320×240 pixels. This cropping and downsampling substantially speeds up subsequent processing by OpenDR, while preserving most of the original color information.

5.0   PHANTOM EXPERIMENT

In this chapter, the tracking performance of ProbeSight is evaluated on both the dinosaur model and flat phantom. For the dinosaur model, the accuracy of the ground truth (Micron tracker) and the suitability of our mutual information metric was validated by deviating the simulated image along each of the 6 DOF, projecting the surface map from each new viewpoint based on that viewpoint's offset from the Micron tracker, and applying the mutual information metric between it and the actual camera image. For the experiment on the flat phantom, the tracking performance of ProbeSight was initially tested on a rotation free environment by mounting it on a 3-axis linear stage, and then extended to a 6-DOF freehand experiment. The estimated motion error compared with ground truth was computed and graphed as the accumulated error in Euclidean distance versus frame number or travel distance.

5.1   PROBESIGHT TRACKING ON THE DINOSAUR MODEL

5.1.1   Experiment setup

The experimental setup for the dinosaur model is shown in Figure 5-1. An optical marker is mounted on the camera to establish ground truth from which to compute the re-projected image, with the optical tracker fixed on a tripod facing down to track the camera's pose. The camera is

mounted on a linear slider base, which enables ProbeSight to acquire data from different regions of the dinosaur model.

Figure 5-1: Experiment setup for tracking the probe's position relative to the dinosaur phantom.

5.1.2 Probe Initialization

With the computed extrinsic parameters and pre-calibrated camera intrinsic parameters established, as described in section 4.4.5, ProbeSight is able to re-project the previously acquired 3D surface map to simulate what would be seen at the initial camera's pose, as determined by the

optical tracker. An example of an observed camera image with its corresponding simulated image is shown in Figure 5-2.

Figure 5-2: Dinosaur model: Pre-acquired surface map **(A)**, rendered from the correct viewpoint, with the actual camera image **(B)** (Wang et al. 2014).

5.1.3   Experimental Protocol

Once the simulated re-projected image is matched with the current camera view at the initial pose, the mutual information value between them is computed. We then translate or rotate the simulated camera along each of the individual 6 DOFs from the initial viewpoint of the 3D surface map, keeping the same physical camera image and computing the MI metric between the original physical image and the new simulated camera image. Because the translation and rotation was achieved in the reprojection of the model, the physical camera was not moved over the course of a single experiment.

71

5.1.4   Experimental Results and Discussion

With an image size of 640×480 pixels, we implemented sequential displacements of 1 pixel or rotations of 1 degree in each DOF. The results are shown in Figure 5-3, with the initial pose marked by the vertical line in the middle of each graph. Along each DOF, the normalized MI metric shows a clear maximum at the ground truth established by the tracking apparatus, and diminishes nearly monotonically as one moves away from the correct pose (Wang et al. 2014).
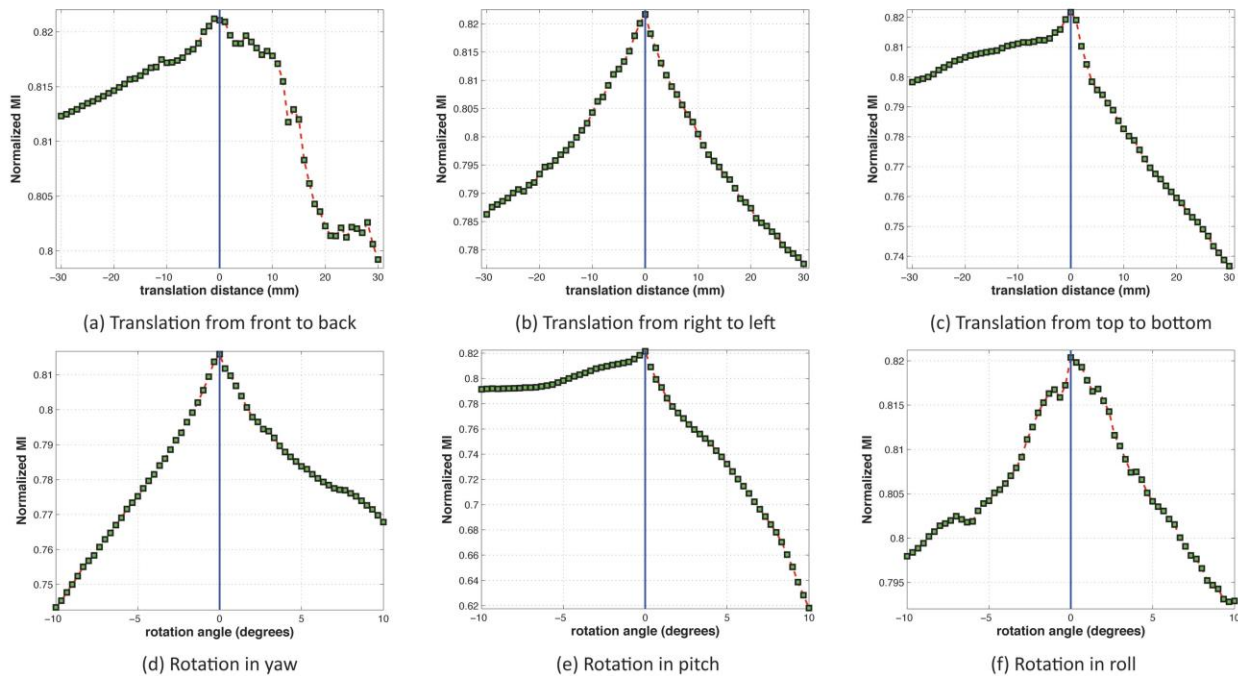
Figure 5-3: Normalized mutual information (MI) for deviations from ground truth in each DOF (Wang et al. 2014).

The dinosaur phantom experiment was just an initial experiment to establish (1) that a previously acquired 3D surface map of a phantom could be projected to match the view through an actual camera, and (2) that MI is an effective metric to determine which viewpoint best matches the camera image.

A major source of error in the matching process is the difference in lighting conditions between the projected surface map and the actual camera image. Although mutual information can cope with lighting differences better than some other metrics (e.g. those based on correlation), lighting differences may still lead to pose-matching errors, especially if there are differences in shading. Lighting-related errors could be reduced by applying image feature enhancement algorithms discussed in section 4.6.1 and using feature based matching discussed in sections 4.5.2 and 4.5.3. Applying a lighting studio as mentioned in section 3.4 would also help remove shading artifacts, mitigating differences between the projected surface map and actual camera image. All these improvements were implemented for both the phantom experiments and the in-vivo human hand experiments, as discussed in later sections.

5.2   PROBESIGHT TRACKING ON THE FLAT PHANTOM

5.2.1   Experimental setup

In this series of experiments, ProbeSight was tested on a flat phantom containing a 2D printed image of actual palm skin. The aim of these experiments was to test ProbeSight's 6-DOF tracking performance in an environment more similar to real human anatomy than the dinosaur model, but still simpler to implement and easier to verify than an actual human hand. Instead of using a MI based metric for matching, we switched to using our customized OpenDR pipeline for pose estimation. Compared with our previous exhaustive 3D search in MI space, the built-in OpenDR optimization and matching method discussed in section 4.5.3.3 effectively reduces the search space and computation time. To make the experimental results consistent, we started with

the simplest 1-DOF translation experiment, then extended it to 2 and 3-DOF translation experiments, and eventually moved onto a full freehand 6-DOF experiment.

For the translation only experiments, ProbeSight was mounted on a 3-axis linear stage, allowing it to translate but not rotate within a one-cubic-inch (25.4 mm) 3D working space. The camera faced down and focused sufficiently close to the printed image on the phantom. The Micron tracker provided the ground truth data for each new location with sub-millimeter accuracy and was also used during in the initialization step as discussed in section 5.1.2. For the freehand 6-DOF experiment, the INS was applied to accurately measure rotation.

5.2.2 One-DOF experiments

During the 1-DOF experiments, an operator translated ProbeSight in steps of 0.05 inch (1.27 mm) in a particular direction. At each step, a camera image was recorded along with the ground-truth data provided by Micron tracker. A total of 20 steps were recorded in each DOF inside the given 3D working space.

In order to examine the effects of the CLAHE image-feature enhancement algorithm, a comparative study was performed on an example 1-DOF dataset before and after applying the algorithm. For the dataset before applying feature enhancement algorithm, the accumulated error was 41.31 ± 14.75 mm over the total 25.4 mm travel distance, which obviously indicates that ProbeSight failed to accurately track the probe's motion. When ProbeSight tracked motion on the same dataset after CLAHE feature enhancement, however, the accumulated error was only 1.18 ± 0.55mm over the total 25.4 mm travel distance. Figure 5-4 shows both results. To accommodate the extremely large error accumulated without image enhancement, a plot with a

logarithmic scale for the accumulated error in Euclidean distance (ED) is used. The result shows

that the CLAHE feature enhancement greatly improves the probe tracking results.
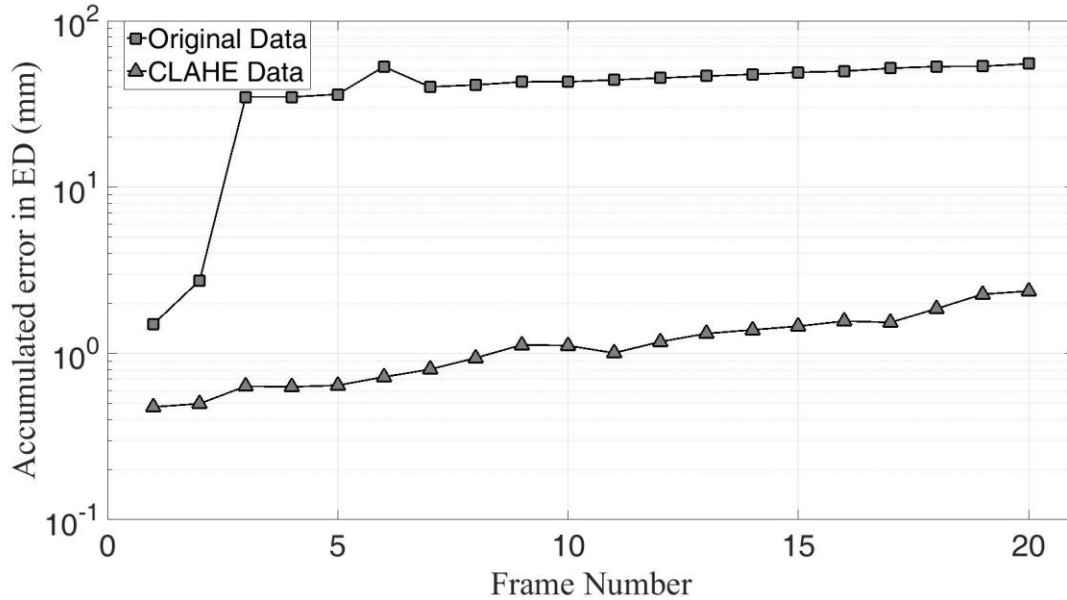
Figure 5-4: Performance comparison of ProbeSight on 1-DOF phantom experiment data with or without

applying CLAHE feature enhancement.

The pose estimation errors on each DOF are shown in Figure 5-5. On average, over the

probe's total travel distance of 25.4 mm in each direction, the accumulated error was

1.00±0.63mm. A major source of the error is the introduction of the background into the camera

image; as Figure 5-5(A) shows, when ProbeSight moves to the boundary of the phantom, the

accumulated error greatly increases. Furthermore, since errors are introduced in directions other

than that of the direction of motion, we show both total ED error (triangles) and error just in the

direction of motion (squares), which are always less than ED. In fact, it seems that most of the

error is in directions other than the direction of motion.
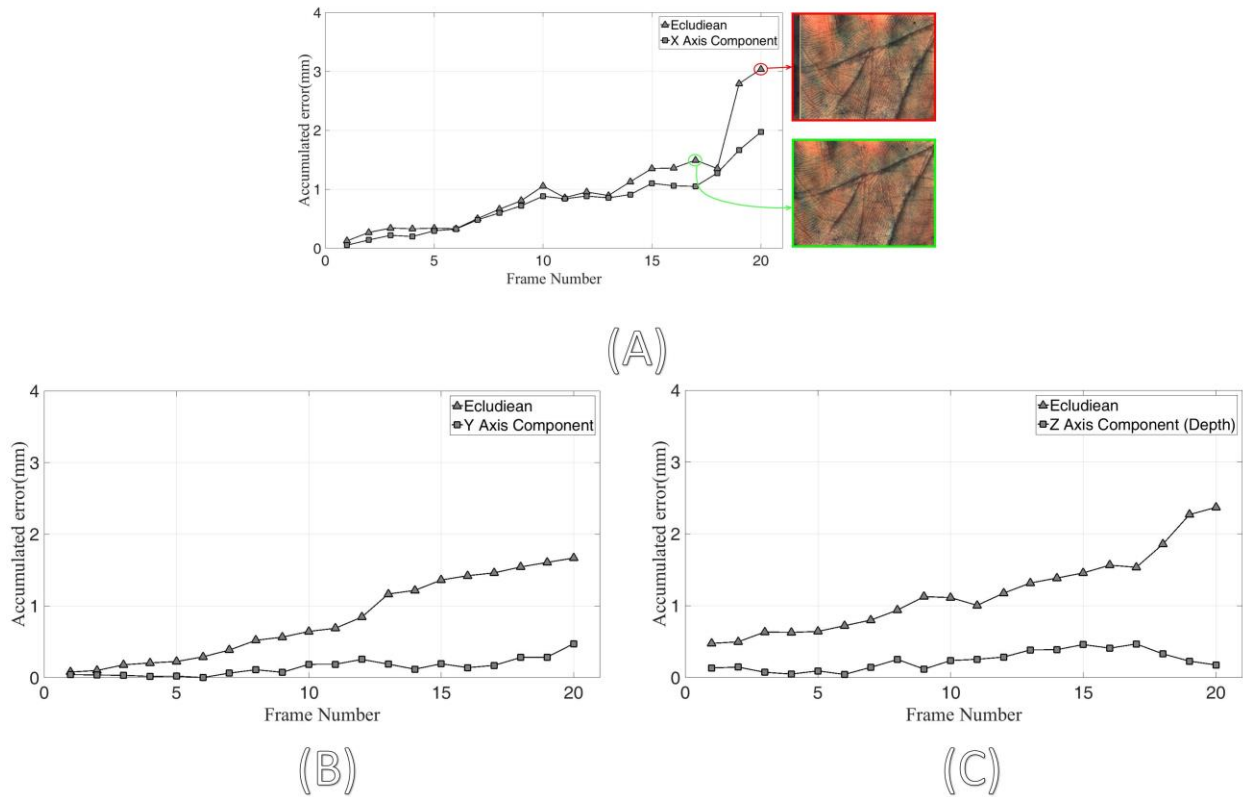
Figure 5-5: (A) 1-DOF x-axis experiment's accumulated error. The comparison of frames with/without background information are circled in the graphs. **(B)** 1-DOF y-axis experiment's accumulated error. **(C)** 1-DOF z-axis experiment's accumulated error (depth direction).

