HIERACHICAL 3-D REGISTRATION OF
ULTRASOUND-COMPUTED TOMOGRAPHY OF
FEMUR SHAFT FRACTURE USING REINFORCEMENT LEARNING

By

XUXIN ZENG

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Biomedical Engineering

Written under the direction of

Ilker Hacihaliloglu

And approved by

_____

_____

_____

New Brunswick, New Jersey

JANUARY 2020

ABSTRACT OF THE THESIS

**HIERARCHICAL 3-D REGISTRATION OF
ULTRASOUND-COMPUTED TOMOGRAPHY OF
FEMUR SHAFT FRACTURE USING REINFORCEMENT LEARNING**

by XUXIN ZENG

Thesis Director: Ilker Hacihaliloglu

Femoral shaft fracture, a bone fracture that involves the femur, typically sustained in high-energy injuries, such as a car crash. Improper treatment of fixation or alignment may cause soft tissue injuries, bone loss and significant high risk on pulmonary compliance. Therefore, the treatments with surgical guidance are of importance in deducing the rate of compliance and improving the accuracy of the operation. Image-guided computer-assisted orthopedic surgery has been explored in improving the outcomes of the femoral shaft fracture treatments. And the domain intra-operative imaging modality fluoroscopy used in CAOS is 2D fluoroscopy. 3D anatomical representation from 2D fluoroscopy requires high volume of 2D data from different directions, which has pool reproductivities into 3D due to the limit field of view in 2D fluoroscopy. Furthermore, the increasing operation time with ionizing radiation exposure from fluoroscopy modality brings essential concerns for the safety of the surgeons and patients. Recently, ultrasound (US) has been investigated as an alternative intra-operative imaging modality due to its real-time, safe and 2D/3D imaging capabilities.

However, lower signal to noise ratio (SNR), imaging artifacts, limited field of view (FOV) and blurred bone boundaries have hindered wide spread adaption of US in CAOS. In order to overcome these limitations, automatic bone segmentation and intra-operative

registration methods have been developed. Accurate, robust and real-time segmentation and registration is necessary for successful guidance in US-based CAOS. This thesis presents an automated hierarchical registration method using reinforcement learning, for accurate, robust and real-time registration of intra-operative US to pre-operative CT data.

The proposed framework consists of: (1) bone shadow region image enhancement and segmentation, (2) point cloud modeling from the segmented bone surface image, and (3) point cloud registration using Q-learning of US-CT. Local phase image features are used as an input to an L1-norm-based regularization framework for enhancement of bone shadow regions. Simple bottom up ray casting method is used to segment the bone surfaces from the enhanced bone shadow images. In addition, CT data was segmented using intensity-based thresholding method. In other words, the complicated cross-modality US-CT registration was transformed into point cloud registration. In the next step, we proposed a hierarchical registration method using supervised Q-learning that learns the optimal sequence of motion action to achieve the optimal alignment. Within this approach, the agent is modeled using PointNet++ framework, with point cloud data obtained by segmenting the US and CT data as the input, and the next optimal action as the output. The quantitative and qualitative evaluations are performed on over 100 test cases and have shown the potential in making ultrasound as an alternative intra-operative image modality in image-guidance. The target registration error (TRE) and fiducial registration error (FRE) range have average value of 4.32 mm and 3.82 mm respectively. And the success rate, which defines as the TRE and FRE are both less than 10mm, is 92.6% with an average time of 0.31 second for each step.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 <u>Thesis Motivation</u>

Improper treatment of fixation or alignment may cause soft tissue injuries, bone loss and significant high risk on pulmonary compliance [1, 2]. The incidence of heterotopic ossification and rotational malalignment after intramedullary nailing (IM) have been reported to range from 10% to 53% and 10% to 37%, respectively [76, 77]. Thus, the successful treatments for femoral fracture is of importance in reducing the complication and malalignment rate significantly. For that, Image-guided CAOS system has been explored in providing improved operative outcomes with computer technology in pre-, intra-, and post-operative procedures. In the image-guided CASO system, the domain intra-operative imaging modality in femoral shaft fracture treatment is 2D fluoroscopy as a wide application in accurate targeting and tracking for image-guidance [3]. However, the cumulative radiation exposure to both surgeons and patients may cause unforeseen harm to them [4, 5]. Furthermore, the pool capabilities in 3-D of the intra-operative fluoroscopy imaging can result in screw mal-placement and injuries in muscles and nerves. And the high expense in 3D units and experience requirement for surgeons have prohibited the wide use of fluoroscopy in intra-operative image-guidance. For solving these issues, US has been investigated as an alternative intra-operative imaging modality in image guidance which is a radiation exposure free and real-time technology.

As the poor tissue/bone boundaries in conventional US images, directly usage of US images in orthopedic surgical guidance cannot deliver reliable results. For that, previous works have focused on bone surface segmentation [6, 7, 8] and the integration of US/CT registration in the intra-operative procedure of intramedullary nailing [9, 10, 11]. Those registration methods used the consecutive sequence of 2-D US images as the input that considered as 3-D volumes. However, the consecutive sequence of 2-D US image was still lack of spatial information that is necessary for surgical navigation.

This thesis aims to provide an accurate 3-D bone surface information intra-operatively from US images in real-time for intramedullary nailing treatment. Also, the reinforcement learning (RL) method is applied in this project to provide the robust and the accuracy of the current registration methods. We believe the 3-D registration using RL for US/CT can provide more accurate and safe treatment for femoral shaft fracture treatment compared to the traditional treatments.

1.2 Treatments of Femoral Shaft Fracture

The surgical treatment of femoral shaft fracture has three different types of surgery, plates and screw, external fixation and intramedullary nailing. External fixation is usually used as an impermanent treatment for femur fractures that uses metal pins and screws to stabilize the bones in the proper position [12, 13]. Intramedullary nailing (IM) is considered as a standard treatment for the femur fractures due to its stable and full-length fixation [14, 15]. Besides, when IM is not possible, the plates and screws technique is used where the fractures extend into either the hip or knee joints [16, 17]. And the example of external fixation and intramedullary nailing were shown in Fig 1.1.

<p style="text-align:center">(a)   (b)</p>
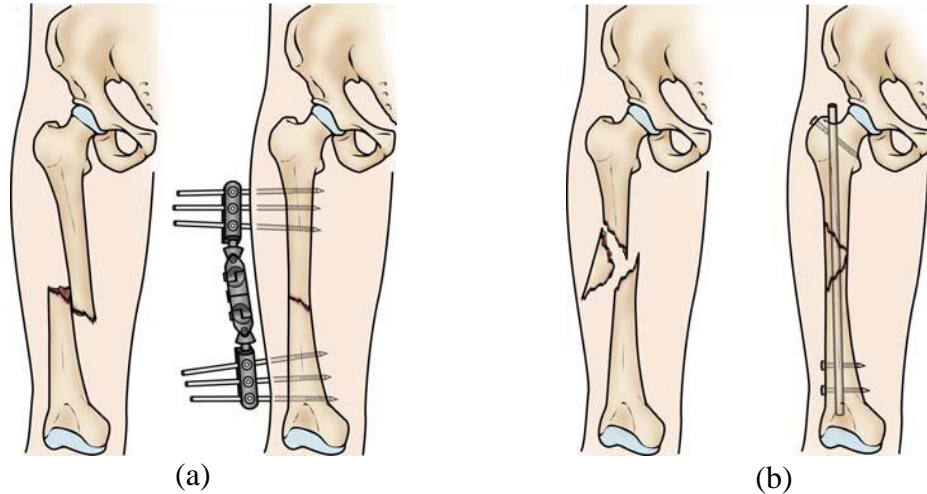
Figure 1.1: (a): External fixation treatment of femur fracture from [71]; (b) Intramedullary nailing treatment of femur fracture from [71]

During the IM procedure, the special intramedullary rod was inserted across the fracture site and the locking screws were placed perpendicular to the rod to secure the intramedullary nailing [18, 72]. From the procedure, accurate tracking and targeting with minimal invasion are secure from the intra-operative fluoroscopy imaging guidance using C-arm [19]. Intra-operative fluoroscopy images enable the surgeons to get access to the good contrast images for both bone and metal implants in real-time. Using these images, the surgeon can fixate the fracture using the intramedullary rod and reduce the fractures with screws under guidance that shown in Fig 1.2.

However, low reproducibility from 2-D to 3-D and the cumulative radiation exposure are still significant challenges for intra-operative fluoroscopy imaging during surgery guidance. The non-uniform exposure variation across the FOV of fluoroscopy, as well as the varying contrast from shot-to-shot, impeded the reproducibility from 2-D to 3-D [20]. Only partial data is suitable for the re-compute the spatial relationship between the pre-operative and intra-operative medical data during the surgery guidance process. The failure
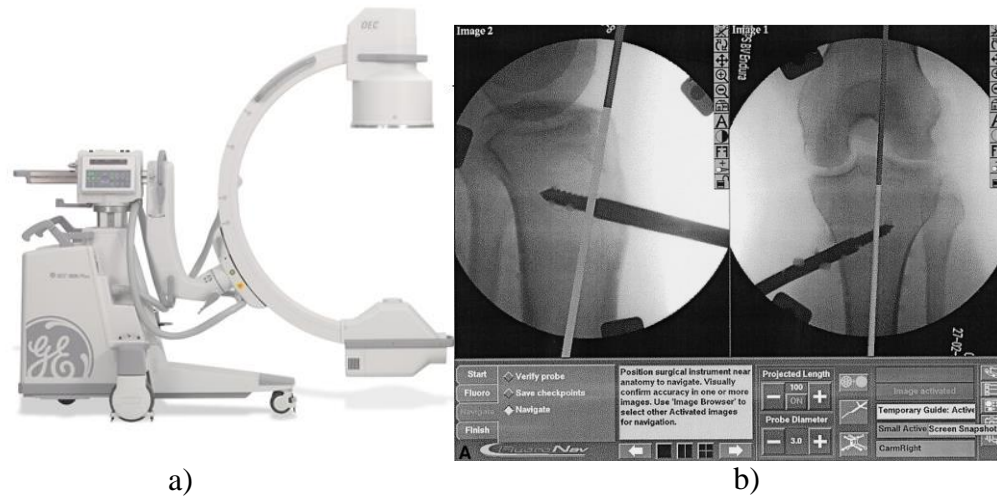
Figure 1.2 (a): A typical C-arm that's used for acquiring intra-operative fluoroscopy images from GE; (b): The acquired fluoroscopy image showing the inserted intramedullary rod and screws through the distal site from Hazan et [19].

in the alignment of spatial relationships can significantly affect the surgery outcomes. Currently, the reported rate of infection for femoral fracture with IM ranges from 0.7% to 2.1% [21]. In addition, the non-uniform rate for femoral fracture with IM ranges from 0.3% to 7.6% [21, 22, 23].

As for the cumulative radiation exposure issue, both surgeons and patients may suffer potentially harm during surgery. In the previous report, the average operation time for IM is 30 minutes and up to 180 minutes [24]. The repeated and long-term radiation exposure over several years has been linked to an increased incidence of cancer [25].

Computer-Assisted Orthopedic Surgery (CAOS) and intra-operative US images with non-radiation exposure have been explored as potential alternative methods to alleviate the issues we discussed above, which are described in detail in the following sections.

1.3 <u>Computer-Assisted Orthopedic Surgery (CAOS)</u>

1.3.1 <u>Bone Segmentation in US images</u>

The bone segmentation in the US can be traced at least as far back as the 1990s. Previous researches on bone segmentation in US images used manually segmentation of bone surface in each 2-D US image for different bone fracture surgical procedures [34-36]. The manual bone segmentation methods typically require an additional 20 minutes of operation time on average for surgical procedures, which could not be applied in surgery procedures [34].

Then, an automatic bone segmentation method was developed for computer-assisted orthopedic surgery using watershed segmentation for pelvic ring fracture [37]. It helped the surgeon to locate the tissue/bone boundaries in US images. Moreover, the other intensity-based segmentation methods were developed to extract the bone surface information from the US image using the intensity feature of bone afterward [38-39]. The intensity-based segmentation algorithms could be accomplished with the required operation time, which is less than 2 minutes, but the accuracy and robustness still need to be improved due to the intensity-variant feature of US images.

The gradient-invariant based bone surface segmentation method was previously demonstrated [40], which has been successfully applied in US-CT registration for computer-assisted orthopedic surgery system [41]. This method extracted phase symmetry (PS) regions that indicate the acoustic interfaces with high intensity using Log-Gabor filters. However, this method cannot be implemented automatically as it has some hand-tuned features. Therefore, the improved automatically segmentations based on PS

optimization has been demonstrated [42, 43] and applied in clinical data with the mean surface error of 0.62 mm on the pelvic surface [44].

Dynamic programming has also been applied in improving the robustness of current intensity- or phase-based segmentation methods by eliminating the effects of disconnected surfaces [45, 46]. Recently, the convolution neural network (CNN) has also been used in bone segmentation on different types of bone [47, 48]. However, the reliable and accurate bone segmentation still remains to be explored and improved for data acquired with different US imaging parameters across different kinds and locations of clinical data set.

1.3.2 <u>Overview of the CAOS system</u>

CAOS is defined as a system integrated with computer-assisted techniques that aim to improve orthopedic surgery outcomes. The frameworks for CAOS vary widely depending on the utilization and availability of technologies. Advances in medical imaging and spatial tracking sensors have made surgical navigation systems possible [26].

The CAOs system consists of three fundamental components, pre-operative modeling, surgical planning, and intra-operative imaging that shown in Fig 1.3. For the pre-operative modeling procedure, the golden standard method is CT scan imaging, which has the highest quality digital representation of the disease or injuries [26, 27]. Due to its high quality, the pre-operative CT can be used as a valuable assessment of the injuries as well as the helpful reference model for intra-operative surgical procedures. In the following surgical planning procedure, the necessarily quantitative analysis, path planning, and target positions can be identified and analyzed from the pre-operative CT model. For example, the desired

reduction and desired rod and screws locations are determined by the surgical expertise for IM.

The intra-operative imaging procedure is the most challenging task in CASO system as it requires to get the spatial relation between the anatomy and the instrument from the referenced preoperative model [27]. Tracking sensors, such as infrared cameras and electromagnetic sensors, are widely used to obtaining the reference position between the trackers and the patients in real-time [28, 29]. And the real-time medical data is obtained from tracked intra-operative imaging modality, such as X-ray and US [74]. After segmenting the anatomy information from the tracked images, the spatial alignment

between the preoperative CT model and the intra-operative images should be made to provide the necessary information for surgery guidance. Also, this registration process is an essential part of the surgery guidance, and we'll discuss the US-based CAOS system in the following sections.
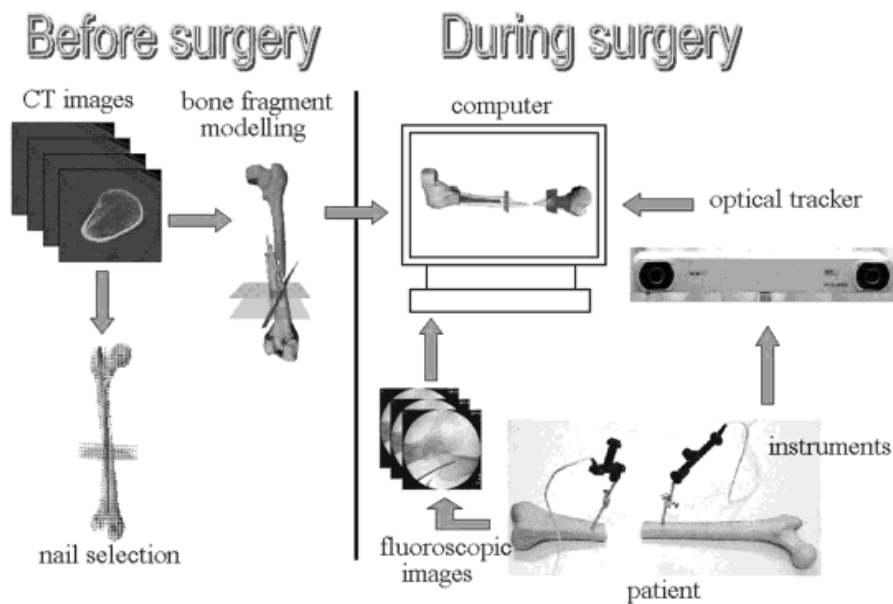
Figure 1.3 Framework of a typical X-ray-Based CAOS system that used CT images as the pre-operative medical data and fluoroscopy as the intra-operative medical data. And the optical tracker provided additional information in intra-operative image guidance from Joskowicz et [73]

1.3.3 <u>US-Based CAOS system</u>

US-based CAOS refers to a CAOS system using the US as an intra-operative image modality to provide spatial anatomy structure information in real-time that shown in Fig 1.4. In the typical US-based CAOS system, it mainly consists of navigator, impactor, virtual visualization monitors, ultrasound machine, and therapeutically subject [59]. The navigator and impactors are used for providing additional spatial information in vitro that improves the image guidance procedure. And the ultrasound machines collect the real-time US images and then sent them to the computer for process and visualization in the monitor [59, 74].

   Segmenting the pixels that contain anatomy structure intra-operatively from US images allows the physical location of the target structure to be known. In other words, for the femoral shaft fracture treatments, the segmentation of bone surface from US images is especially vital but challenging due to the low quality of the US. Moreover, the reliable and robust outcome of the US-CT registration process secures the accuracy of anatomy location targeting and the success rate of the surgery, which refers to compute the optimal alignment between the pre-operative CT data and intra-operative US data.
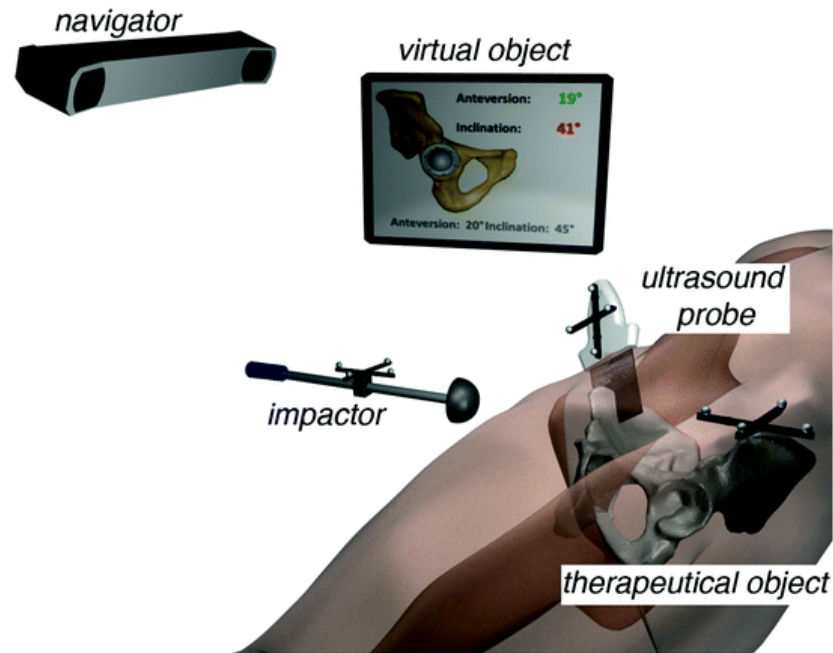
Figure 1.4 Example of a typical US-Based CAOS system that composes with the infrared camera, medical instrument, ultrasound probe, monitor, and the subject from [59].

1.4 Literature Review in Previous work of US-Based CAOS

Several advanced studies have been explored towards achieving US-Based surgical interventions intra-operatively. These researches mainly focus on two different aspects, bone segmentation in US images and multi-modal registration algorithms, which committed to overcoming the challenges we discussed in Section 3.1.

1.4.1 US-CT registration

Previous studies on the US-CT registration have shown that the intra-operative US image is feasible to apply in surgical operations. According to the type of data, the existing registration of the fusion of CT and US can be divided into three catalogs, landmark-based digitization, surface-based registration, and volume-based registration [75].