5.2.3 Two DOF experiment

The 1-DOF experiments only allowed ProbeSight to move over a linear distance in space, which limited our study. A 2-DOF experiment was performed to further analyze ProbeSight's tracking performance, this time over a path tracing a square. This allowed us especially to test the algorithms ability to self-correct when moving from between regions with and without background visible.

During the 2-DOF experiment, the subject was instructed to translate ProbeSight 0.05 inch (1.27 mm) at each step in the positive direction along the x axis. Once the movement reached a maximum 1-inch displacement, ProbeSight was then translated along the positive y axis, the negative x axis and finally the negative y axis, to complete a square path. No z axis displacement (depth change) was permitted in this experiment. For the movement along each axis, 20 camera images were recorded, making a total of 80 frames for the entire experiment. Nearly half of the image frames contain background information. ProbeSight's 2-DOF motion trace computed by OpenDR was compared with ground truth data reported by the Micron tracker. The upper panel of Figure 5-6 shows the accumulated ED error for an example trace, and the actual trace is shown in the lower panel of (triangles) compared to the ground truth (circles). We observed that the accumulated error markedly increased when the probe was in a region with background (see representative frames 15 and 26 in the figure). It is also shown that ProbeSight was able to self-correct from the error, and the accumulated error drops when the probe moves back to a region without background as the representative frame 46 shows.

In this experiment, the error was mainly caused by specular artifacts from the phantom's glossy background surface and by various unmatched features from the real-time camera images, such as the dark threaded holes in the optical table. Either of these could mislead the algorithm to make false matches.
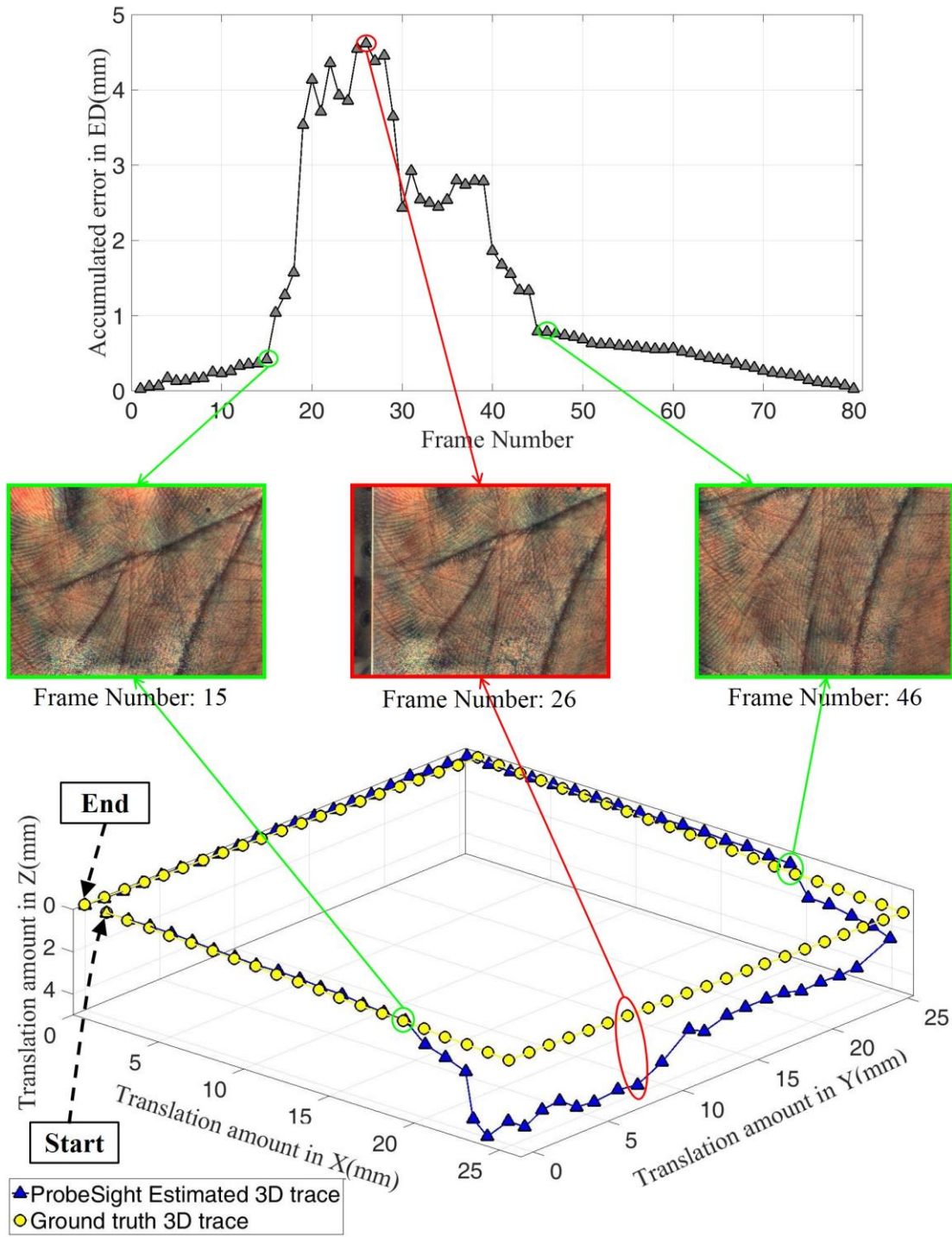
Figure 5-6: Upper panel: Accumulated error in Euclidian Distance. **Lower panel:** Estimated probe trajectory (blue triangles) compared with ground truth data (yellow circles) for the 2-DOF translation only experiment data.

5.2.4 Three DOF experiment

For a final experiment using the 3-DOF translation stages, instead of translating ProbeSight a fixed amount one direction each step, the experimenter translated ProbeSight an arbitrary amount in all 3-DOF directions at each step. An example 3-DOF trace with ground truth is shown in Figure 5-7(A), with the accumulated ED error shown in Figure 5-7(B). As with the initial 1-DOF experiment, the accumulated error was only 0.64 ± 0.49mm, over the probe's total travel distance of approximately 200mm in 3-DOF space.
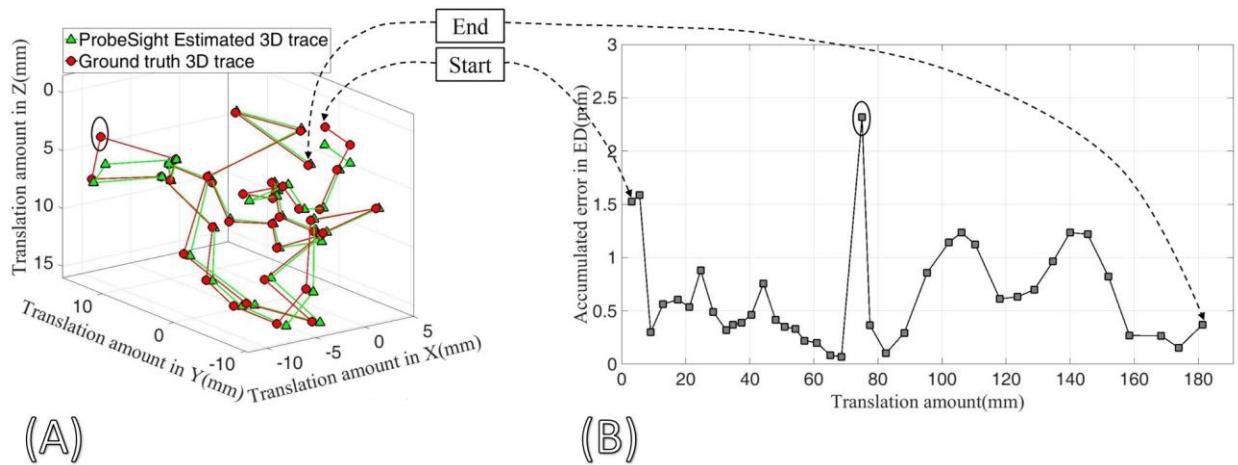
Figure 5-7: (A) Estimated probe trajectory (green triangles) compared with ground truth data (red circles) for the pure translation experiment data. **(B)** Accumulated error in Euclidian Distance. The outlier is circled in both graphs.

In summary, ProbeSight achieved a reliable performance in transplantation-only experiments using the phantom. Overall, the major source of error was due to misleading background information, not surprisingly, since the backgrounds differed between the prior scan and the real-time video images. Other sources of errors include inaccuracies in pre-calibration, manual error when turning linear stage knobs, and ground truth error, the last of which will be discussed in section 5.2.6.

5.2.5 The 6-DOF freehand experiment

Finally, we extended the experiment to a full 6-DOF range of motion. The experimenter held the device free-hand and moved it around the phantom to various locations and orientations. The camera, tracker, and INS synchronously acquired data at approximately 30 frames per second. During the experiment, orientation change was directly measured by the INS, leaving the OpenDR pipeline to estimate only translation between subsequent frames.

For the freehand experiment, the problem with background remained challenging. Therefore, we applied the method discussed in section 4.6.2 to segment foreground from background. Figure 5-8 shows the accumulated ED error for the same original dataset with and without the foreground-background segmentation procedure. For the unsegmented dataset, the accumulated error increases when the background becomes visible (circled in red with corresponding red-framed image), as representative frame 29 shows. The error climbs when more background "features" emerge, especially in frame 36 where the background contains dark screw-holes in the optical table, which are not present in the pre-scanned model. By comparison, the segmented dataset, in which the background is set to zeros, (circled in green with corresponding green-framed images) shows a drastic reduction in error and improvement in tracking performance.
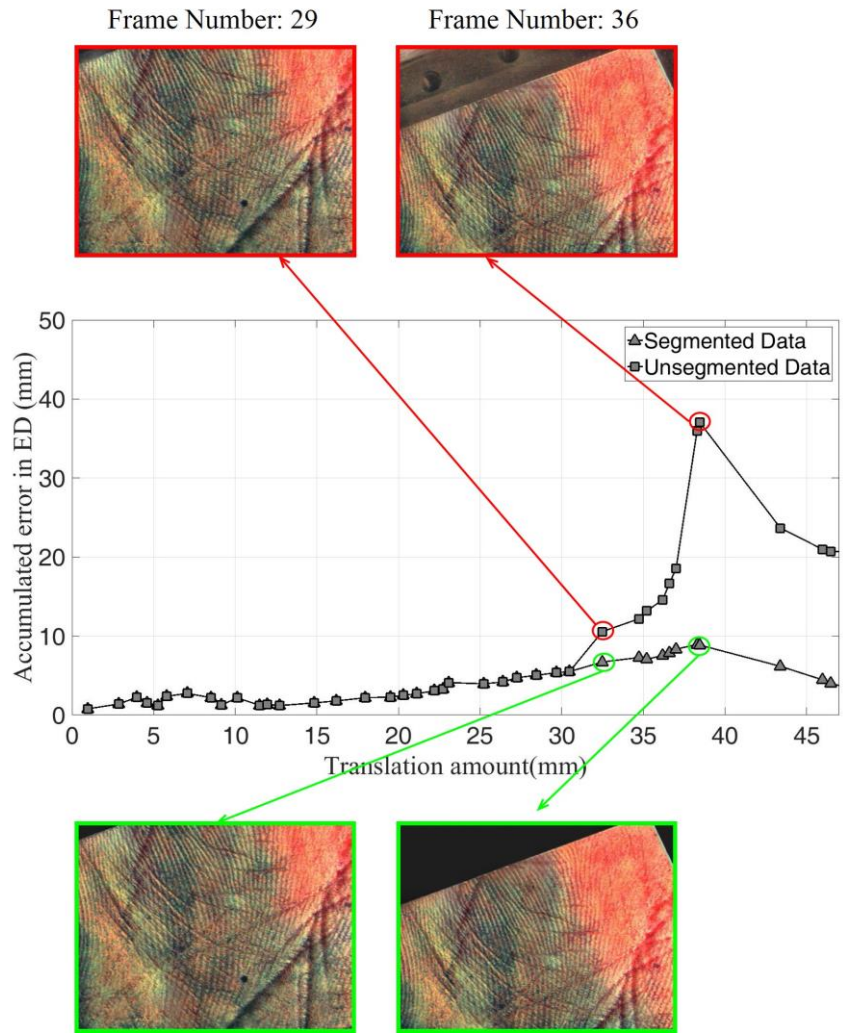
Figure 5-8: Performance comparison of ProbeSight using the freehand flat-phantom experiment data, with or without background segmentation.