For the landmark-based digitization, the digitized landmarks were generally selected to set up a reference plane that the intra-operative reference spatial transformation with preoperative CT data could be computed [60, 61, 62]. Jaramaz et al. [63] designed and employed the reference plane with three landmarks in hip replacement operation, which referred as anterior pelvic plane (APP) [64]. The orientation of the implants with respect to the APP can be measured during the surgical procedure. In the further reported results, this method demonstrated the average error on inclination and anteversion are $-0.1 \pm 1.0°$ and $-0.4 \pm 2.7°$ comparing with the ground truth [60, 61, 62]. However, setting up the APP with three landmarks is not accurate for most of the surgical treatments. Accurate spatial information should be determined during surgical guidance.

One of the first papers in surface-based registration proposed a US-CT registration method used intensity feature for surgical guidance of a lumbar operation [49, 58]. Nevertheless, the bone surface information still required manual US segmentation for it. Other surface-based registration techniques have been explored to achieve automatic bone segmentation for US-CT registration [50-52]. Ionescu et al proposed one of the first surface-based registration methods in US-CT by extracting the cortical aspect of the bone in US images [50]. But most of these methods require alignment estimation or fiducial points identification before the real-time registration as the high computation expensive for the surface-based registration convergence. In order to improve the computation expensive of surface-based registration methods, point-based registration techniques have been developed to achieve time-efficiency [53-55]. Rasoulian et al. [65] employed the biomechanical spring model to simulate the point-based US spine and femur models and then aligned the point-based US models with the CT models using iterative closet point

(ICP) registration. The average target registration error of this proposed method was 2.2 mm in vivo and 1.99 mm in vitro.

Sometimes, the bone segmentation is not necessary for surgical guidance in US-Based CAOS as the volume-based registration can be employed to get the spatial information [66, 67, 70]. Brendel et al. [67] proposed a volume-based registration method for US-CT of the lumbar spine. Before surgery, they segmented the CT surface model and designed the scanning path for the orientation of position of US probe with respect to the CT. Then the intra-operative registration can be done with the preoperative scanning path data using intensity-based registration. Besides, another application has been developed based on this idea [68]. They simulated the US images from the abdominal CT model as the preoperative medical data and then used intensity-based registration to achieve the alignment with the intra-operative US data. The reported RMS mean target registration error and success rate for the relative researches was 8.1 mm in the liver and kidney of 25 patients [69].

1.5 <u>Thesis Objective</u>

Our main objective is the development of a robust, accurate and real-time intra-operative registration method using RL for US-based CASO with specific focus on femoral shaft fracture. Our specific aims, listed below, once achieved will improve our understanding of the challenges faces in US-based guidance and provide a new solution which has not been investigated previously. Finally, the outcomes achieved in this work will help us in the future development of an US-based surgery guidance system which could provide a potential alternative to X-ray as the intra-operative image modality during surgery.

The specific aims of this thesis are listed below:

1. Develop and evaluate a robust bone surface segmentation method, specific to femur bones, to overcome the cross-modality problem encountered in US-CT based CAOS system.

2. Develop and evaluate a point cloud-based registration method for intraoperative fusion of US and CT using reinforcement learning for femoral shaft fracture surgery.

Reference

[1] Harwood, Paul John, Peter V. Giannoudis, Martijn Van Griensven, Christian Krettek, and Hans-Christoph Pape. "Alterations in the systemic inflammatory response after early total care and damage control procedures for femoral shaft fracture in severely injured patients." Journal of Trauma and Acute Care Surgery 58, no. 3 (2005): 446-454.

[2] Schemitsch, Emil H., Rina Jain, Diana C. Turchin, J. Brendan Mullen, Robert J. Byrick, Gail I. Anderson, and Robin R. Richards. "Pulmonary effects of fixation of a fracture with a plate compared with intramedullary nailing. A canine model of fat embolism and fracture fixation." JBJS 79, no. 7 (1997): 984-96.

[3] Whatling, Gemma Marie, and Leonard Derek Martin Nokes. "Literature review of current techniques for the insertion of distal screws into intramedullary locking nails." Injury 37, no. 2 (2006): 109-119.

[4] Nelson, Elisha M., Shafagh M. Monazzam, Kee D. Kim, J. Anthony Seibert, and Eric O. Klineberg. "Intra-operative fluoroscopy, portable X-ray, and CT: patient and operating room personnel radiation exposure in spinal surgery." The Spine Journal 14, no. 12 (2014): 2985-2991.

[5] Hall, E. J., and D. J. Brenner. "Cancer risks from diagnostic radiology." The British journal of radiology 81, no. 965 (2008): 362-378.

[6] Hacihaliloglu, Ilker, Rafeef Abugharbieh, Antony Hodgson, and Robert Rohling. "Bone segmentation and fracture detection in ultrasound using 3D local phase features." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 287-295. Springer, Berlin, Heidelberg, 2008.

[7] Hacihaliloglu, Ilker, Rafeef Abugharbieh, Antony J. Hodgson, and Robert N. Rohling. "2A-4 Enhancement of bone surface visualization from 3D ultrasound based on local phase information." In 2006 IEEE Ultrasonics Symposium, pp. 21-24. IEEE, 2006.

[8] Hacihaliloglu, Ilker, Rafeef Abugharbieh, Antony J. Hodgson, and Robert N. Rohling. "Bone surface localization in ultrasound using image phase-based features." Ultrasound in medicine & biology 35, no. 9 (2009): 1475-1487.

[9] Amin, Devin V., Takeo Kanade, Anthony M. Digioia, and Branislav Jaramaz. "Ultrasound registration of the bone surface for surgical navigation." Computer Aided Surgery 8, no. 1 (2003): 1-16.

[10] Mahaisavariya, Banchong. "Ultrasound-guided intramedullary nailing." Techniques in Orthopaedics 13, no. 1 (1998): 61-70.

[11] Masson-Sibut, Agnès, Amir Nakib, Eric Petit, and François Leitner. "Computer-assisted intramedullary nailing using real-time bone detection in 2D ultrasound images." In International Workshop on Machine Learning in Medical Imaging, pp. 18-25. Springer, Berlin, Heidelberg, 2011.

[12] Haidukewych, George J. "Temporary external fixation for the management of complex intra-and periarticular fractures of the lower extremity." Journal of orthopaedic trauma 16, no. 9 (2002): 678-685.

[13] Scalea, Thomas M., Sharon A. Boswell, Jane D. Scott, Kimberly A. Mitchell, Mary E. Kramer, and Andrew N. Pollak. "External fixation as a bridge to intramedullary nailing for patients with multiple injuries and with femur fractures: damage control orthopedics." Journal of Trauma and Acute Care Surgery 48, no. 4 (2000): 613-623.

[14] Winquist, R. A., Sigvard Jr, H. T., & Clawson, D. K. (2005). Closed intramedullary nailing of femoral fractures. A report of five hundred and twenty cases. Orthopedic Trauma Directions, 3(04), 29-31.

[15] Hansen, Sigvard T., and Robert A. Winquist. "Closed intramedullary nailing of the femur. Küntscher technique with reaming." Clinical orthopaedics and related research 138 (1979): 56-61.

[16] Wolfgang, Gary L., Michael H. Bryant, and James P. O'Neill. "Treatment of intertrochanteric fracture of the femur using sliding screw plate fixation." Clinical orthopaedics and related research 163 (1982): 148-158.

[17] Egol, Kenneth A., Erik N. Kubiak, Eric Fulkerson, Frederick J. Kummer, and Kenneth J. Koval. "Biomechanics of locked plates and screws." Journal of orthopaedic trauma 18, no. 8 (2004): 488-493.

[18] Slone, R. M., M. M. Heare, R. A. Vander Griend, and W. J. Montgomery. "Orthopedic fixation devices." Radiographics 11, no. 5 (1991): 823-847.

[19] Hazan, Eric J., and Leo Joskowicz. "Computer-assisted image-guided intramedullary nailing of femoral shaft fractures." Techniques in Orthopaedics 18, no. 2 (2003): 191-200.

[20] Yaniv, Ziv, Leo Joskowicz, Ariel Simkin, Maria Garza-Jinich, and Charles Milgrom. "Fluoroscopic image processing for computer-aided orthopaedic surgery." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 325-334. Springer, Berlin, Heidelberg, 1998.

[21] Karadimas, Efthimios J., George Papadimitriou, Gerasimos Theodoratos, Anastasios Papanikolaou, and John Maris. "The effectiveness of the antegrade reamed technique: the experience and complications from 415 traumatic femoral shaft fractures." Strategies in Trauma and Limb Reconstruction 4, no. 3 (2009): 113-121.

[22] Malik, M. H. A., P. Harwood, P. Diggle, and S. A. Khan. "Factors affecting rates of infection and nonunion in intramedullary nailing." The Journal of bone and joint surgery. British volume 86, no. 4 (2004): 556-560.

[23] El Moumni, M., P. A. Leenhouts, H. J. Ten Duis, and K. W. Wendt. "The incidence of non-union following unreamed intramedullary nailing of femoral shaft fractures." Injury 40, no. 2 (2009): 205-208.

[24] Kirousis, George, Harry Delis, Panagiotis Megas, Elias Lambiris, and George Panayiotakis. "Dosimetry during intramedullary nailing of the tibia: patient and occupational exposure." Acta orthopaedica 80, no. 5 (2009): 568-572.

[25] Brenner, David J., Richard Doll, Dudley T. Goodhead, Eric J. Hall, Charles E. Land, John B. Little, Jay H. Lubin et al. "Cancer risks attributable to low doses of ionizing radiation: assessing what we really know." Proceedings of the National Academy of Sciences 100, no. 24 (2003): 13761-13766.

[26] Ecker, Timo Michael, J. Jost, J. L. Cullmann, Wolf-Dieter Zech, Valentin Djonov, M. J. B. Keel, Lorin Michael Benneker, and Johannes Dominik Bastian. "Percutaneous screw fixation of the iliosacral joint: a case-based preoperative planning approach reduces operating time and radiation exposure." Injury 48, no. 8 (2017): 1825-1830.

[27] Cleary, Kevin, and Terry M. Peters. "Image-guided interventions: technology review and clinical applications." Annual review of biomedical engineering 12 (2010): 119-142.

[28] Schlenzka, Dietrich, Timo Laine, and Teija Lund. "Computer-assisted spine surgery." European Spine Journal 9, no. 1 (2000): S057-S064.

[29] Roessler, K., K. Ungersboeck, W. Dietrich, M. Aichholzer, K. Hittmeir, C. H. Matula, T. H. Czech, and W. Th Koos. "Frameless stereotactic guided neurosurgery: clinical experience with an infrared based pointer device navigation system." Acta neurochirurgica 139, no. 6 (1997): 551-559.

[30] Szabo, Thomas L. Diagnostic ultrasound imaging: inside out. Academic Press, 2004. Page 7.

[31] Azhari, Haim. "Appendix A: typical acoustic properties of tissues." (2010).

[32] Noble, Alison, and Djamal Boukerroui. "Ultrasound image segmentation: a survey." IEEE Transactions on medical imaging 25, no. 8 (2006): 987-1010.

[33] Quader, Niamul, Antony Hodgson, Kishore Mulpuri, Anthony Cooper, and Rafeef Abugharbieh. "Towards reliable automatic characterization of neonatal hip dysplasia from 3D ultrasound images." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 602-609. Springer, Cham, 2016.

[34] Carrat, Lionel, Jerome Tonetti, Stephane Lavallee, Philippe Merloz, Laurence Pittet, and Jean-Paul Chirossel. "Treatment of pelvic ring fractures: percutaneous computer assisted iliosacral screwing." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 84-91. Springer, Berlin, Heidelberg, 1998.

[35] Carrat, Lionel, Jérome Tonetti, Philippe Merloz, and Jocelyne Troccaza. "Percutaneous computer assisted iliosacral screwing: clinical validation." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 1229-1237. Springer, Berlin, Heidelberg, 2000.

[36] Hata, Nobuhiko, Makoto Suzuki, Takeyoshi Dohi, Hiroshi Iseki, Kintomo Takakura, and Daijo Hashimoto. "Registration of ultrasound echography for intra-operative use: a newly developed multiproperty method." In Visualization in Biomedical Computing 1994, vol. 2359, pp. 251-259. International Society for Optics and Photonics, 1994.

[37] Thomas, Judith G., Richard Alan Peters, and Philippe Jeanty. "Automatic segmentation of ultrasound images using morphological operators." IEEE Transactions on Medical Imaging 10, no. 2 (1991): 180-186.

[38] Ionescu, Gelu, Stéphane Lavallée, and Jacques Demongeot. "Automated registration of ultrasound with CT images: Application to computer assisted prostate radiotherapy and orthopedics." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 768-777. Springer, Berlin, Heidelberg, 1999.

[39] Amin, Devin V., Takeo Kanade, Anthony M. Digioia, and Branislav Jaramaz. "Ultrasound registration of the bone surface for surgical navigation." Computer Aided Surgery 8, no. 1 (2003): 1-16.

[40] Hacihaliloglu, Ilker, Rafeef Abugharbieh, Antony Hodgson, and Robert Rohling. "Bone segmentation and fracture detection in ultrasound using 3D local phase features." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 287-295. Springer, Berlin, Heidelberg, 2008.

[41] Anas, Emran Mohammad Abu, Alexander Seitel, Abtin Rasoulian, Paul St John, Tamas Ungi, Andras Lasso, Kathryn Darras et al. "Bone enhancement in ultrasound based on 3D local spectrum variation for percutaneous scaphoid fracture fixation." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 465-473. Springer, Cham, 2016.

[42] Anas, Emran Mohammad Abu, Alexander Seitel, Abtin Rasoulian, Paul St John, Tamas Ungi, Andras Lasso, Kathryn Darras et al. "Bone enhancement in ultrasound based on 3D local spectrum variation for percutaneous scaphoid fracture fixation." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 465-473. Springer, Cham, 2016.

[43] Hacihaliloglu, Ilker. "Enhancement of bone shadow region using local phase-based ultrasound transmission maps." International journal of computer assisted radiology and surgery 12, no. 6 (2017): 951-960.

[44] Hacihaliloglu, Ilker. "Enhancement of bone shadow region using local phase-based ultrasound transmission maps." International journal of computer assisted radiology and surgery 12, no. 6 (2017): 951-960.

[45] Foroughi, Pezhman, Emad Boctor, Michael J. Swartz, Russell H. Taylor, and Gabor Fichtinger. "P6d-2 ultrasound bone segmentation using dynamic programming." In 2007 IEEE Ultrasonics Symposium Proceedings, pp. 2523-2526. IEEE, 2007.

[46] Hacihaliloglu, Ilker. "Localization of bone surfaces from ultrasound data using local phase information and signal transmission maps." In International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging, pp. 1-11. Springer, Cham, 2017.

[47] Salehi, Mehrdad, Raphael Prevost, José-Luis Moctezuma, Nassir Navab, and Wolfgang Wein. "Precise ultrasound bone registration with learning-based segmentation and speed of sound calibration." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 682-690. Springer, Cham, 2017.

[48] Wang, Puyang, Vishal M. Patel, and Ilker Hacihaliloglu. "Simultaneous segmentation and classification of bone surfaces from ultrasound using a multi-feature guided CNN." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 134-142. Springer, Cham, 2018.

[49] Brendel, Bernhard, S. Winter A. Rick, M. Stockheim, and H. Ermert. "Registration of 3D CT and ultrasound datasets of the spine using bone structures." Computer Aided Surgery 7, no. 3 (2002): 146-155.

[50] Ionescu, Gelu, Stéphane Lavallée, and Jacques Demongeot. "Automated registration of ultrasound with CT images: Application to computer assisted prostate radiotherapy and orthopedics." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 768-777. Springer, Berlin, Heidelberg, 1999.

[51] Penney, Graeme P., Dean C. Barratt, Carolyn SK Chan, Mike Slomczykowski, Timothy J. Carter, Philip J. Edwards, and David J. Hawkes. "Cadaver validation of intensity-based ultrasound to CT registration." Medical image analysis 10, no. 3 (2006): 385-395.

[52] Barratt, Dean C., Graeme P. Penney, Carolyn SK Chan, Mike Slomczykowski, Timothy J. Carter, Philip J. Edwards, and David J. Hawkes. "Self-calibrating 3D-ultrasound-based bone registration for minimally invasive orthopedic surgery." IEEE transactions on medical imaging 25, no. 3 (2006): 312-323.

[53] Brounstein, Anna, Ilker Hacihaliloglu, Pierre Guy, Antony Hodgson, and Rafeef Abugharbieh. "Towards real-time 3D US to CT bone image registration using phase and curvature feature based GMM matching." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 235-242. Springer, Berlin, Heidelberg, 2011.

[54] Hacihaliloglu, Ilker, Anna Brounstein, Pierre Guy, Antony Hodgson, and Rafeef Abugharbieh. "3D ultrasound-CT registration in orthopaedic trauma using GMM registration with optimized particle simulation-based data reduction." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 82-89. Springer, Berlin, Heidelberg, 2012.

[55] Brounstein, Anna, Ilker Hacihaliloglu, Pierre Guy, Antony Hodgson, and Rafeef Abugharbieh. "Fast and accurate data extraction for near real-time registration of 3-D ultrasound and computed tomography in orthopedic surgery." Ultrasound in medicine & biology 41, no. 12 (2015): 3194-3204.

[56] Hacihaliloglu, I. Enhancement of bone shadow region using local phase-based ultrasound transmission maps Int J CARS (2017) 12: 951.

[57] Wein W, Brunke S, Khamene A, et al. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention.[J]. Medical Image Analysis, 2008, 12(5):577-585.

[58] Fanti, Zian, Fabian Torres, Eric Hazan-Lasri, Alfonso Gastelum-Strozzi, Leopoldo Ruiz-Huerta, Alberto Caballero-Ruiz, and F. Arámbula Cosío. "Improved Surface-Based

Registration of CT and Intra-operative 3D Ultrasound of Bones." Journal of healthcare engineering 2018 (2018).

[59] Schumann, Steffen. "State of the art of ultrasound-based registration in computer assisted orthopedic interventions." In Computational Radiology for Orthopaedic Interventions, pp. 271-297. Springer, Cham, 2016.

[60] Kiefer, Hartmuth, and Aneis Othman. "Ultrasound vs pointer palpation based method in THA navigation: a comparative study." Orthopedics 30, no. 10 (2007): S153.

[61] Parratte, S., P. Kilian, V. Pauly, P. Champsaur, and J-NA Argenson. "The use of ultrasound in acquisition of the anterior pelvic plane in computer-assisted total hip replacement: a cadaver study." The Journal of bone and joint surgery. British volume 90, no. 2 (2008): 258-263.

[62] Wassilew, Georgi I., Markus O. Heller, Olaf Hasart, Carsten Perka, Ingrid Südhoff, Viktor Janz, Georg N. Duda, and Christian König. "Ultrasound-based computer navigation of the acetabular component: a feasibility study." Archives of orthopaedic and trauma surgery 132, no. 4 (2012): 517-525.

[63] Jaramaz, Branislav, Anthony M. DiGioia III, Mike Blackwell, and Constantinos Nikou. "Computer assisted measurement of cup placement in total hip replacement." Clinical Orthopaedics and Related Research® 354 (1998): 70-81.

[64] Lewinnek, George E., J. L. Lewis, R. I. C. H. A. R. D. Tarr, C. L. Compere, and J. R. Zimmerman. "Dislocations after total hip-replacement arthroplasties." The Journal of bone and joint surgery. American volume 60, no. 2 (1978): 217-220.

[65] Rasoulian, Abtin, Purang Abolmaesumi, and Parvin Mousavi. "Feature‐based multibody rigid registration of CT and ultrasound images of lumbar spine." Medical physics 39, no. 6Part1 (2012): 3154-3166.