We also analyzed two possible ProbeSight configurations, either using the INS or OpenDR to track 3D rotation (OpenDR was used in both cases to track translation). Without the INS, the accumulated error was 141 ± 70 mm over the first 100 mm of travel. Clearly, OpenDR failed to accurately track rotation, being unable to disambiguate between camera rotation and translation (a well-known problem in computer vision). When OpenDR determined translation and relied on the INS to determine rotation, however, the accumulated error was only 4.06 ± 2.38mm over the first 100 mm of travel. Figure 5-9 shows both results. To accommodate the extremely large error accumulated without INS guidance, a plot with a logarithmic scale for the accumulated error in ED is used.
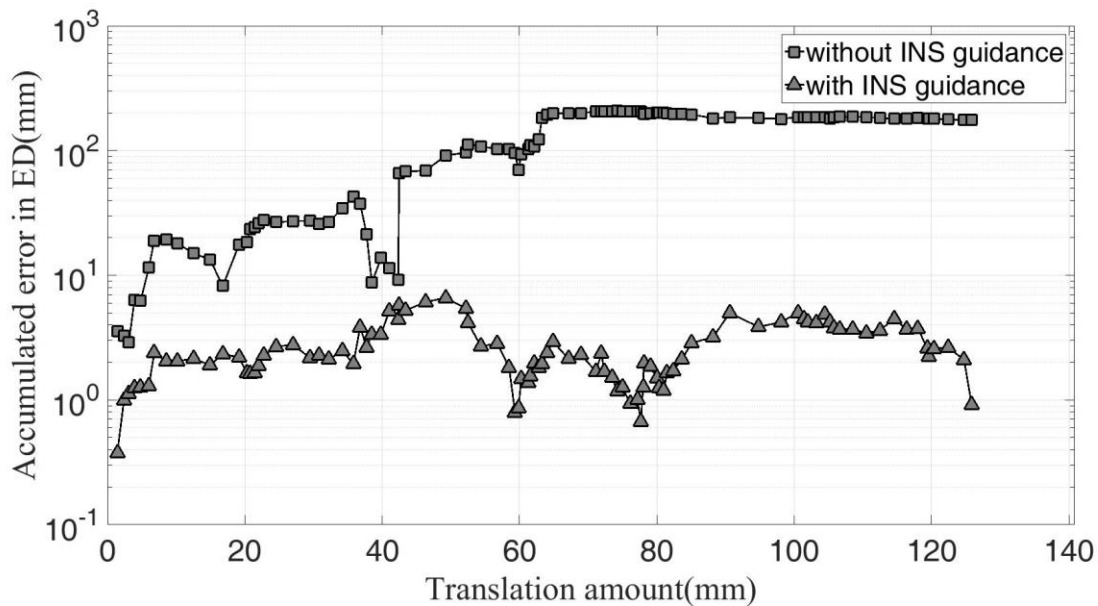
Figure 5-9: Performance comparison of ProbeSight using the freehand flat-phantom experiment data, with or without using INS guidance for rotation.

The 6-DOF freehand experiment further validated ProbeSight's tracking performance and also demonstrated that natural skin features could potentially be directly tracked for freehand

probe localization. The application of both hardware (light studio) and software (image enhancement and foreground-background segmentation algorithms) further improves the tracking accuracy of ProbeSight. We also discovered that OpenDR itself is not able to address the problem of distinguishing between camera rotation and translation, but it did show high accuracy for translation-only estimation when combined with INS-derived rotational pose.

5.2.6 Optical tracker jitter effect analysis

So far we have discussed ProbeSight tracking with the Micron optical tracker as ground truth. However, the Micron tracker suffers from jitter and latency, which degrades its tracking performance (Teather et al. 2009). Consequently, the ProbeSight tracking results may not accurately reflect their potential. For the current applied Micron Tracker device, the latency incorporating processing time and lag is about 27 ms, which has almost no apparent effect on our experiment, since the camera runs at 30 frame per second (fps) and we don't require higher speed tracking. However, spatial jitter does affect performance. It is straightforward to see jitter by observing the vibration of the 2D re-projected image when under the sole guidance of the Micron Tracker. Since translation data provided by tracker is very stable (approximately ±1mm), the jitter effect is mainly caused by errors in orientation. It is not surprising that orientation error is greater, given that it is calculated from small changes in the location of individual features in a relatively small, printed-pattern marker on the target. In this section, we investigate this error by keeping camera pose unchanged and analyzing the corresponding orientation data. The orientation data reported by the INS was recorded simultaneously for comparison purposes.

5.2.6.1 Drift comparison between INS and tracker under the same illumination condition

The experimental setup is similar to the phantom experiment. The camera was kept motionless under constant illumination while a total of 100 orientation readings were recorded simultaneously from both the INS and optical tracker. Although values in the two data sets are in different coordinate systems, the orientation changes are in the same units. Therefore, changes in orientation at each step for both the tracker and INS can be converted to Euler angles (yaw, pitch and roll). These are separately plotted in Figure 5-10, clearly indicating that the optical tracker data has greater fluctuations compared with INS data. Statistical analysis (Student's unpaired two-tailed t test) was performed to confirm that there is a significant difference between the two groups of data in each DOF as Table 5-1 shows.
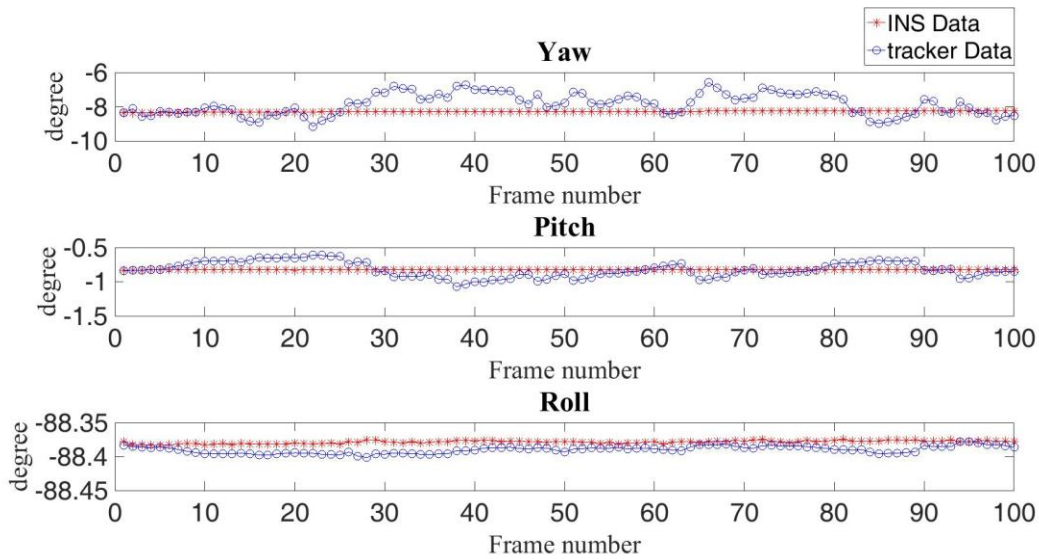
Figure 5-10: The graph comparison between INS and tracker orientation data in each DOF.

Table 5-1: *t-test* comparison between INS and tracker orientation data in each DOF

Measure	INS			Tracker			*Error t* test
	N	*Range*	*SD*	*N*	*Range*	*SD*	*p*
Yaw (degree)	30	0.113	0.0072	30	2.570	0.4581	<.0001
Pitch (degree)	30	0.012	0.0018	30	0.461	0.0817	<.0001
Roll (degree)	30	0.009	0.0015	30	0.024	0.0033	<.0001

Note. Mean, standard deviation, and results of two-tailed t test assuming unequal variances.
30 time series data from middle of the data have been used for t test.

To further explore the effect of jitter on the simulated 2D re-projected image, 100 frames of re-projected images were acquired with the camera motionless, either using the INS or optical tracker to guide 3D rotation (the optical tracker was used in both cases to guide translation). For each dataset, A GPU-based dense optical flow algorithm was applied to calculate flow for every pixel in both x and y. Figure 5-11(A) shows an arrow map to demonstrate the flow, which reflects the jitter between two consecutive frames.

To quantify the effect of jitter on the 2D simulated re-projected image, we computed the average flow along both x and y axes for all pixels, defined as drift error Δx and Δy, and then projected them for all 100 frames into a 2D plane. Figure 5-11(B) shows clearly that the effect of jitter from the Micron tracker is much larger than from the INS, with Micron's errors more widely dispersed in the graph.

The Euclidean distance (EU) error in pixel can be computed based on drift error in each axes, as $\Delta d = \sqrt{\Delta x^2 + \Delta y^2}$. Statistical analysis (Student's unpaired two-tailed t test) for the Euclidean pixel distance error from the two datasets is summarized in Table 5-2, which further validates that there is a significant difference in the effect of jitter on the re-projected images between INS and optical tracker guidance.

(A)   (B)
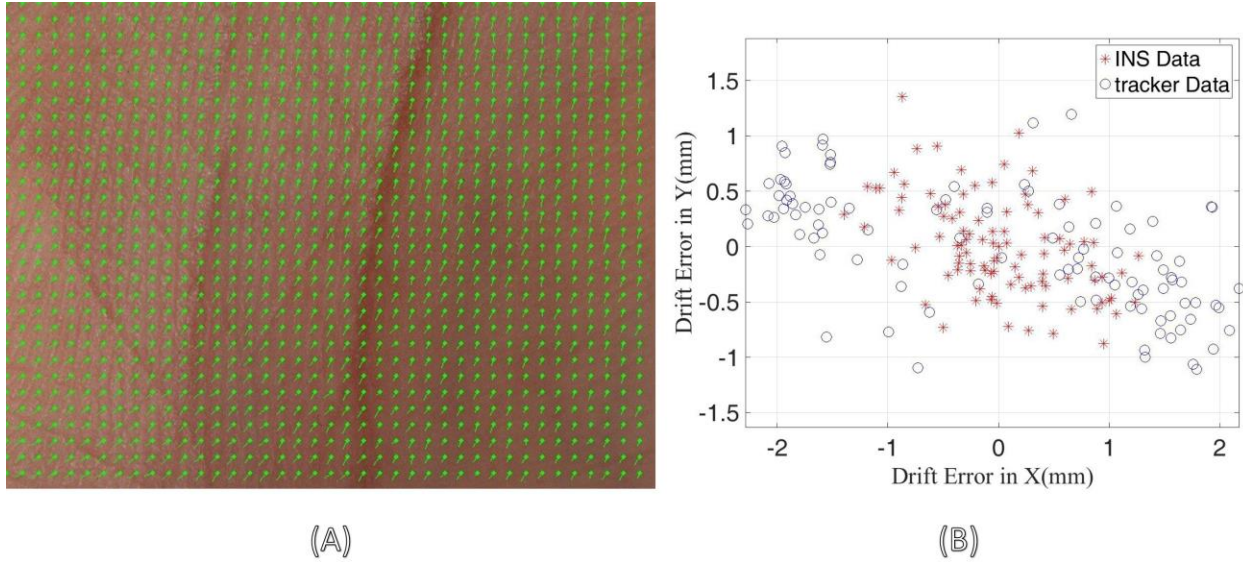
Figure 5-11: (A) Arrow map visualization of optical flow between two consecutive frames. For visualization

purposes, the arrow map is plotted for every 10 pixels along both x and y axis. **(B)** The projected 2D plot of drift

error among 100 consecutive re-projected images guided separately by INS vs. tracker.

Table 5-2: *t-test* comparison of Euclidean distance error on 2D simulated re-projected image between INS

and optical tracker guidance

Measure	INS			Tracker			*t* test
	N	*Mean*	*SD*	*N*	*Mean*	*SD*	*p*
EU error (pixel)	30	0.14	0.1366	30	1.72	1.2547	<.0001

Note. Mean, standard deviation, and results of two-tailed t test assuming unequal variances.
30 data points from the middle of the time series have been used for the t test

5.2.6.2 Drift comparison between INS and tracker under different illumination conditions

The optical tracker relies on a Sum of Squared Differences (SSD) correlation-based stereo vision

algorithm, and is sensitive to changes in illumination and shading across the printed markers

(Hager and Belhumeur 1996). We performed an experiment to explore the optical tracker's jitter

86

under different illumination conditions. During the experiment, the camera was again kept motionless. A set of 100 consecutive re-projected images were collected guided by the optical tracker under two illumination conditions. The aim of the experiment was to test whether there was a significant difference of error in Micron's "ground truth" data under different illumination conditions. INS data was collected as a control group, and we predicted that different illumination would not significantly change the drift error when the re-projected images were guided by INS.

The results when using the INS for rotation are shown in Figure 5-12(a). As expected, there is not much difference of drift error under different illumination conditions when guided by the INS. However, if the Micron optical tracker is used to measure rotation instead of the INS, the drift error is more widely spread under low illumination compared with high illumination as shown in Figure 5-12(b).
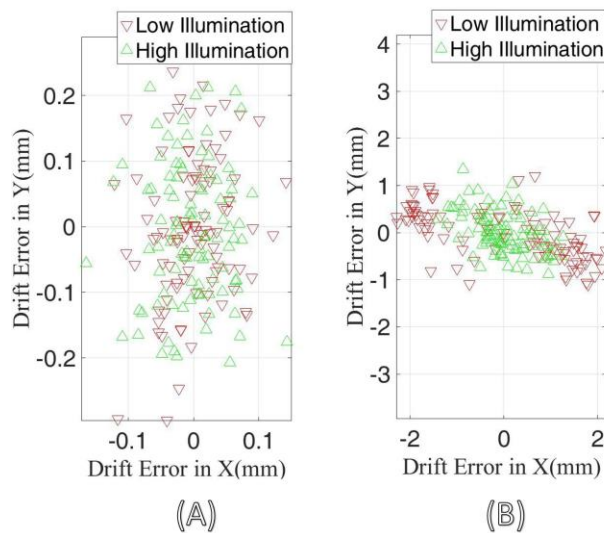
Figure 5-12: The projected 2D plane plot of drift error among 100 consecutive re-projected images guided by INS under two illumination conditions. Figure 5-12(b): The projected 2D plane plot of drift error among 100 consecutive re-projected images guided by optical tracker under two illumination conditions. (Note different scales)

Statistical analysis (Student's unpaired two-tailed t test) for the Euclidean pixel distance errors under two lighting conditions from both INS and optical tracker datasets are shown in Table 5-3. There is a significant difference of Euclidean pixel-distance error under different illumination conditions for the purely optical tracker data, while no significant difference for the INS data.

Table 5-3: *t-test* comparison of Euclidean distance error between two illumination conditions

Measure	Low Illumination			High Illumination			t test
	N	*Mean*	*SD*	*N*	*Mean*	*SD*	*p*
INS EU error (pixel)	30	0.1122	0.0832	30	0.0879	0.0531	.1854
Tracker EU error (pixel)	30	1.3867	0.6375	30	0.6504	0.3960	<.0001

Note. Mean, standard deviation, and results of two-tailed t test assuming unequal variances.
30 data points from the middle of the time series have been used for the t test

6.0   IN-VIVO HUMAN HAND EXPERIMENT

In this chapter, the performance of ProbeSight is evaluated on three datasets containing various skin textures and geometry from different anatomical regions of the upper extremity: finger, palm and wrist. Motion estimation errors in probe localization have been quantified. For tracking on regions with high accumulated error, we have proposed a method to incorporate more features from nearby regions to reduce error.