[66] Brendel, Bernhard, S. Winter A. Rick, M. Stockheim, and H. Ermert. "Registration of 3D CT and ultrasound datasets of the spine using bone structures." Computer Aided Surgery 7, no. 3 (2002): 146-155.

[67] Winter, Susanne, Bernhard Brendel, Ioannis Pechlivanis, Kirsten Schmieder, and Christian Igel. "Registration of CT and intra-operative 3-D ultrasound images of the spine using evolutionary and gradient-based methods." IEEE Transactions on Evolutionary Computation 12, no. 3 (2008): 284-296.

[68] Wein, Wolfgang, Ali Khamene, Dirk-André Clevert, Oliver Kutter, and Nassir Navab. "Simulation and fully automatic multimodal registration of medical ultrasound." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 136-143. Springer, Berlin, Heidelberg, 2007.

[69] Wein, Wolfgang, Shelby Brunke, Ali Khamene, Matthew R. Callstrom, and Nassir Navab. "Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention." Medical image analysis 12, no. 5 (2008): 577-585.

[70] Penney, Graeme P., Dean C. Barratt, Carolyn SK Chan, Mike Slomczykowski, Timothy J. Carter, Philip J. Edwards, and David J. Hawkes. "Cadaver validation of intensity-based ultrasound to CT registration." Medical image analysis 10, no. 3 (2006): 385-395.

[71] https://orthoinfo.aaos.org/en/diseases--conditions/femur-shaft-fractures-broken-thighbone/

[72] Huang, Kui-Chou, Kwok-Man Tong, Yu-Min Lin, El-Wui Loh, and Cheng-En Hsu. "Evaluation of methods and timing in nail dynamisation for treating delayed healing femoral shaft fractures." Injury 43, no. 10 (2012): 1747-1752.

[73] Joskowicz, Leo, Charles Milgrom, Ariel Simkin, Lana Tockus, and Ziv Yaniv. "FRACAS: A system for computer‑aided image‑guided long bone fracture surgery." Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery (ISCAS) 3, no. 6 (1998): 271-288.

[74] Amin, Devin V., Takeo Kanade, Anthony M. Digioia, and Branislav Jaramaz. "Ultrasound registration of the bone surface for surgical navigation." Computer Aided Surgery 8, no. 1 (2003): 1-16.

[75] Zheng, Guoyan, and Shuo Li, eds. Computational radiology for orthopaedic interventions. New York: Springer, 2016.

[76] Meyers, Carolyn, Jeffrey Lisiecki, Sarah Miller, Adam Levin, Laura Fayad, Catherine Ding, Takashi Sono, Edward McCarthy, Benjamin Levi, and Aaron W. James. "Heterotopic Ossification: A Comprehensive Review." JBMR plus 3, no. 4 (2019): e10172.

[77] Ricci, William M., Carlo Bellabarba, Robert Lewis, Bradley Evanoff, Dolfi Herscovici, Thomas DiPasquale, and Roy Sanders. "Angular malalignment after intramedullary nailing of femoral shaft fractures." Journal of orthopaedic trauma 15, no. 2 (2001): 90-95.

CHAPTER 2

US-BASED CAOS CHALLENGES

This section represents the overview of current challenges for US-based CAOS, which includes the US imaging artifacts, bone surface segmentation in US image, and the US-CT registration methods. Besides, the basic concepts of the critical techniques that we used in this work to overcome the limitation of the current challenges are described in the following section.

2.1 US imaging artifacts

The main physical principle of US images is the acoustic reflection. In an ultrasound system, a transducer placed on the skin superficial to the interest area and emitted acoustic energy throughout the body. Also, the transducer receives the reflected acoustic signal which defines the intensity of the images, and the time delay between sending and receiving determines the position in the images [1]. The reflected intensity directly depends on the changes in the acoustic impedance along with the acoustic wave as shown in Fig 2.1. As the bone has the highest acoustic impedance, the reflected signal from bone typically has a high intensity as most of the incoming acoustic energy is reflected back to the transducer [2]. Although the US provides non-ionizing image modality in real-time, US images still have significant speckle noise and confounding soft tissue data compared to X-ray images. Thus, pre-processing is necessary before applying it in practice.

<center>(a)   (b)</center>
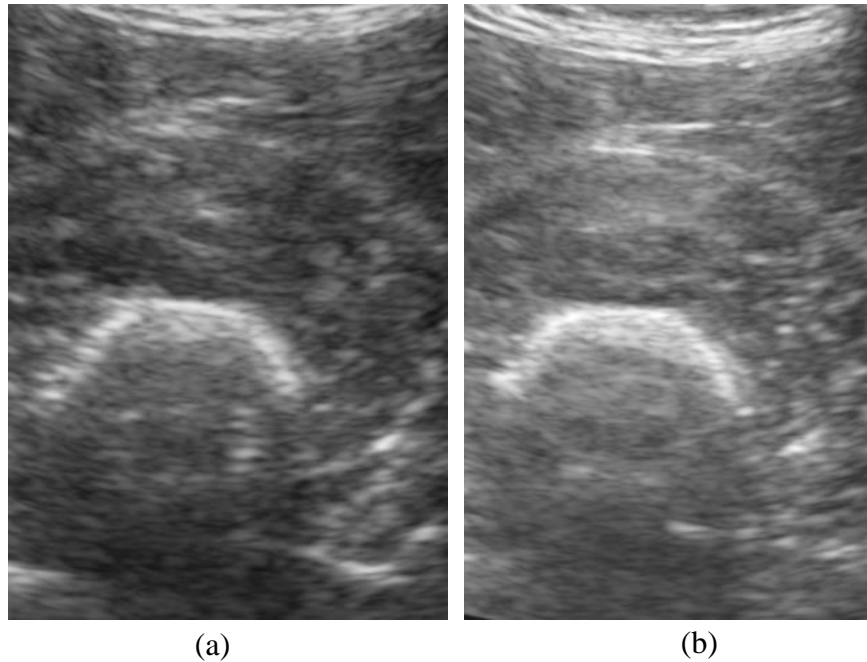
Figure 2.1 2D femur US image obtained with different settings in vivo has different speckle noises

2.2 Bone surface segmentation in US images

Secondly, the bone segmentation is the essential, but challenging pre-process procedure for CAOS system. As soft-tissue can also appear hyper-echogenic and image similarly to bone as high-intensity pixels, it is challenging to classify them correctly. Moreover, the noise and speckles in US images impede the differentiating process in the anatomical boundaries [3]. Furthermore, in the hand-held US system, the orientation of the US probe changes irregularly with respect to the bone surface can alter the appearance drastically [4]. Thus, the robust and accurate bone surface segmentation should be secured in a functional CAOS system for femoral shaft fracture.
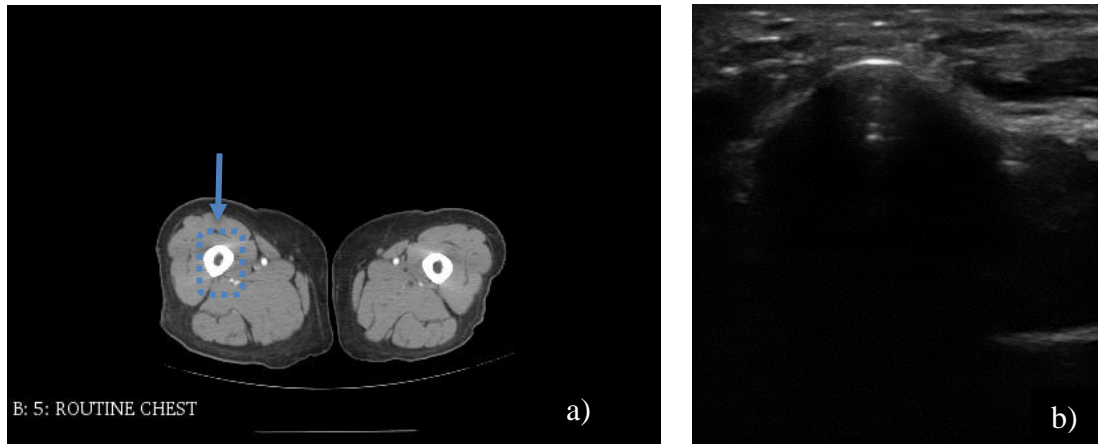
Figure 2.2 (a): example CT scan slice from a femur in cortical view. The blue dashed rectangle and the arrow highlight the approximate location and rotation of the US image, respectively. (b): US image of the bone from the same femur in the left.

2.3 US-CT registration

Last but not least, US-CT registration is another challenging task in the CAOS system for femoral fracture as it provides real-time spatial relationships by integrating different image modalities. As the US and CT have different imaging principles, the appearances of those images are significant different in the resolution and the FOV as shown in Fig 2.2. While the CT scan captures the entire cortical view of the body with high quality, the US can only get the narrow one scanned side with low quality. The cross-modality for the US-CT registration must be accurate enough to align them rigidly in 3-D space despite these challenges.

2.4 <u>Local phase image features</u>

Local phase features are pivotal in analyzing the structure of the image as they contribute to the visual appearance of the image [5, 34]. The phase image can be utilized in the feature extraction for an image as it contains the edges and detail of the image.

From the previous studies, local phase features have been applied in feature extraction successfully using the local energy model [6]. This proposed method extracted the phase features from the prominent features located in the maximal phase of the Fourier components. In addition, the local phase features could also be extracted from the US image in the frequency domain using band-pass quadrature filters [7]. Afterward, some studies in the local phase feature demonstrated that the potential of the local phase features in bone surface enhancement for detecting the fracture [8-15].
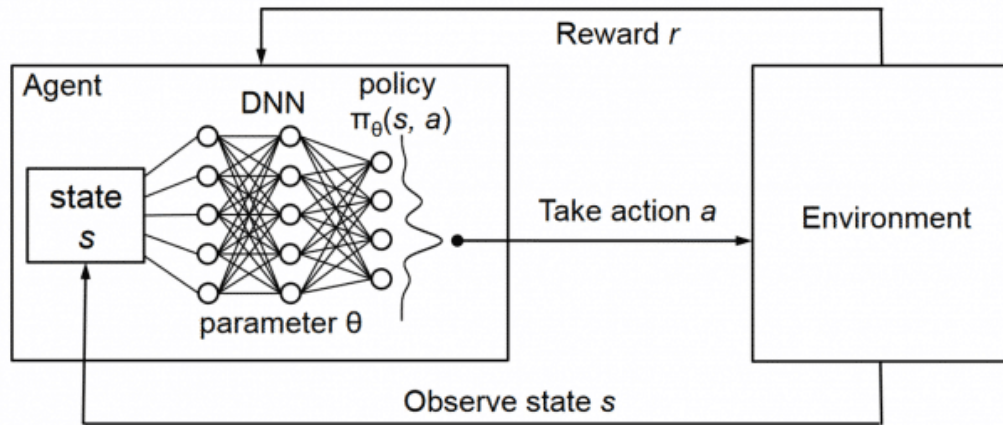
Figure 2.3 A typical workflow for the Q-learning method, which is a reinforcement learning method interacting with the environment from [35].

2.5 <u>Hierarchical Reinforcement Learning in registration</u>

Reinforcement learning (RL) has been widely explored in operation researches, such as gaming reinforcement and registration procedure, in the machine learning and AI field due to its self-adaptation and self-learning [16, 33]. The agent in RL learns and updates according to the interactions with its environment [17]. Hierarchical reinforcement learning (HRL) refers to an RL algorithm that can be decomposed into some sub-problems to solve that could be more powerful than solving the entire problem [18-23]. And Q-learning is one of the classic HRL methods that have been utilized in different areas: face recognition, web-based education, etc [24, 25].

Q-learning aims to find the optimal action for each state by estimate the action-value function (Q-value function) using iteration estimation. In Figure 2.3, we can know that the agent learns the action policy, which is modeled using the deep neural network (DNN), and then observes and learns the state and DNN from the interaction of the environment. To improve the efficiency of the accuracy in Q-value estimation, some studies have been explored in this field by improving the model of the Q-learning, such as doubly deep Q-learning, prioritized replay, and dueling network [26-29, 33]. And they also demonstrated the potential of Q-learning in registration [30-32], which can be utilized in our work potentially.

Reference

[1] Szabo, Thomas L. Diagnostic ultrasound imaging: inside out. Academic Press, 2004. Page 7.

[2] Azhari, Haim. "Appendix A: typical acoustic properties of tissues." (2010).

[3] Noble, Alison, and Djamal Boukerroui. "Ultrasound image segmentation: a survey." IEEE Transactions on medical imaging 25, no. 8 (2006): 987-1010.

[4] Quader, Niamul, Antony Hodgson, Kishore Mulpuri, Anthony Cooper, and Rafeef Abugharbieh. "Towards reliable automatic characterization of neonatal hip dysplasia from 3D ultrasound images." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 602-609. Springer, Cham, 2016.

[5] Hacihaliloglu, Ilker, Abtin Rasoulian, Robert N. Rohling, and Purang Abolmaesumi. "Local phase tensor features for 3-D ultrasound to statistical shape+ pose spine model registration." IEEE transactions on Medical Imaging 33, no. 11 (2014): 2167-2179.

[6] Felsberg, Michael, and Gerald Sommer. "The monogenic signal." IEEE transactions on signal processing 49, no. 12 (2001): 3136-3144.

[7] Roerdink, Jos BTM, and Arnold Meijster. "The watershed transform: Definitions, algorithms and parallelization strategies." Fundamenta informaticae 41, no. 1, 2 (2000): 187-228.

[8] Karamalis, Athanasios, Wolfgang Wein, Tassilo Klein, and Nassir Navab. "Ultrasound confidence maps using random walks." Medical image analysis 16, no. 6 (2012): 1101-1112.

[9] Sharma, Neeraj, and Lalit M. Aggarwal. "Automated medical image segmentation techniques." Journal of medical physics/Association of Medical Physicists of India 35, no. 1 (2010): 3.

[10] Belaid, Ahror, and Djamal Boukerroui. "α scale spaces filters for phase based edge detection in ultrasound images." In 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 1247-1250. IEEE, 2014.

[11] Kamdi, Shilpa, and R. K. Krishna. "Image segmentation and region growing algorithm." International Journal of Computer Technology and Electronics Engineering (IJCTEE) 2, no. 1 (2012).

[12] Mohammed M Abdelsamea, \An enhancement neighborhood connected segmentation for 2d-cellular image," arXiv preprint arXiv:1407.3664, 2014.

[13] Pohle, Regina, and Klaus D. Toennies. "Segmentation of medical images using adaptive region growing." In Medical Imaging 2001: Image Processing, vol. 4322, pp. 1337-1346. International Society for Optics and Photonics, 2001.

[14] Bozkurt, Ferhat, Cemal Köse, and Ahmet San. "Comparison of seeded region growing and random walk methods for vessel and bone segmentation in CTA images." In

2017 10th International Conference on Electrical and Electronics Engineering (ELECO), pp. 561-567. IEEE, 2017.

[15] Ng, H. P., S. Huang, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski. "Medical image segmentation using watershed segmentation with texture-based region merging." In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4039-4042. IEEE, 2008.

[16] Wang, Jian, Lei Zuo, Xin Xu, and Chun Li. "A hierarchical representation policy iteration algorithm for reinforcement learning." In International Conference on Intelligent Science and Intelligent Data Engineering, pp. 735-742. Springer, Berlin, Heidelberg, 2012.

[17] Barto, Andrew G., and Sridhar Mahadevan. "Recent advances in hierarchical reinforcement learning." Discrete event dynamic systems 13, no. 1-2 (2003): 41-77.

[18] Kawano, Hiroshi. "Hierarchical sub-task decomposition for reinforcement learning of multi-robot delivery mission." In 2013 IEEE International Conference on Robotics and Automation, pp. 828-835. IEEE, 2013.

[19] Xiaoqin, Du, Li Qinghua, and Han Jianjun. "Applying hierarchical reinforcement learning to computer games." In 2009 IEEE International Conference on Automation and Logistics, pp. 929-932. IEEE, 2009.

[20] Gil, Paulo, and Luís Nunes. "Hierarchical reinforcement learning using path clustering." In 2013 8th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-6. IEEE, 2013.

[21] Stulp, Freek, and Stefan Schaal. "Hierarchical reinforcement learning with movement primitives." In 2011 11th IEEE-RAS International Conference on Humanoid Robots, pp. 231-238. IEEE, 2011.

[22] Dietterich, Thomas G. "Hierarchical reinforcement learning with the MAXQ value function decomposition." Journal of artificial intelligence research 13 (2000): 227-303.

[23] Cuayáhuitl, Heriberto, and Nina Dethlefs. "Spatially-aware dialogue control using hierarchical reinforcement learning." ACM Transactions on Speech and Language Processing (TSLP) 7, no. 3 (2011): 5.

[24] Rodrigues Gomes, Eduardo, and Ryszard Kowalczyk. "Dynamic analysis of multiagent Q-learning with ε-greedy exploration." In Proceedings of the 26th annual international conference on machine learning, pp. 369-376. ACM, 2009.

[25] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." Machine learning 8, no. 3-4 (1992): 279-292.

[26] Wang, Ziyu, Tom Schaul, Matteo Hessel, Hado Van Hasselt, Marc Lanctot, and Nando De Freitas. "Dueling network architectures for deep reinforcement learning." arXiv preprint arXiv:1511.06581 (2015).

[27] Jaderberg, Max, Volodymyr Mnih, Wojciech Marian Czarnecki, Tom Schaul, Joel Z. Leibo, David Silver, and Koray Kavukcuoglu. "Reinforcement learning with unsupervised auxiliary tasks." arXiv preprint arXiv:1611.05397 (2016).

[28] Wang, Ziyu, Tom Schaul, Matteo Hessel, Hado Van Hasselt, Marc Lanctot, and Nando De Freitas. "Dueling network architectures for deep reinforcement learning." arXiv preprint arXiv:1511.06581 (2015).

[29] Van Hasselt, Hado, Arthur Guez, and David Silver. "Deep reinforcement learning with double q-learning." In Thirtieth AAAI conference on artificial intelligence. 2016.

[30] Liao, Rui, Shun Miao, Pierre de Tournemire, Sasa Grbic, Ali Kamen, Tommaso Mansi, and Dorin Comaniciu. "An artificial agent for robust image registration." In Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[31] Ma, Kai, Jiangping Wang, Vivek Singh, Birgi Tamersoy, Yao-Jen Chang, Andreas Wimmer, and Terrence Chen. "Multimodal image registration with deep context reinforcement learning." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 240-248. Springer, Cham, 2017.

[32] Luo, Biao, Derong Liu, Tingwen Huang, and Ding Wang. "Model-free optimal tracking control via critic-only Q-learning." IEEE transactions on neural networks and learning systems 27, no. 10 (2016): 2134-2144.

[33] Chaturvedi, Anshul, and Nidhi Mishra. "An exhaustive survey of reinforcement learning with hierarchical structure."

[34] Desai, Prajna, and Ilker Hacihaliloglu. "Knee-Cartilage Segmentation and Thickness Measurement from 2D Ultrasound." Journal of Imaging 5, no. 4 (2019): 43.

[35] Mao, Hongzi, Mohammad Alizadeh, Ishai Menache, and Srikanth Kandula. "Resource management with deep reinforcement learning." In Proceedings of the 15th ACM Workshop on Hot Topics in Networks, pp. 50-56. ACM, 2016.