6.1   EXPERIMENTAL SETUP

The experimental setup is shown Figure 6-1. Both the molded hand splint and Micron optical tracker (not visible in the picture) have been screwed down to an optical table-top, which is secured on the desk inside a light studio, so that the relative pose between the hand splint and the optical tracker remains unchanged during experiments. The transform from anatomical coordinates to tracker coordinates requires calibration only once during the probe initialization process. The light studio provides sufficient diffused lighting to avoid motion blur and specular reflections, as discussed in section 3.4. The high-resolution ultrasound (HRUS) system is seen on the right side of the light studio with its 70MHz transducer incorporated into an assembly containing both camera and INS, as discussed in section 3.1. The ultrasound image is transmitted to a processing computer system on the left side of the light studio by using a video

frame grabber (DVI2USB, Epiphan Video, Ottawa, Canada). Although integration of the HFUS data into the ProbeSight system is not an essential part of the current thesis, concurrent work by other graduate students in our research group aims to merge the analysis of ultrasound and video data to further the utility of each, in much the same way that this happens for the human operator. The current system is being used for this other research as well, hence the incorporation of the ultrasound data.
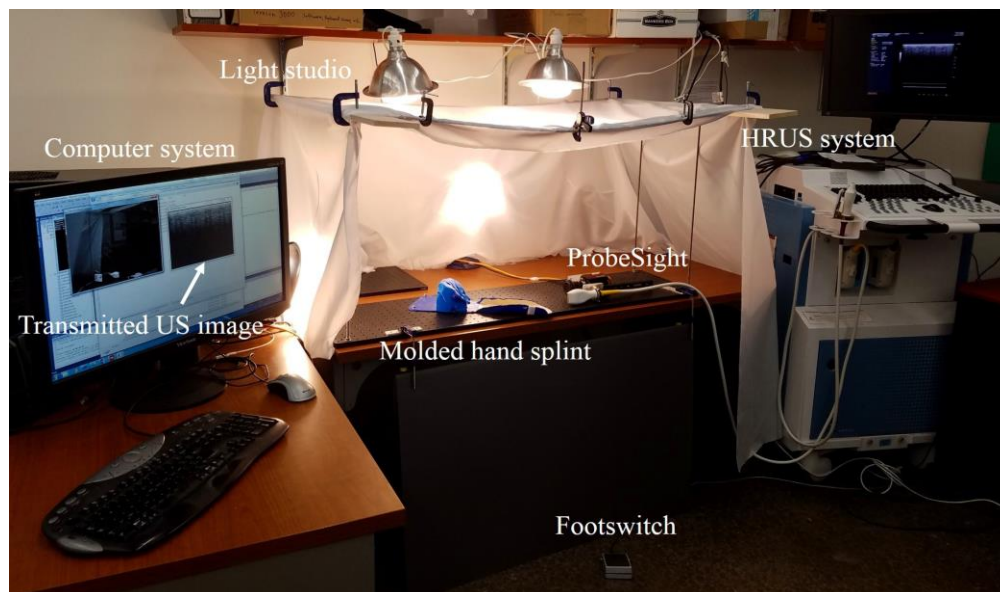
Figure 6-1: The In-vivo human hand experiment setup.

During the experiment, we can tap the footswitch under the desk to start and stop the data acquisition process. Once the experiment is done, all data in each camera frame, the INS data, the ground truth optical tracker pose, and the corresponding ultrasound image are stored on disk. The appropriate data are then transmitted and processed through the OpenDR pipeline in python for offline pose estimation.

6.2   DATASET DESCRIPTIONS
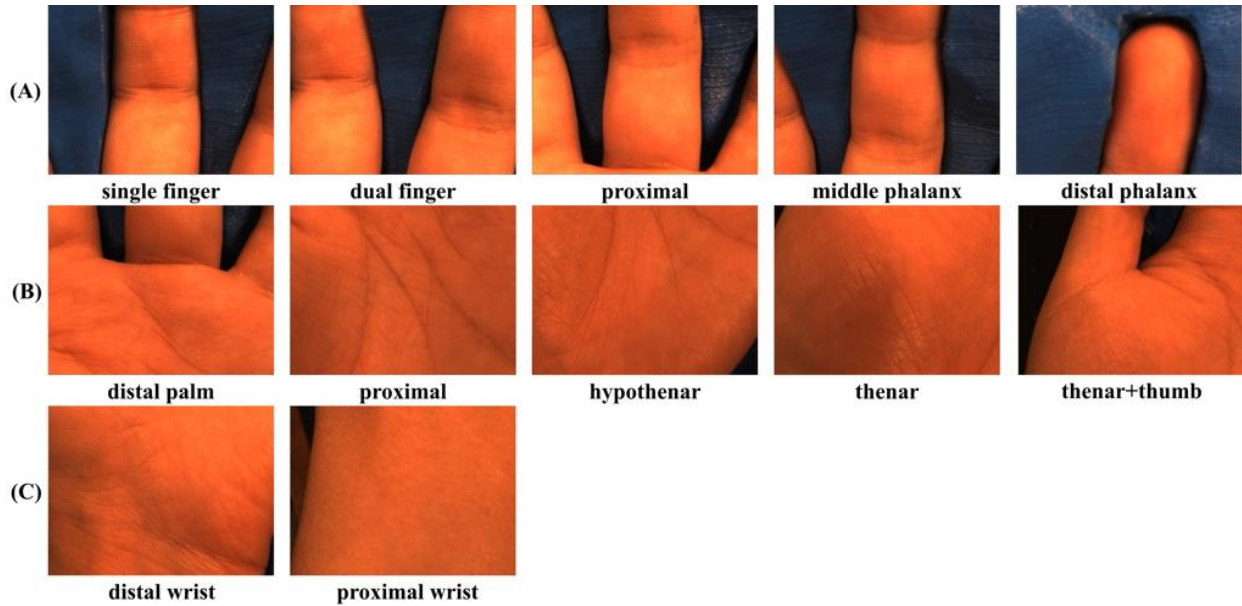
Figure 6-2: Example frames of three datasets: **(A)** Finger Dataset, **(B)** Palm Dataset and **(C)** Wrist Dataset.

We have discovered that ProbeSight's performance depends critically on the particular anatomical region being scanned and path along which the scanner is moved. Thus we divide our scans of the hand into major datasets (Finger, Palm, and Wrist), each of which are further divided into smaller datasets according to direction and particular location of the scan.

Some representative frames from these datasets are shown in Figure 6-2, and each of these regions and creases are labeled in Figure 6-3. Our *finger dataset* contains 392 images continuously acquired from ProbeSight's camera while moving across the fingers. The dataset consists of two parts, (1) for side-to-side motion across the proximal region of the fingers, and (2) for motion along the length of the fingers. In the first case, the *finger medial-lateral* (FML) dataset, the probe shifts from viewing both the middle and index fingers (*dual finger*), to viewing

just the index finger (*single finger*), and then back to viewing both fingers (*dual finger*). In the second case, the *finger proximal-distal* (FPD) dataset, the probe moves along the middle finger from the proximal phalanx to the middle phalanx to the distal phalanx and then back again (with a view that also partially includes the index finger). Both finger edges and coarse creases are visible, including the *distal digital crease*, *middle digital crease*, and *proximal digital crease*. CLAHE preprocessing accentuates the creases in these images, thereby facilitating their use during OpenDR optimization.
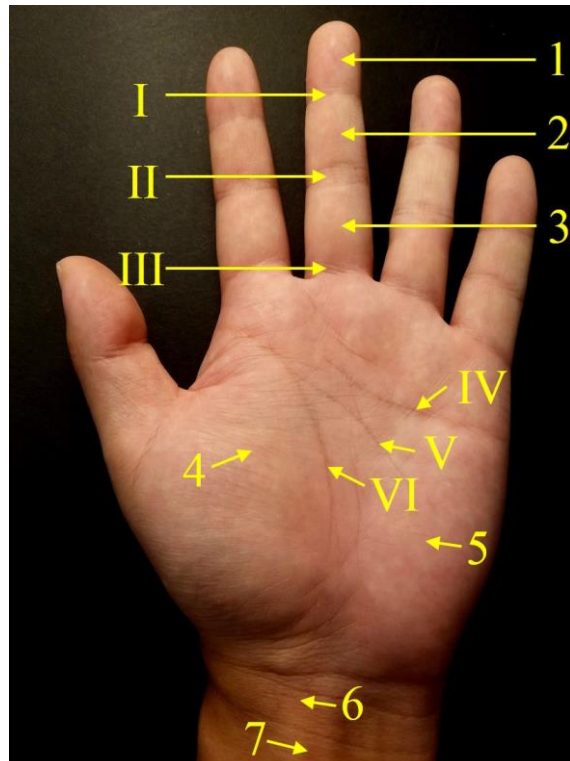
Figure 6-3: Upper extremity regions and creases. The regions are (1) distal phalanx, (2) middle phalanx, (3) proximal phalanx, (4) thenar, (5) hypothenar, (6) distal wrist and (7) proximal wrist. Major creases are (I) distal digital crease, (II) crease, (III) proximal digital crease, (IV) distal transverse creases, (V) proximal transverse crease and (VI) radial longitudinal crease.

92

The *Palm Dataset* contains 610 images continuously acquired from ProbeSight's camera while moving across five regions of the palm. Again, there are two parts to this dataset, *palm medial-lateral* (PML) and *palm proximal-distal* (PPD). PML motion is from the hypothenar to thenar to thenar+thumb regions, and PPD motion is from the distal to proximal palm regions. Visible coarse creases include the distal transverse crease, the proximal transverse crease and the radial longitudinal crease on the palm area. Only the thenar region lacks obvious coarse creases and edge information.

The *Wrist Dataset* contains 161 images continuously acquired from ProbeSight's camera while moving along the wrist. This *wrist proximal-distal* (WPD) motion is from the distal wrist region to the proximal wrist region. Almost no coarse creases or obvious edges are visible in these regions.

6.3   EXPERIMENT RESULTS AND DISCUSSION

The efficacy of ProbeSight varies with the particular anatomical region and direction of camera motion.

Results from the FML dataset are shown in Figure 6-4(A). The accumulated error climbs when the probe moves from dual fingers to a single finger, but then drops when the probe moves back to the dual finger region. Both of the regions have similar finger crease patterns (see representative frames 46 and 63 in the figure), and so we conclude, unsurprisingly, that ProbeSight performs better when more features (e.g., coarse creases, edges, and skinfolds) are visible.

93

FPD results are shown in Figure 6-4(B). The accumulated error increases when the probe moves from the proximal phalanx to the distal phalanx, but decreases when the probe goes back towards the proximal phalanx. The proximal and middle phalanx contain strong and clear creases, more than the distal phalanx (see representative frames 77, 93 and 102 in the figure). Again, we conclude that ProbeSight performs better given more of these features.

PML results are shown in Figure 6-4(C). The accumulated error in the hypothenar region is much lower than in the thenar+thumb region. Looking at frames 43 and 99, we observe much less coarse creasing in the thenar region compared with hypothenar region. ProbeSight shows poor tracking ability without these coarse creases.

PPD results are shown in Figure 6-4(D). Since both distal palm and proximal palm regions contain similarly strong crease patterns, as shown in example frames 77 and 119, ProbeSight has low accumulated error when tracking in both regions.

WPD results are shown in Figure 6-4(E). The accumulated error markedly increases when the probe moves from distal to proximal across the wrist region. Looking at frames 49 and 69, we observe that neither region has sufficient coarse creases, while the proximal wrist region contains almost no coarse creases. This explains why the accumulated error in both regions are very high, with the proximal wrist region having the highest error.
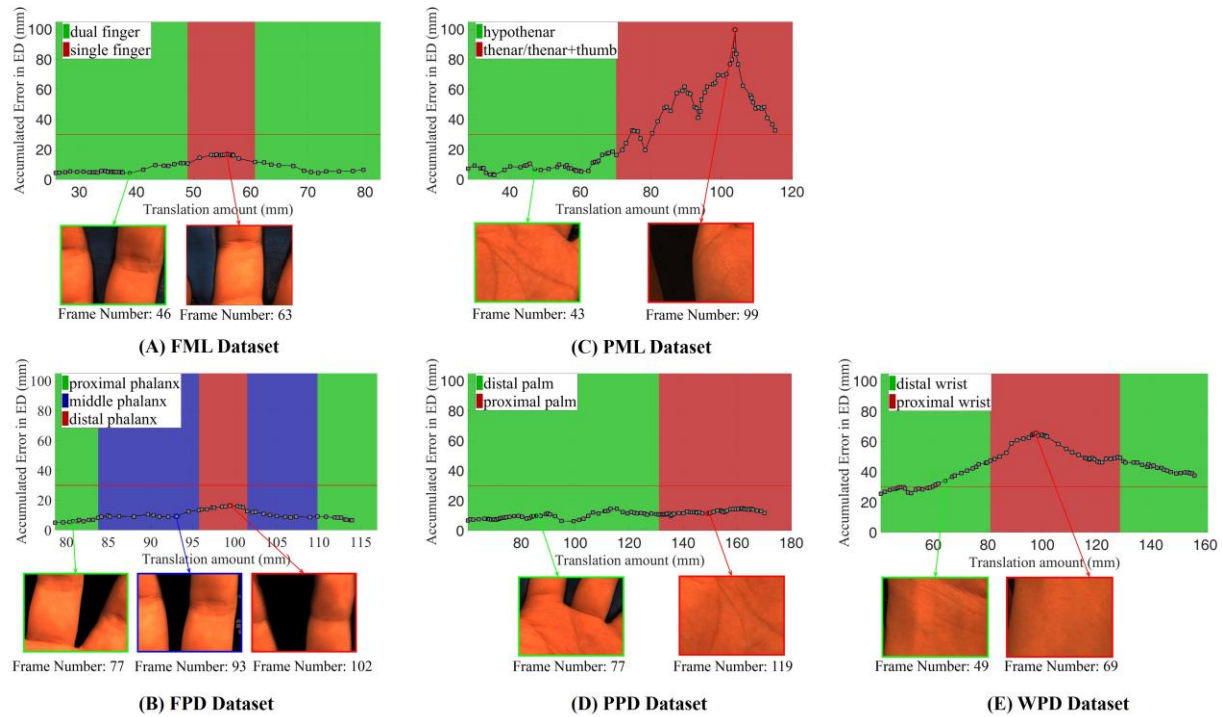
Figure 6-4: ProbeSight performance when tracking in different regions.

Figure 6-5 is a graph from (Sun 2014), showing the motion errors in translation of Sun's probe tracking system during a freehand scan on a lower-leg region. His system appears to suffer from a drift problem, judging from the accumulated error in the graph. He also mentions in the paper, that "the drifting effect due to accumulated errors is observable, which is common for dead-reckoning approaches and is a shortcoming of the proposed system compared to optical or electromagnetic tracking" (Sun 2014). Comparing ProbeSight's experiment results, shown in Figure 6-4, with Sun's system, our system's tracking error only accumulates over the short term, because ProbeSight is able to self-correct based on the prior 3D surface map. This comparison is under similar anatomical structures (upper extremity vs. lower leg) with approximately the same probe travel distance (approximately 100 mm), and is quantified using the same motion error analysis method.
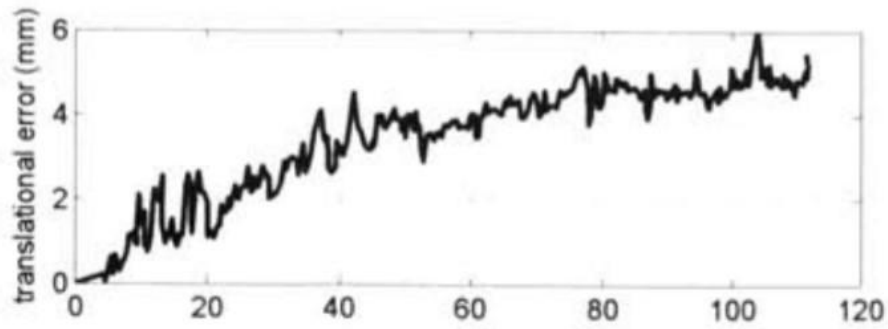
Figure 6-5: The motion errors in translation of Sun's tracking system during a freehand scan on a lower-leg region.

Table 6-1 uses average accumulated error to further illustrate how ProbeSight's performance varies across different regions. In general, the accumulated error increases in the regions where there are fewer clear features such as coarse creases and edges (e.g., thenar, distal wrist and proximal wrist regions). Conversely, ProbeSight's error generally decreases when viewing regions with clear coarse creases and strong edges (e.g., dual finger, proximal phalanx and distal palm regions). In extreme cases, ProbeSight's error may drift outside OpenDR's convergence range, requiring its estimated pose to be reset before ProbeSight can regain tracking, even when it moves back to a relatively easy tracking region.

Datasets	Finger				
Experiments	FML		FPD		
Regions	Single Finger	Dual Finger	Proximal Phalanx	Middle Phalanx	Distal Phalanx
Average Accumulated Error (mm)	15.627	6.628	6.961	9.952	14.847

Datasets	Palm				Wrist	
Experiments	PML		PPD		WPD	
Regions	Hypo Thenar	Thenar	Distal Palm	Proximal Palm	Distal Wrist	Proximal Wrist
Average Accumulated Error (mm)	8.926	51.659*	9.757	12.571	54.818*	37.248*

*✱ *Average accumulated error is above defined thresholding value*

We have labeled regions in which ProbeSight's average accumulated tracking error is above 30 mm to be challenging regions, while we labeled other regions as unchallenging regions. This 30mm error threshold is graphed as a horizontal red line in Figure 6-4.

In an attempt to improve ProbeSight's performance in challenging regions, we proposed a method to give ProbeSight a wider field of view, so as to always have more creases and features in view. Instead of directly center-cropping the 640×480 region from the camera's Bayer-pattern sensor, we first center-cropped a 1280 x 960 "double-size" region from the original camera image and then downsampled it to 640×480 pixels before preprocessing (as previously described in section 4.6.3). Using this larger FOV in challenging regions incorporates more features and edges from near-by regions than with the camera's original smaller FOV.

Figure 6-6 compares the results when applying our original cropping method (line with square markers) vs. down-sampling a wider view (line with triangle markers) on the same

datasets. The average accumulated error is reduced, especially when looking at challenging regions.

In the *hand medial-lateral* (HML) dataset, the probe starts from the unchallenging proximal palm region, moves toward the thenar region and then moves back again. As Figure 6-6(A) shows, the reduction in error is greatest in the challenging thenar region. Looking at camera frame 36 in Figure 6-6(A), in the proximal palm region both methods have ample coarse creases to analyze. These creases provide sufficient features during the OpenDR optimization. However, when the probe moves to the thenar region, the image processed by the cropping method (in the dashed box of frame 99) has insufficient coarse creases and also lacks prominent edges. By comparison, the image processed by the down-sample method (in the solid box of frame 99) incorporates coarse creases and edges from nearby regions, improving tracking performance.

The results on the *hand proximal-distal* (HPD) experiment can be explained in the same manner. When the probe starts from the unchallenging distal palm region as frame 20 shows, images processed by both methods contain coarse creases and strong edges. Therefore, in Figure 6-6(B), there is no significant difference between these two methods. When the probe moves to the challenging wrist area, however, the image processed by the down-sample method (in the solid box of frame 65) contains more edges and parts of coarse creases from the proximal palm region, reducing the accumulated error in comparison to the non-downsampled cropped image (in the dashed box of frame 65).
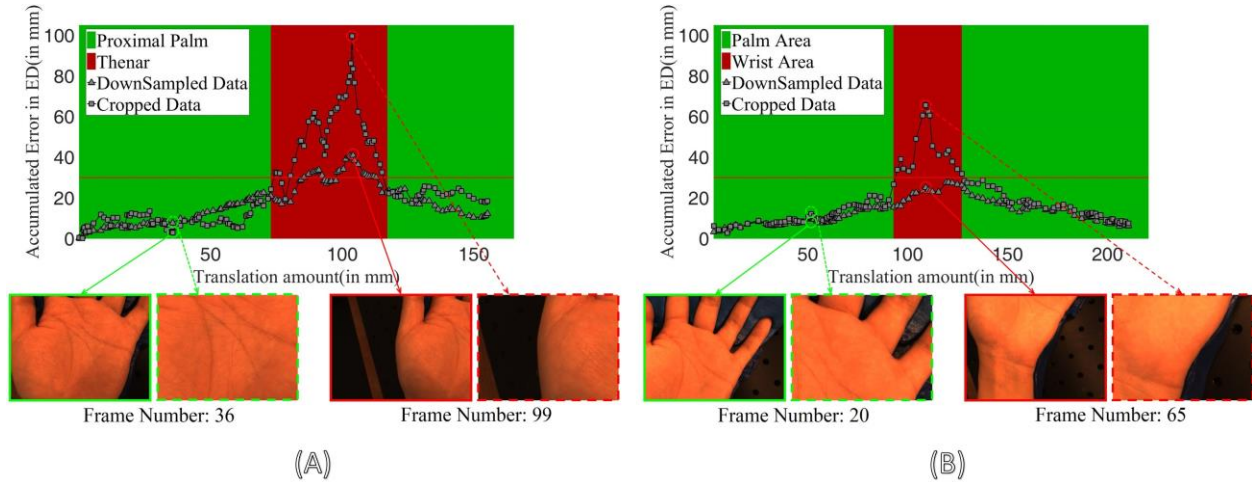
Figure 6-6: Performance comparison of ProbeSight when cropping vs. down-sampling a wider view before preprocessing. **(A)** Accumulated error in HML experiment. **(B)** Accumulated error in HPD experiment.

We used a Student's unpaired two-tailed t test to further compare down-sampling's improvement in challenging vs. unchallenging regions. Table 6-2 and Table 6-3 illustrate that on both experiments there are significant differences in the measurements of the Normalized Absolute Difference (NAD) of the accumulated error in the two datasets between challenging (green in table) and unchallenging regions (red in table). The Normalized Absolute Difference (NAD) of accumulated error across the frames of the two datasets can be computed as

$$\Delta\varepsilon = \sum_{n \in \{\text{all camera-image frames so far in this region}\}} \frac{\left|error_{crop}(n) - error_{downsample}(n)\right|}{error_{downsample}(n)} \qquad \textbf{Equation 6-1}$$

99

Table 6-2: *t-test* results by comparing normalized absolute difference of cropping and downsampling at proximal (unchallenging) and thenar (challenging) regions on HML Experiment

Regions	Proximal Palm		Thenar		
	M	*SD*	*M*	*SD*	*p*
NAD	0.487	0.365	0.773	0.267	<.0001

Note. Mean (M), Standard Deviation (SD), and results of two-tailed t test assuming unequal variances.
n=60 for proximal palm region and n=40 for thenar region.

Table 6-3: *t-test* results by comparing normalized absolute difference of cropping and down-Sample method at palm and wrist regions on HPD Experiment

Regions	Palm		Wrist		
	M	*SD*	*M*	*SD*	*p*
NAD	0.185	0.150	0.976	0.485	<.0001

Note. Mean (M), Standard Deviation (SD), and results of two-tailed t test assuming unequal variances.
n=50 for palm region and n=15 for wrist region.

From the graph and statistical analysis, we conclude that down sampling from a wider FOV improves performance in all cases, but the improvement is significantly larger for challenging regions.

By analyzing the results qualitatively, it can be seen that with the assistance of the INS, OpenDR is able to use a variety of skin features (e.g., coarse creases, edges, and skinfolds) to iteratively match the prior 3D surface map with a real-time 2D camera image, resulting in an accurate pose estimation in all regions of a real human subject's upper extremity.

7.0   REAL-TIME 3D INTERACTIVE VISUALIZATION FOR MEDICAL DATA

The many impressive developments in medical imaging technology during the last decades provided physicians with an increasing amount of high-resolution, patient-specific anatomical and functional data. In addition, the increasing use of non-ionizing real-time imaging, in particular ultrasound and optical imaging, during surgical procedures, has created the need for design and development of new visualization and display technology allowing physicians to take full advantage of rich sources of heterogeneous preoperative and intraoperative data (Sielhorst, Feuerstein, and Navab 2008). It has long been recognized that computer graphics might be an effective means for presenting medical data to the clinician (Fuchs, Levoy, and Pizer 1989). Various methods have been developed for rendering 3D data onto a stationary 2D display, with or without special hardware for stereovision, such as Lorensen's marching cubes algorithm for rendering the surfaces of pre-segmented 3D objects and Levoy's conversion of intensity gradient to rendered opacity and surface normal (Lorensen and Cline 1987; Levoy 1988). In the meantime, the advent of larger computer memories and faster processors has spawned a period of rapid growth in software (e.g., OpenGL, VTK and ITK) and hardware techniques (e.g., CUDA hardware (GPU)) for data visualization.

An important challenge in visualization is how to superimpose the display of multiple datasets. One solution is to apply image overlay, which is a computer display technique that superimposes images over the user's direct view of the real world. The images are transformed

in real-time so they appear to the user to be an integral part of the surrounding environment. By using image overlay with medical data such as CT, MRI reconstructions or ultrasound images, a surgeon can visualize the data 'in-situ', exactly positioned within the patient's anatomy, and potentially enhance the surgeon's ability to perform a complex procedure (Blackwell et al. 2000). The underlying concept of ProbeSight is somewhat similar, except that the user's direct view of the world is replaced with the real-time camera, and the medical data (in this case, ultrasound) is merged with the video first in the computer rather than the human mind. Nonetheless, it is subsequently essential to communicate these fused data to the human operator using some form of visualization.

In this chapter, the real-time interactive visualization system of ProbeSight is introduced and demonstrated.

7.1   REGISTRATION OF 3D ANATOMY WITH ULTRASOUND DATA

As discussed in section 4.3, the relative pose between anatomical model and the camera can be computed via the Micron optical tracker on the initial frame. If the optical tracker's origin is defined as the origin of global coordinates, we can render a 3D scene based on its coordinate system C_C. We call this rendering window the global view window. With a known geometry of the camera and the probe, along with the pre-measured transform between them, we can register the 3D anatomy with both the camera and probe in the global view window. Since the size (depth and width) of the US beam plane is known, and since it is always perpendicular to the long axis of probe surface, as Figure 7-1 shows, we can further use the graphical method of *texture mapping* to render the ultrasound slice superimposed at its correct pose in the window.

102
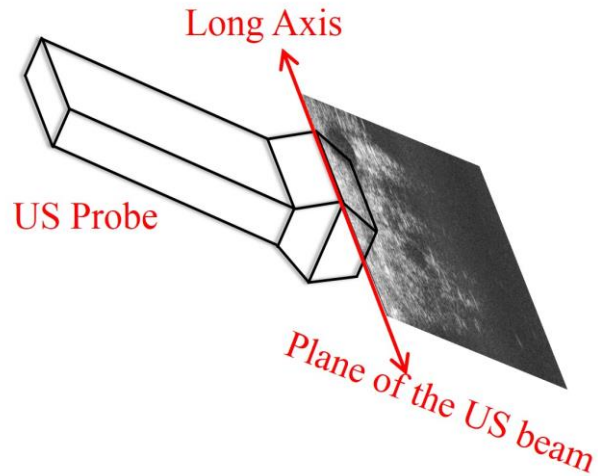
Figure 7-1: Illustration of relative pose between US beam plane and US probe.

7.2   THE OPENGL IMPLEMENTATION OF THE VISUALIZATION SYSTEM

We implemented our system for rendering a merged display in OpenGL, the standard graphics library available with hardware acceleration on most computers. OpenGL's flexibility permits the definition of arbitrary coordinate systems for graphics. There are two OpenGL real-time rendering windows in the current system. One is the *global view window* discussed in section 7.1, and the other is the *local view window*, which simulates what the actual view from the camera should look like. Both windows use perspective projection to produce a more realistic looking image.

For the global view window, ProbeSight provides three view modes and can be switched in real time using keyboard control. The different view modes are shown in Figure 7-2(A)(B) and (C), described as follows: (A) Front view: This mode simulates the view from the point of

view of the operator performing the ultrasound study with ProbeSight. All probe movements (rotation and translation) relative to the patient's anatomy correspond to what the operator would be seeing while performing a freehand scan. (B) Top view: This mode simulates the view from the top of ProbeSight, providing an overlay image with the underlying ultrasound slice fused with a transparent 3D anatomy. (C) Following view: This mode simulates a view from beneath the anatomical surface along the normal vector to the US image plane. Thus in this last mode, no matter how the user steers the transducer during the scan, the line of sight of view is always perpendicular to the US image plane, which provides a "following" view along the probe's trace. For all the view modes described above, the operator can interact with the scene by using mouse to zoom, rotate, and translate the view port, which gives maximum flexibility to observe each component of the scene. The views are available in real time during the operation of ProbeSight, or may be replayed after the scan. This last feature is particularly novel and important for ultrasound, which normally does not store the 3D anatomical context of its image data.