CHAPTER 3

METHODS

3.1 Overview

As explained previously, we aim to develop a fully automatic, robust and real-time intra-operative US-CT registration method for femur fracture treatment. For that purpose, we proposed an RL-based registration method that consists of five main stages illustrated in Fig 3.1. It consists of data acquisition, bone shadow enhancement, point cloud modeling, hierarchical registration based on RL, and data augmentation.
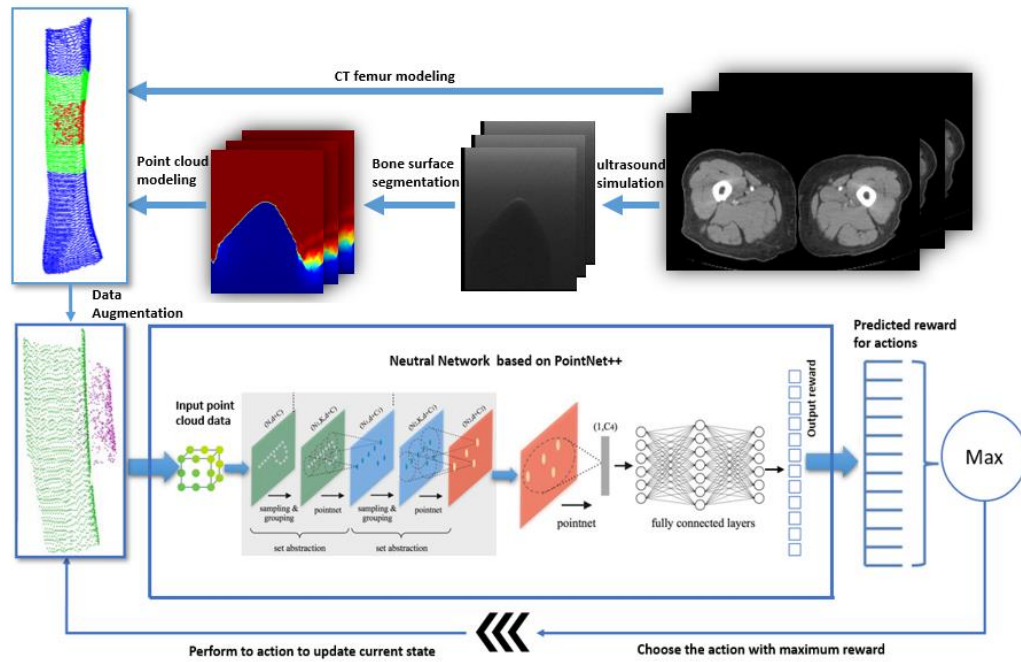
Figure 3.1 The framework of the proposed registration method using Q-learning, with CT data and simulated point clouds as the input and output for data acquisition, and the misalignment point clouds and the optimal action as the input and output for the Q-learning network.

3.2 <u>Data Acquisition</u>

3.2.1   <u>Overview</u>

When training a reliable registration agent, it requires a great number of labeled pairs with the known transformation matrix. However, it is difficult to obtain such a dataset in the medical domain. We proposed to generate the dataset mainly using 3D slicer [26] illustrated in Fig 3.2. For the femoral CT model acquisition, 3D slicer was used to visualize and segment the femur model from DICOM format and then transform them into ply format, which is a format of the point-cloud file. After getting the femoral CT data, the ultrasound simulator working on 3D slicer and Public software Library for Ultrasound (PLUS) toolkit was used to generate the consecutive simulated US image with has known transformation matrix of the CT femur model.

femur CT model acquisition from TCIA → ultrasound simulation in 3D slicer+PLUC → Generated paired US-CT
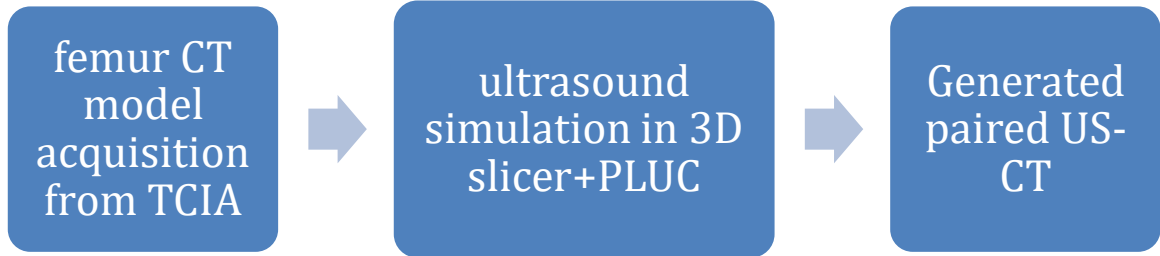
Figure 3.2 Flowchart of the data acquisition process that consists of three steps, femur model acquisition, US images simulation and pairwise US-CT dataset generation.

3.2.2   <u>Femoral CT model acquisition</u>

TCIA is a service that de-identifies and hosts an extensive archive of medical images of cancer accessible for public download [25]. The data covers different types of image modality, such as MRI, CT, digital histopathology, with different common cancer diseases.

For our search focus, those CT covers the femur could be used in further research as the cancer tissue and the other soft tissue could be eliminated later.

3D slicer is an open-source software platform for medical image informatics, image processing, and three-dimensional visualization [26]. After loading the DICOM files from TCIA into 3D slicer, the 3D model was visualized in the Volume Rendering module in CT-Bone mode and was segmented in the Segmentations module [27].
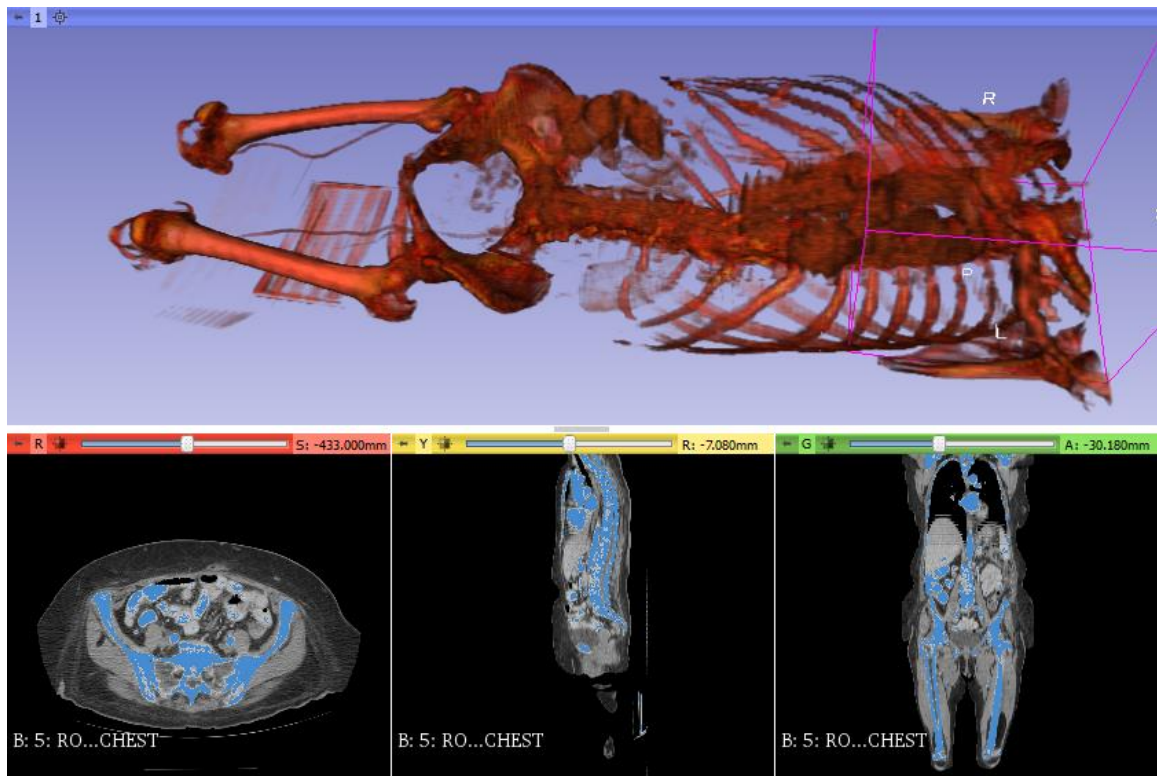
Figure 3.3 3D slicer in the Segmentation module. The 3D human models were shown in the upper in CT-Bone mode that was segmented from the CT data. And the CT data from TCIA was shown in the lower where the blue portion indicated the segmented area based on the threshold range.

In the Segmentations module, we first used a threshold effect in intensity to create a surface map, in which the intensity range is from 150 to 1600 after trails as shown in Fig 3.3. Then the model was segmented and visualized based on the surface map. Next, the Draw and Grow-Cut effects were used to manual select the region of interest of the bone we wanted and then generate the surface map again optionally. The operation process mentioned above could be repeated until the femur bones were segmented clearly, the model was exported and saved as a point cloud file in the Segmentation module and 10 different CT femur models were generated from 8 different patients in this procedure.

3.2.3   Ultrasound images simulation

The PLUS is an open-source toolkit for US acquisition, pre-processing and calibration for navigated image-guided intervention [28]. The simulated US images were acquired using Ultrasound Simulator in 3D slicer, SlicerIGT [29] and PLUS toolkit.

In the US Simulator [33], PLUS toolkit generated and then sent the synthetic US images to 3D slicer in real-time, with a configuration file as the input. The configuration file contains the synthesized ultrasound properties, which consists of three main components, the acoustic properties and surface meshes of the models, the properties of the synthetic ultrasound machine and the properties of the algorithm. The scanlines were computed based on the properties in the configuration file, which includes the attenuation and absorption of the models, the refraction and the speed of sound [28].

Then 3D slicer visualized the simulated US images from PLUS toolkit with the same models loaded into PLUS toolkit that contained the linear C5-2_60 transducer and the segmented CT femur model in the previous process. The positions and the orientations

information of each object in 3D slicer can be obtained in Transform Module in real-time, which means the different position and orientation of the simulated US image can also be obtained with respect to the transducer.

In order to simulate the tracked 3-D US images, the 3D US images can be collected using Transform module in 3D slicer as shown in Fig 3.4. As the transducer moved along one direction with specific spacing, a consecutive sequence of simulated US images with known spatial information were generated which can be considered as 3D volumetric US images. And the spacing between two consecutive simulated US images in this work was set as 0.5 ~ 0.6 mm which was comparable to the actual tracking US system.