For the local view window, the FOV is set to that of the real camera. All coordinates are transformed from anatomical coordinates C_A to camera coordinates C_C. Using the pre-calibrated camera intrinsic parameters (camera matrix and distortion coefficients), we can thus render the 3D scene to simulate what would be seen by the real camera. The corresponding local view window at the particular probe poses in Figure 7-2(A)(B) and (C) are shown in Figure 7-2(D)(E) and (F).
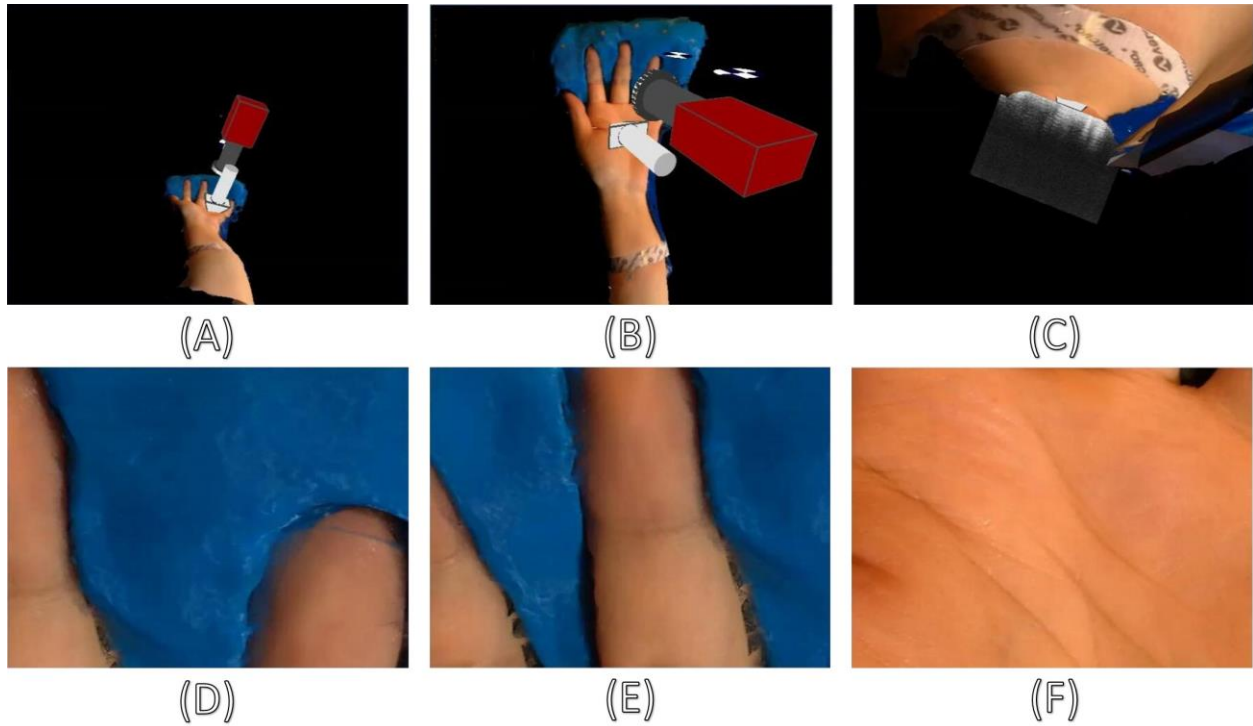
Figure 7-2: Upper row: The global view windows (A) in front view mode. **(B)** in top view mode. **(C)** in following view mode. **Lower row:** The corresponding local view windows.

At present, the entire 3D interactive visualization system is guided by optical tracker, since the OpenDR image optimization is currently being processed offline. We are working on translating our algorithms from python to C++ to make ProbeSight a fully real-time system, in which translation computed by the computer vision algorithms and rotation determined by INS will be directly applied to determine the probe pose in real time.

We believe combining ProbeSight tracking with 3D interactive visualization will provide a profound contribution to assist surgeons to quickly navigate the probe to the correct pose in the follow-up US studies. Our 3D interactive visualization system is not limited to US probe tracking, but could be extended to the guidance of surgical tools in general relative to the anatomy. Thus our system of augmented reality image overlay may enhance hand-eye

coordination in an intuitive and non-intrusive manner, improving the safety and effectiveness over a broad range of invasive procedures.

8.0   CONCLUSION

Ultrasound probe localization is essential for (1) volumetric imaging with a 2D ultrasound probe, (2) establishing a recorded anatomical context for US-guided surgery and (3) longitudinal studies. The existing techniques for probe localization, however, require external tracking devices, making them inconvenient for clinical use. In addition, the probe pose is typically measured with respect to a fixed coordinate system independent of the patient's anatomy, making it difficult to correlate ultrasound studies across time.

This dissertation describes the development and validation of our ProbeSight system, in which a self-navigating ultrasound probe operates in patient space, enabled by an onboard video camera from which images are matched with a prior high-resolution 3D model. The tracking performance of ProbeSight was validated through a series of phantom and in-vivo experiments. Overall, we have shown that ProbeSight tracks the probe precisely and robustly in different regions on both a phantom and an actual human upper extremity. ProbeSight demonstrated accurate tracking in unchallenging regions as well as challenging regions, even after temporarily accumulating small amounts of error. Combining ProbeSight tracking with a real-time interactive visualization system provides an anatomical context for the ultrasound data in terms of the exterior of the patient, permitting accurate anatomic localization of the ultrasound data in real time, as well as later for longitudinal ultrasound studies. Knowing the exact anatomic location from which images of nerves, arteries, ectopic ossifications, and other structures have

been obtained will greatly facilitate understanding their condition after injury and during healing, helping determine a timely and accurate treatment strategy. Such localized data can be correlated across time, as well as registered with other modalities, such as MRI diffusion tensor imaging of regenerating nerves in transplanted limbs. We may thus provide hand-transplant patients with an important new ability to follow nerve regeneration and chronic rejection.

8.1   INVENTIONS AND CONTRIBUTIONS

The inventions and contributions of ProbeSight, discussed in the preceding sections, can be summarized as follows:

(1) **Tracking natural skin and subdermal cues:** We developed an effective method to enhance skin features (section 4.6.1) so that we are able to efficiently track and match them with a prior high-resolution 3D map (section 4.2). Through in-vivo experiments on a real human upper extremity, the extracted natural skin features such as edges, fine ridges, and coarse creases were further analyzed (section 6.2). We further found that subdermal information such as veins, which are invisible in the original image, can be used as important cues for tracking once the image is enhanced. The distinctiveness of skin features in different regions of an upper extremely was also explored (Section 6.3).

(2) **Using a prior high-resolution 3D surface map for pose estimation:** Prior systems that use an on-board camera to track skin features have done so using frame-by-frame pose estimation, requiring that skin features be overlapping between neighboring frames (Rafii-Tari, Abolmaesumi, and Rohling 2011; Wang et al. 2012; Sun,

Gilbertson, and Anthony 2013). From this fact arise two problems. One is that the probe cannot move fast, otherwise there will be fewer overlapping features between neighboring frames, potentially causing tracking to fail. The other issue is error accumulation over extended periods of tracking. ProbeSight overcomes these problems by using a known 3D surface model established a-priori. The tracking error does not accumulate and the probe's speed of motion is not restricted to keep particular features in view from one frame to the next (section 6.3). To our knowledge, ProbeSight is the first such system to include a pre-scanned 3D surface map for pose estimation.

(3) Applying an INS to distinguish rotation from translation: It is well known that camera pose estimation methods have difficulty distinguishing between small translations and small rotations (Dannilidis and Nagel, n.d.). This problem has previously contributed to accumulated error in other single-camera probe tracking systems, such as Sun et al. presented (Sun, Gilbertson, and Anthony 2014; Sun 2014). In our system, we fused hardware-INS orientation data with translation estimated from the camera images (section 3.1). We have demonstrated that orientation measurement from the INS is more robust compared to the Micron optical tracker (section 5.2.6), and we discovered that OpenDR, while itself is not able to address the problem of distinguishing rotation and translation, shows increased accuracy if orientation data is instead acquired from the INS (section 5.2.5). As far as we know, ProbeSight is the first system to track an US probe by combining an INS with camera based pose estimation.

(4) Validating probe tracking on the upper extremity of a human subject: We have systematically evaluated ProbeSight on three datasets from the finger, palm, and wrist, and the corresponding results show that ProbeSight successfully tracks the probe in different regions in these datasets, without requiring any external tracking device. For tracking in challenging regions, we have proposed using a wider field of view (in this instance via down-sampling instead of cropping) to incorporate more features from nearby regions, reducing the accumulated error, allowing significantly better probe tracking (section 6.3).

8.2   LIMITATIONS

In this section, the limitations of the current ProbeSight system are summarized, and potential remedies are described.

8.2.1   Skin Deformation

Although ProbeSight has broad potential uses across clinical ultrasound, we have chosen a compelling niche for our initial application: the monitoring of rejection after hand transplant surgery. The upper limb region is relatively rigid, so we are not overly concerned about skin deformation in the current system. However, if ProbeSight is further applied on other regions, such as upper leg, abdomen, or particularly, the breast, the issue of skin deformation due to probe contact would be relatively large and need to be considered. Since a static 3D surface model is

acquired before the US study, the skin deformation caused by probe contact would not be reflected in the model. Therefore, we would expect an error caused by deformation when ProbeSight tracks on these regions. A potential solution would be to directly use the ultrasound image to correct the skin deformation by applying a deformable registration framework. OpenDR might also be extended to enable solving for deformation of the 3D surface map. Another solution is to develop a force-sensing ultrasound probe, as presented by Gilbertson Sun et al., who mounted force sensors on the probe so that the device is able to measure forces to more accurately estimate position relative to the skin surface (Gilbertson and Anthony 2015).

8.2.2 Probe Initialization and guidance without the optical tracker

As discussed in section 5.1.2, for the current ProbeSight system, we are using the Micron optical tracker to initialize ProbeSight's pose for the first frame only, avoiding for now the problems associated with the large search space for the initial pose. In order to completely exclude the tracker in a real clinical application, a probe initialization method needs to be developed to find the initial pose from which to render the prior 3D model to match the real-time camera image. This could be realized by requiring the probe be initially placed at a known position with particularly clear features in the anatomic atlas, such as those upper extremity creases shown in Figure 6-3 in Chapter 6.0 . Another possible approach is to guide the probe to an initial pose via the 3D visualization system. If we initialize the probe once at the beginning of an ultrasound study using the Micron optical tracker, we could able to render a wireframe virtual probe at that location in the 3D visualization of the anatomy. For following studies, the clinician could be instructed to navigate the probe so that it's visualized rendering fits the virtual wireframe probe, under the guidance of the visualization system. The same routine could be applied to navigate

the probe to the same target as a previous scan. Once the surgeon identifies a target in the initial study, the corresponding optimal probe pose is recorded and rendered relative to the anatomy. In a subsequent follow-up study, the wireframe representation in the visualization system at the same location could enable quick guidance of the probe to the same pose to view the original target in a corresponding US image for comparison.

8.2.3   Defocus aberration

The fixed-focus camera in our current hardware set-up of ProbeSight has a limited depth of focus, as discussed in section 3.1, and defocus aberration can cause severe blur in the camera image when the probe is titled beyond a certain range of angles. Figure 8-1 shows two probe tilting conditions with corresponding camera images. The camera image obtains clear focus with all coarse and fine creases visible when the probe has a small tilt angle, but many creases are obscured due to blur when the probe is tilted at a large angle.

To address the defocus aberration issue, an auto-focus camera could be used on the next version of ProbeSight. The corresponding algorithm would need to be adapted to dynamically compensate for both the image center and focal length changes caused by the auto focus, so that the camera model could maintain accuracy during automatic focus changes.
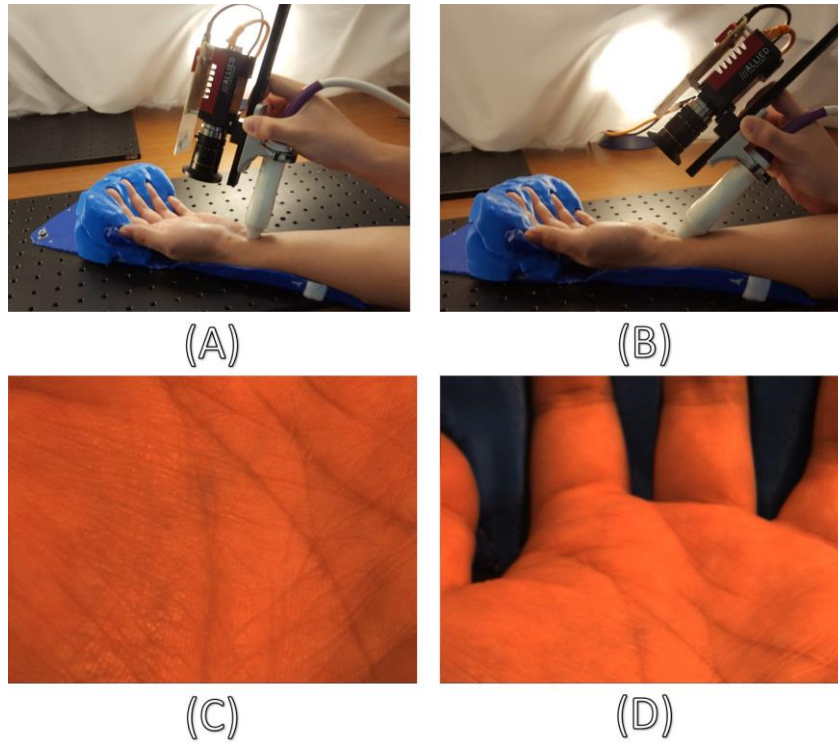
Figure 8-1: Upper row: The illustration of two probe tilting conditions. **(A)** Small tilt (Less than 15 degree). **(B)** Large tilt (Larger than 45 degree). **Lower row:** The corresponding camera image under the two conditions. Many more creases are visible in **(C)** than in **(D).**

8.3   OUR LAB'S RELATED RESEARCH AND FUTURE WORK

In this section, ProbeSight-related research by others in our laboratory is first discussed and then directions for future work are outlined for this combined effort.

8.3.1   3D ultrasound volume reconstruction using external tracking

In a fully functional ProbeSight system, the tracking developed in this thesis will be combined with the 3D ultrasound volume reconstruction techniques developed by Vikas Shivaprabhu, who received his Ph.D. in our laboratory in 2015. In that work, the location and orientation of the probe is determined by the external Micron optical tracker with a marker mounted on the US probe, as Figure 8-2(A) shows. The 3D reconstruction process is described as follows. For each B-scan, individual 2D images are stamped with both the time of acquisition and the location of images obtained from the tracking system. A 3D volume is then reconstructed by placing the 2D images within a 3D space based on the tracking information (Figure 8-2(B)). When the 2D images are consolidated into a 3D space, a particular voxel in the 3D volume may either be intersected by pixels from more than one 2D image or may not be intersected by any scan. As suggested in (Rohling, Gee, and Berman 1999), the former problem, known as bin-filling, can be solved by combining data in the overlapping pixels (compounding). The latter problem, known as hole-filling, can be solved by inferring values for the missing data, using information of the voxel's neighbors (interpolation). The algorithm employs bin-filling and hole-filling techniques described in (Dewi et al. 2009) and (Gobbi and Peters 2002). In order to correctly localize the data captured, temporal and spatial calibration are required, as the method described in (Rousseau, Hellier, and Barillot 2006). Temporal calibration finds the latency between the image acquisition and the tracking system, and spatial calibration finds the transformation between pixels in the 2D image and the location of the tracked marker on the probe in 3D space. The system used an established N-wire phantom developed specifically for this purpose (Chen et al. 2009). The algorithms for calibration, image acquisition, and volume reconstruction use the Public Software Library for Ultrasound (PLUS) toolkit (Lasso, Heffter, and Pinter 2012).

Reconstructed image slices that correspond to the current location of the US probe may then be retrieved from previously stored US data. To illustrate this, an US phantom containing tubing is used to simulate vasculature (Blue Phantom, Inc., FL, USA). The phantom is tracked with reference markers. Figure 8-2(D) shows an image slice retrieved from a reconstructed 3D US volume of the phantom corresponding to the live US image seen in Figure 8-2(C).
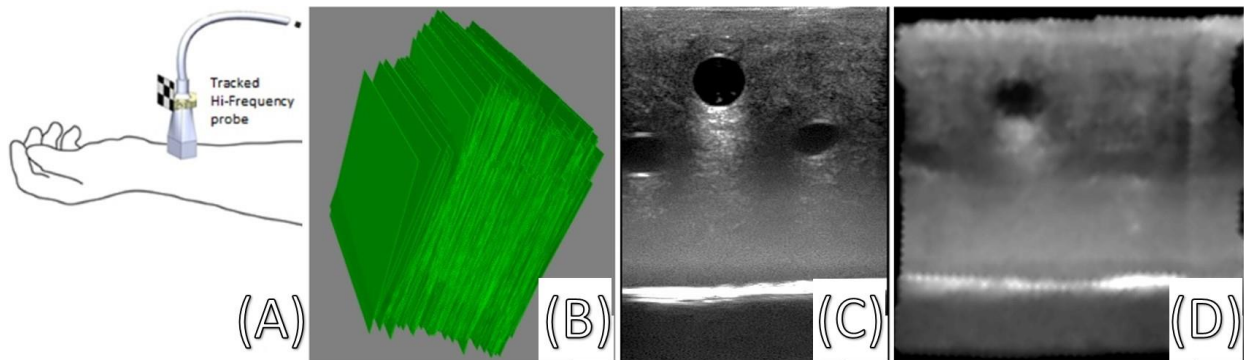
Figure 8-2: **(A)** US probe tracked by markers mounted on it. **(B)** 3D model of individual 2D US images displayed in a 3D space based on the recorded location and orientation of the US probe. **(C)** The Live US image. **(D)** US image slice retrieved from a previously reconstructed US volume corresponding to the live US image.

In the future work, the Micron optical tracking system will be replaced by the ProbeSight tracking system described in the present thesis, providing the benefits of a self-contained anatomically-based system. Further work will also combine data from the reconstructed 3D ultrasound volume into the 3D visualization system discussed on Chapter 7.0 making it possible for the surgeon to make use of that data in its correct 3D anatomical context for guidance and comparison with real-time US data.

8.3.2   HFUS in-plane registration of deformable finger vessels

Our initial clinical application involves hand transplants, which calls for longitudinal studies of finger vasculature to detect chronic rejection. Due to the complex internal structure of the finger and the inherent noise of HFUS, it is difficult to register 2D images of finger vascular tissue, especially under deformation. Although mono-modality ultrasound registration frameworks have been developed over the past few decades, few of them focused on small anatomical structures such as fingers or were validated using a HFUS imaging system (FOROUGHI, ABOLMAESUMI, and HASHTRUDIZAAD 2006).

Several other graduate students in our lab (Tejas Mathai and Chengqian Che) are developing an in-plane registration framework for HFUS of deformable finger tissue. We have studied a variety of similarity measurements with different pre-processing techniques to find which registration metrics are best suited for HFUS vessel tracking. The overall best performance obtained so far uses a normalized correlation metric coupled with HFUS downsampling and a one-plus-one evolutionary optimizer, to yield a mean registration error of 0.05 mm. We are currently using HFUS to study how finger tissue deforms under the ultrasound transducer by comparing internal motion vs. transducer motion. This will potentially address the deformation issue of ProbeSight, mentioned in section 8.2.1. Improving HFUS registration and tissue modeling may lead to new basic image analysis research and benefit treatments for peripheral vascular disorders.

8.3.3 Future directions

Future directions of ProbeSight will involve overcoming skin deformation (section 8.2.1), developing a complete navigation system to guide the probe to the same pose used for a particular target in a previous scan (section 8.2.2), possibly mounting lights on the probe, and using a smaller camera with either an auto-focus lens or a larger depth of focus (section 8.2.3).

Toward the clinical application of ProbeSight, we will develop automatic initialization technology as discussed in section 8.2.2, so as to no longer need the optical tracker, making ProbeSight a fully self-contained device. We will also convert the OpenDR pipeline code from python to C++ and CUDA, to make ProbeSight a fully real-time system.

Eventually, we want to merge ProbeSight tracking with 3D ultrasound volume reconstruction and analysis, and include the ultrasound data in the 3D visualization system to provide a system that can combine the ultrasound and visual data in both the mind of the operator and in the software of the computer. We then intend to explore how the human and the machine can help each other to understand what each other is seeing, enabling more holistic awareness for the benefit of the patient.

BIBLIOGRAPHY

Adelson, E H, C H Anderson, J R Bergen, P J Burt, and J M Ogden. 1984. "Pyramid Methods in Image Processing." *RCA Engineer* 29 (6): 33–41. doi:10.1.1.59.9419.

Apel, Peter J., Jianjun Ma, Michael Callahan, Casey N. Northam, Timothy B. Alton, William E. Sonntag, and Zhongyu Li. 2010. "Effect of Locally Delivered IGF-1 on Nerve Regeneration during Aging: An Experimental Study in Rats." *Muscle & Nerve* 41 (3): 335–41. doi:10.1002/mus.21485.

Birkfellner, W, F Watzinger, F Wanschitz, G Enislidis, C Kollmann, D Rafolt, R Nowotny, R Ewers, and H Bergmann. 1998. "Systematic Distortions in Magnetic Position Digitizers." *Medical Physics* 25 (11): 2242. doi:10.1118/1.598425.

Blackall, J. M., D Rueckert, C. R. Maurer, G. P. Penney, D. L. G. Hill, and D. J. Hawkes. 2000. "An Image Registration Approach to Automated Calibration for Freehand 3D Ultrasound." In *Medical Image Computing and Computer-Assisted Intervention*, 462–71. doi:10.1007/978-3-540-40899-4_47.

Blackwell, M, C Nikou, a M DiGioia, and T Kanade. 2000. "An Image Overlay System for Medical Data Visualization." *Medical Image Analysis* 4 (1): 67–72. doi:10.1016/S1361-8415(00)00007-4.

Boctor, Emad M, Gabor Fichtinger, Russell H Taylor, and Michael A Choti. 2003. "Tracked 3D Ultrasound in Radio-Frequency Liver Ablation." In *Medical Imaging 2003: Ultrasonic Imaging and Signal Processing, Proceedings of SPIE*, edited by William F. Walker and Michael F. Insana, 5035:174. doi:10.1117/12.479952.

Bouguet, Jean-yves. 1999. "Visual Methods for Three-Dimensional Modeling." California Institute of Technology.

Brown, Duane C. 1966. "Decentering Distortion of Lenses." *Photogrammetric Engineering* 32: 444–62.

Cash, C J C, a M Sardesai, L H Berman, M J Herrick, G M Treece, R W Prager, and a H Gee. 2005. "Spatial Mapping of the Brachial Plexus Using Three-Dimensional Ultrasound." *The British Journal of Radiology* 78 (936): 1086–94. doi:10.1259/bjr/36348588.

Chan, Candice, Felix Lam, and Robert Rohling. 2005. "A Needle Tracking Device for Ultrasound Guided Percutaneous Procedures." *Ultrasound in Medicine & Biology* 31 (11): 1469–83. doi:10.1016/j.ultrasmedbio.2005.07.014.

Chen, Thomas Kuiran, Adrian D. Thurston, Randy E. Ellis, and Purang Abolmaesumi. 2009. "A Real-Time Freehand Ultrasound Calibration System with Automatic Accuracy Feedback and Control." *Ultrasound in Medicine & Biology* 35 (1): 79–93. doi:10.1016/j.ultrasmedbio.2008.07.004.

Dannilidis, K., and H.H. Nagel. n.d. "The Coupling of Rotation and Translation in Motion Estimation of Planar Surfaces." In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 188–93. IEEE Comput. Soc. Press. doi:10.1109/CVPR.1993.340990.

de La Gorce, Martin, Nikos Paragios, and David J. Fleet. 2008. "Model-Based Hand Tracking with Texture, Shading and Self-Occlusions." In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. IEEE. doi:10.1109/CVPR.2008.4587752.

Dekker, Don L., Robert L. Piziali, and Eugene Dong. 1974. "A System for Ultrasonically Imaging the Human Heart in Three Dimensions." *Computers and Biomedical Research* 7 (6): 544–53. doi:10.1016/0010-4809(74)90031-7.

Dewi, D. E. O., M. H. F. Wilkinson, T. L. R. Mengko, I. K. E. Purnama, P. M. A. van Ooijen, A. G. Veldhuizen, N. M. Maurits, and G. J. Verkerke. 2009. "3D Ultrasound Reconstruction of Spinal Images Using an Improved Olympic Hole-Filling Method." In *International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering 2009*, 1–5. IEEE. doi:10.1109/ICICI-BME.2009.5417235.

Duric, Nebojsa, Peter Littrup, Lou Poulo, Alex Babkin, Roman Pevzner, Earle Holsapple, Olsi Rama, and Carri Glide. 2007. "Detection of Breast Cancer with Ultrasound Tomography: First Results with the Computed Ultrasound Risk Evaluation (CURE) Prototype." *Medical Physics* 34 (2): 773. doi:10.1118/1.2432161.

Eberly, D., R. Gardner, B. Morse, S. Pizer, and C. Scharlach. 1994. "Ridges for Image Analysis." *Journal of Mathematical Imaging and Vision* 4 (4): 353–73. doi:10.1007/BF01262402.

Fischler, Martin a., and Robert C. Bolles. 1981. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Communications of the ACM* 24 (6): 381–95. doi:10.1145/358669.358692.

Flaccavento, Giselle, Peter Lawrence, and Robert Rohling. 2004. "Patient and Probe Tracking During Freehand Ultrasound." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 585–93. Springer Berlin Heidelberg. doi:10.1007/978-3-540-30136-3_72.

FOROUGHI, P, P ABOLMAESUMI, and K HASHTRUDIZAAD. 2006. "Intra-Subject Elastic Registration of 3D Ultrasound Images." *Medical Image Analysis* 10 (5): 713–25. doi:10.1016/j.media.2006.06.008.

Fried, Nathaniel M., and Arthur L. Burnett. 2015. "Novel Methods for Mapping the Cavernous Nerves during Radical Prostatectomy." *Nature Reviews Urology* 12 (8): 451–60. doi:10.1038/nrurol.2015.174.

Fuchs, H, M Levoy, and SM Pizer. 1989. "Interactive Visualization of 3D Medical Data," 1–9. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA208091.

Gauch, J.M., and S.M. Pizer. 1993. "Multiresolution Analysis of Ridges and Valleys in Grey-Scale Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15 (6): 635–46. doi:10.1109/34.216734.

Ghannam, Sherin, and AL Abbott. 2013. "Cross Correlation versus Mutual Information for Image Mosaicing." *International Journal of Advanced Computer Science and Applications* 4 (11): 94–102.

Gilbertson, Matthew W., and Brian W. Anthony. 2015. "Force and Position Control System for Freehand Ultrasound." *IEEE Transactions on Robotics* 31 (4): 835–49. doi:10.1109/TRO.2015.2429051.

Girlanda, Raffaele. 2013. "Complications of Post-Transplant Immunosuppression." In *Regenerative Medicine and Tissue Engineering*. InTech. doi:10.5772/55614.

Gobbi, David G., Roch M. Comeau, and Terry M. Peters. 1999. "Ultrasound Probe Tracking for Real-Time Ultrasound/MRI Overlay and Visualization of Brain Shift." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 920–27. Springer Berlin Heidelberg. doi:10.1007/10704282_100.

Gobbi, David G, and Terry M Peters. 2002. "Interactive Intra-Operative 3D Ultrasound Reconstruction and Visualization." *Medical Image Computing and ComputerAssisted Intervention—MICCAI 2002* 2489 (2489): 156–63. doi:10.1007/3-540-45787-9_20.

Gorantla, Vijay S., and Anthony J. Demetris. 2011. "Acute and Chronic Rejection in Upper Extremity Transplantation: What Have We Learned?" *Hand Clinics* 27 (4): 481–93. doi:10.1016/j.hcl.2011.08.006.

Hager, G.D., and P.N. Belhumeur. 1996. "Real-Time Tracking of Image Regions with Changes in Geometry and Illumination." In *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 403–10. IEEE Comput. Soc. Press. doi:10.1109/CVPR.1996.517104.

Han, Lianghao, J.Alison Noble, and Michael Burcher. 2003. "A Novel Ultrasound Indentation System for Measuring Biomechanical Properties of in Vivo Soft Tissue." *Ultrasound in Medicine & Biology* 29 (6): 813–23. doi:10.1016/S0301-5629(02)00776-7.

Hastenteufel, Mark, Marcus Vetter, Hans-Peter Meinzer, and Ivo Wolf. 2006. "Effect of 3D Ultrasound Probes on the Accuracy of Electromagnetic Tracking Systems." *Ultrasound in Medicine & Biology* 32 (9): 1359–68. doi:10.1016/j.ultrasmedbio.2006.05.013.

Huang, Q.H., Y.P. Zheng, M.H. Lu, and Z.R. Chi. 2005. "Development of a Portable 3D Ultrasound Imaging System for Musculoskeletal Tissues." *Ultrasonics* 43 (3): 153–63. doi:10.1016/j.ultras.2004.05.003.

Hummel, Johann, Michael Figl, Michael Bax, Helmar Bergmann, and Wolfgang Birkfellner. 2008. "2D/3D Registration of Endoscopic Ultrasound to CT Volume Data." *Physics in Medicine and Biology* 53 (16): 4303–16. doi:10.1088/0031-9155/53/16/006.

J.G. Fryer; D.C. Brown. 1986. "Lens Distortion for Close-Range Photogrammetry." *Photogrammetric Engineering and Remote Sensing* 52(1): 51–58.

Kaufman, Christina L., R. Ouseph, B. Blair, J. E. Kutz, T. M. Tsai, L. R. Scheker, H. Y. Tien, et al. 2012. "Graft Vasculopathy in Clinical Hand Transplantation." *American Journal of Transplantation* 12 (4): 1004–16. doi:10.1111/j.1600-6143.2011.03915.x.

Klimczak, Aleksandra, and Maria Siemionow. 2007. "Immune Responses in Transplantation: Application to Composite Tissue Allograft." *Seminars in Plastic Surgery* 21 (4): 226–33. doi:10.1055/s-2007-991192.

Kneip, Laurent, Roland Siegwart, and Marc Pollefeys. 2012. "Finding the Exact Rotation between Two Images Independently of the Translation." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7577 LNCS:696–709. doi:10.1007/978-3-642-33783-3_50.

Kostereva, Nataliya V., Yong Wang, Derek R. Fletcher, Jignesh V. Unadkat, Jonas T. Schnider, Chiaki Komatsu, Yang Yang, et al. 2016. "IGF-1 and Chondroitinase ABC Augment Nerve Regeneration after Vascularized Composite Limb Allotransplantation." Edited by William D Phillips. *PLOS ONE* 11 (6): e0156149. doi:10.1371/journal.pone.0156149.

Lasso, Andras, Tamas Heffter, and Csaba Pinter. 2012. "Implementation of the PLUS Open-Source Toolkit for Translational Research of Ultrasound-Guided Intervention Systems." *MICCAI-Systems and Architectures for Computer Assisted Interventions (2012): 1-12*.

Levoy, Marc. 1988. "Display of Surfaces from Volume Data." *IEEE Computer Graphics and Applications* 8 (3): 29–37. doi:10.1109/38.511.

Loper, Matthew M. "Chumpy." https://github.com/mattloper/chumpy/blob/master/optimization.py.

Loper, Matthew M., and Michael J. Black. 2014. "OpenDR: An Approximate Differentiable Renderer." In *European Conference on Computer Vision*, 8695 LNCS:154–69. Springer International Publishing. doi:10.1007/978-3-319-10584-0_11.

Lorensen, William E., and Harvey E. Cline. 1987. "Marching Cubes: A High Resolution 3D Surface Construction Algorithm." *ACM SIGGRAPH Computer Graphics* 21 (4): 163–69. doi:10.1145/37402.37422.

Lourakis, M.L.A., and A.A. Argyros. 2005. "Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment?" In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, II:1526–1531 Vol. 2. IEEE. doi:10.1109/ICCV.2005.128.

Lovasik, Darlene, Daniel E. Foust, Joseph E. Losee, W.P. Andrew Lee, Gerald Brandacher, and Vijay S. Gorantla. 2011. "Helping Hands: Caring for the Upper Extremity Transplant Patient." *Critical Care Nursing Clinics of North America* 23 (3): 505–17. doi:10.1016/j.ccell.2011.09.002.

Maes, F, A Collignon, D Vandermeulen, G Marchal, and P Suetens. 1997. "Multimodality Image Registration by Maximization of Mutual Information." *IEEE Transactions on Medical Imaging* 16 (2): 187–98. doi:10.1109/42.563664.

Mercier, Laurence, Thomas Langø, Frank Lindseth, and D. Louis Collins. 2005. "A Review of Calibration Techniques for Freehand 3-D Ultrasound Systems." *Ultrasound in Medicine & Biology* 31 (4): 449–71. doi:10.1016/j.ultrasmedbio.2004.11.015.

Mills, P.H., and H Fuchs. 1990. "3D Ultrasound Display Using Optical Tracking." In *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, 490–97. IEEE Comput. Soc. Press. doi:10.1109/VBC.1990.109361.

Moritz, W.E., and P.L. Shreve. 1976. "A Microprocessor-Based Spatial-Locating System for Use with Diagnostic Ultrasound." *Proceedings of the IEEE* 64 (6): 966–74. doi:10.1109/PROC.1976.10250.

Moritz, William E, Alan S Pearlman, Daniel H. Mc Cabe, Douglas K Medema, Mark E Ainsworth, and Mark S Boles. 1983. "An Ultrasonic Technique for Imaging the Ventricle in Three Dimensions and Calculating Its Volume." *IEEE Transactions on Biomedical Engineering* BME-30 (8): 482–92. doi:10.1109/TBME.1983.325151.

Morse, Bryan S., Stephen M. Pizer, and Alan Liu. 1994. "Multiscale Medial Analysis of Medical Images." *Image and Vision Computing* 12 (6): 327–38. doi:10.1016/0262-8856(94)90057-4.

Nakamoto, Masahiko, Kazuhisa Nakada, Yoshinobu Sato, Kozo Konishi, Makoto Hashizume, and Shinichi Tamura. 2008. "Intraoperative Magnetic Tracker Calibration Using a Magneto-Optic Hybrid Tracker for 3-D Ultrasound-Based Navigation in Laparoscopic Surgery." *IEEE Transactions on Medical Imaging* 27 (2): 255–70. doi:10.1109/TMI.2007.911003.

Pizer, S.M., R.E. Johnston, J.P. Ericksen, B.C. Yankaskas, and K.E. Muller. 1990. "Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness." In *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, 337–45. IEEE Comput. Soc. Press. doi:10.1109/VBC.1990.109340.

Pizer, Stephen M., E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. 1987. "Adaptive Histogram Equalization and Its Variations." *Computer Vision, Graphics, and Image Processing* 39 (3): 355–68. doi:10.1016/S0734-189X(87)80186-X.

Pizer, Stephen M., Christina A. Burbeck, James M. Coggins, Daniel S. Fritsch, and Bryan S. Morse. 1994. "Object Shape before Boundary Shape: Scale-Space Medial Axes." *Journal of Mathematical Imaging and Vision* 4 (3): 303–13. doi:10.1007/BF01254105.

Poon, Tony C., and Robert N. Rohling. 2007. "Tracking a 3-D Ultrasound Probe with Constantly Visible Fiducials." *Ultrasound in Medicine & Biology* 33 (1): 152–57. doi:10.1016/j.ultrasmedbio.2006.07.017.

Rafii-Tari, Hedyeh, Purang Abolmaesumi, and Robert Rohling. 2011. "Panorama Ultrasound for Guiding Epidural Anesthesia: A Feasibility Study." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6689 LNCS:179–89. doi:10.1007/978-3-642-21504-9_17.

Rohling, Robert, Andrew Gee, and Laurence Berman. 1999. "A Comparison of Freehand Three-Dimensional Ultrasound Reconstruction Techniques." *Medical Image Analysis* 3 (4): 339–59. doi:10.1016/S1361-8415(99)80028-0.

Rousseau, François, Pierre Hellier, and Christian Barillot. 2006. "A Novel Temporal Calibration Method for 3-D Ultrasound." *IEEE Transactions on Medical Imaging* 25 (8): 1108–12. doi:10.1109/TMI.2006.877097.

Sauer, F., Khamene, A. 2003. "Video-Assistance for Ultrasound Guided Needle Biopsy" 1 (12): 0–4. doi:10.1016/j.(73).

Schneeberger, Stefan, Vijay S. Gorantla, Gerald Brandacher, Adriana Zeevi, Anthony J. Demetris, John G. Lunz, Diana M. Metes, et al. 2013. "Upper-Extremity Transplantation Using a Cell-Based Protocol to Minimize Immunosuppression." *Annals of Surgery* 257 (2): 345–51. doi:10.1097/SLA.0b013e31826d90bb.

Shulman, H M, G E Sale, K G Lerner, E a Barker, P L Weiden, K Sullivan, B Gallucci, E D Thomas, and R Storb. 1978. "Chronic Cutaneous Graft-versus-Host Disease in Man." *The American Journal of Pathology* 91 (3): 545–70. http://www.ncbi.nlm.nih.gov/pubmed/26221.

Sielhorst, Tobias, Marco Feuerstein, and Nassir Navab. 2008. "Advanced Medical Displays: A Literature Review of Augmented Reality." *Journal of Display Technology* 4 (4): 451–67. doi:10.1109/JDT.2008.2001575.

Stoll, Guido, and Hans Werner Müller. 2006. "Nerve Injury, Axonal Degeneration and Neural Regeneration: Basic Insights." *Brain Pathology* 9 (2): 313–25. doi:10.1111/j.1750-3639.1999.tb00229.x.

Studholme, Colin, Derek L G Hill, and David J Hawkes. 1997. "Automated Three-Dimensional Registration of Magnetic Resonance and Positron Emission Tomography Brain Images by Multiresolution Optimization of Voxel Similarity Measures." *Medical Physics* 24 (1): 25. doi:10.1118/1.598130.

Sun, Shih-Yu. 2014. "Ultrasound Probe Localization by Tracking Skin Features." Massachusetts Institute of Technology.

Sun, Shih-Yu, Matthew Gilbertson, and Brian W. Anthony. 2013. "6-DOF Probe Tracking via Skin Mapping for Freehand 3D Ultrasound." In *2013 IEEE 10th International Symposium on Biomedical Imaging*, 780–83. IEEE. doi:10.1109/ISBI.2013.6556591.

Sun, Shih-Yu, Matthew Gilbertson, and Brian W. Anthony. 2014. "Probe Localization for Freehand 3D Ultrasound by Tracking Skin Features." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8674 LNCS:365–72. doi:10.1007/978-3-319-10470-6_46.

Svennigsen, Åsa, and Lars Dahlin. 2013. "Repair of the Peripheral Nerve—Remyelination That Works." *Brain Sciences* 3 (3): 1182–97. doi:10.3390/brainsci3031182.

Teather, Robert J., Andriy Pavlovych, Wolfgang Stuerzlinger, and I. Scott MacKenzie. 2009. "Effects of Tracking Technology, Latency, and Spatial Jitter on Object Movement." In *2009 IEEE Symposium on 3D User Interfaces*, 43–50. IEEE. doi:10.1109/3DUI.2009.4811204.

Tian, Mingzhi, Jihang Wang, Samantha Horvath, John Galeotti, Vijay Gorantla, and George Stetten. 2016. "Ridge Matching Based on Maximal Correlation in Transform Space." Pittsburgh.

Valenzuela, Nicole M., and Elaine F. Reed. 2011. "The Link between Major Histocompatibility Complex Antibodies and Cell Proliferation." *Transplantation Reviews* 25 (4): 154–66. doi:10.1016/j.trre.2011.04.001.

Viola, P., and W.M. Wells. 1995. "Alignment by Maximization of Mutual Information." In *Proceedings of IEEE International Conference on Computer Vision*, 24:16–23. IEEE Comput. Soc. Press. doi:10.1109/ICCV.1995.466930.

Wakefield, R. 1999. "The Current Status of Ultrasonography in Rheumatology." *Rheumatology* 38 (3): 195–98. doi:10.1093/rheumatology/38.3.195.

Wang, Jihang, Samantha Horvath, George Stetten, Mel Siegel, and John Galeotti. 2012. "Real-Time Registration of Video with Ultrasound Using Stereo Disparity." In *SPIE Medical Imaging*, edited by David R. Holmes III and Kenneth H. Wong, 83162D. doi:10.1117/12.911222.

Wang, Jihang, Vikas Shivaprabhu, John Galeotti, Samantha Horvath, Vijay Gorantla, and George Stetten. 2014. "Towards Video Guidance for Ultrasound, Using a Prior High-Resolution 3D Surface Map of the External Anatomy." In *Lecture Notes in Computer Science*, 51–59. doi:10.1007/978-3-319-10437-9_6.

Webb, Andrew, and George C. Kagadis. 2003. "Introduction to Biomedical Imaging." *Medical Physics* 30 (8): 2267. doi:10.1118/1.1589017.

Wells, W. 1996. "Multi-Modal Volume Registration by Maximization of Mutual Information." *Medical Image Analysis* 1 (1): 35–51. doi:10.1016/S1361-8415(96)80004-1.

Willson, Reg G., Mark W. Maimone, Andrew E. Johnson, and Larry M. Scherr. 2005. "An Optical Model for Image Artifacts Produced by Dust Particles on Lenses." *European Space Agency, (Special Publication) ESA SP* 2005 (603): 801–8.

Yan, Ying, Matthew R MacEwan, Daniel A Hunter, Scott Farber, Piyaraj Newton, Thomas H Tung, Susan E Mackinnon, and Philip J Johnson. 2013. "Nerve Regeneration in Rat Limb Allografts." *Plastic and Reconstructive Surgery* 131 (4): 499e–511e. doi:10.1097/PRS.0b013e31828275b7.