(a)   (b)
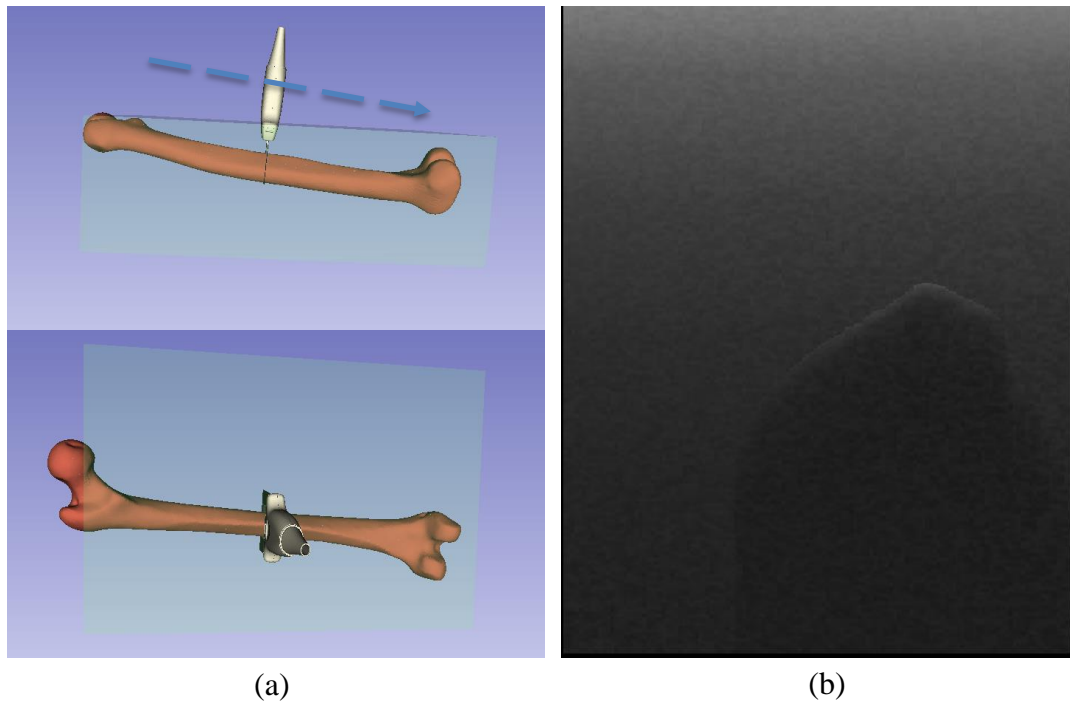
Figure 3.4 (a) 3D layout view of US simulator in 3D slicer. The linear probe in the initial position was shown in two different views. The blue arrow indicates the data collection direction of the transducer. (b) The simulated ultrasound image from the current position of the transducer.

For the whole femur model, we used US simulator to generate 500 ~ 600 US images from one femoral CT model we generated previously, which depends on the length of the CT femur model. Here, And the size of the output US image was 860x616, where the resolutions were 0.084mm and 0.087mm, respectively. The initial positions of the femur model and the probe are [0, 0, 0] and [-10, -80, -15.46] respectively in the system coordination. Since there are four different coordination systems in the ultrasound simulator, the system, the transducer, the US image, and the femur bone coordination system. The referenced position between US images and femur bone would be described in the following point cloud modeling section.

3.3 <u>Femur bone surface enhancement</u>

Acoustic shadows occur in interfaces where there is a high impedance difference, such as air-tissue, tissue-bone, and tissue-lesion in 2D US [1]. Bone shadow can aid in the interpretation, such as identification of spinal levels, and has been incorporated as an additional feature to improve the segmentation of bone surfaces from US data. Real-time feedback of bone shadow information can also be used to guide the clinician to a standardized diagnostic viewing plane with minimal artifacts. In other words, the enhancement effectively removes the irrelevant tissue information during image guidance for femoral shaft treatment.

The traditional approaches for bone shadow enhancement in 2D US images based on intensity or gradient-variant have shown its significant unreliable and limitations [2-4]. As the gradient information is unpredictable and could be affected by the imaging conditions,

<div align="center">(a)   (b)</div>
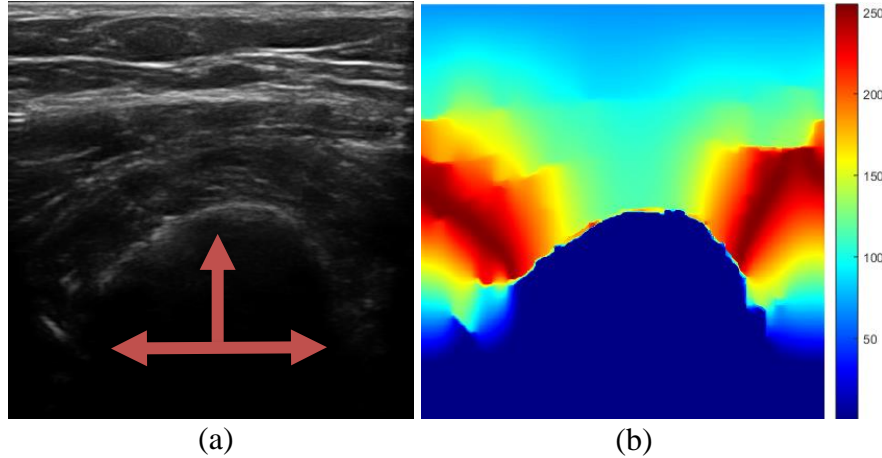
Figure 3.5 (a) B-mode US image of femur obtained in vivo. Red arrows indicate the bone shadow region. (b) The bone shadow region enhancement is shown in jet color coding. The bone shadow region in the enhanced image is color coded in blue.

the soft tissue and bone surfaces that are both denoted as high-intensity values would be both enhanced. The detection and segmentation of bone surface could be affected significantly. Thus, instead of using the traditional intensity-variant based method, we used local phase features that are intensity-invariant to enhance the tissue/bone interface in 2D US images as shown in Fig 3.5. Local phase features are crucial for analyzing the structural information of the image, as they could be considered as the interpretation of the visual appearance of the image [5].

The US response profile is highly correlated with the orientation of the transducer in a bone and dominant ridge-like edge along the scan line. In order to localize the presence of tissue/femur interface, the images are enhanced with the combination of three different local phase features, local phase tensor $LPT(x, y)$, local phase energy $LPE(x, y)$ and local weighted mean phase angle $L\omega PA(x, y)$.

The $LPT(x, y)$ is calculated using the gradient energy tensor GET filter that provides simultaneous analysis of local orientation and phase information [6]. The GET filter response is defined as:

$$GET\big(US_{DB}(x, y)\big) = T_{even} + T_{odd} = \begin{bmatrix} GET_{11} & GET_{12} \\ GET_{21} & GET_{22} \end{bmatrix} \tag{3.1}$$

where

$$T_{even} = \big[H\big(US_{DB}(x, y)\big)\big] \big[H\big(US_{DB}(x, y)\big)\big]^T \tag{3.2}$$

$$T_{odd} = -0.5 \,([\nabla US_{DB}(x, y)] \, [\nabla\nabla^2 US_{DB}(x, y)]^T +$$

$$[\nabla\nabla^2 US_{DB}(x, y)] \, [\nabla US_{DB}(x, y)]^T) \tag{3.3}$$

The equations of T_{even} and T_{odd} represent symmetric and asymmetric features respectively. And H, ∇ and ∇^2 denotes the Hessian, Gradient and Laplacian operations. $US_{DB}(x, y)$ is a distance weighted and band-pass filter that improves the enhancement result for the bone surface located deeper in the image and reduces the error enhancement for the soft tissue interfaces that located closer to the transducer surface. As band-pass filtering was obtained based on the even and odd symmetry response of Log-Gabor filter, the $LPT(x, y)$ is calculated as:

$$LPT(x, y) = \sqrt{T_{even}^2 + T_{odd}^2} \times \cos \emptyset \tag{3.4}$$

Here, $\emptyset$ represents the instantaneous phase obtained from the even and odd feature response [7].

In order to get a more compact bone surface representation with less soft tissue information, the $LPE(x, y)$ and $L\omega PA(x, y)$ are computed using monogenic signal theory [8, 9], which is determined using Reisz filter in Fourier domain representation:

$$H_1(u_1, u_2) = \frac{u_1}{\sqrt{u_1^2 + u_2^2}}, \quad H_2(u_1, u_2) = \frac{u_2}{\sqrt{u_1^2 + u_2^2}} \tag{3.5}$$

And the monogenic signal image $US_M(x, y)$ is formed by the combination of $LPT(x, y)$ and Reisz filtered component as follows:

$$US_M(x, y) = [LPT_B(x, y), LPT_B(x, y) \times h_1(u_1, u_2), LPT_B(x, y) \times h_2(u_1, u_2)] \tag{3.6}$$

Where $h_1(u_1, u_2)$ and $h_2(u_1, u_2)$ denote as the spatial representation of Reisz filter and * is convolution operator. In order to further enhance the edge information in the image, the α-scale space derivative quadrature filter (ASSD) is used for band-pass filtering, which is defined as [10]:

$$ASSD(\omega) = \begin{cases} n_c \omega^\alpha exp(-\sigma \omega^{2\alpha}) & \omega \geq 0 \\ 0 & otherwise \end{cases} \tag{3.7}$$

In the above equation, In the above equation, a is a constant derivative parameter which is chosen to be α > 1 for the filters to satisfy the DC condition [1]. n_c is a unit normalization constant calculated from the filter α value, and σ is the filter α-scale parameter [10]. The $LPE(x, y)$ and $L\omega PA(x, y)$ are calculated as:

$$LPE(x, y) = \sum_{sc} |US_{M1}(x, y)| - \sqrt{US_{M2}^2(x, y) + US_{M2}^3(x, y)} \tag{3.9}$$

$$L\omega PA(x, y) = \arctan \frac{\sum_{sc} US_{M1}(x, y)}{\sqrt{\sum_{sc} US_{M1}^2 + \sum_{sc} US_{M2}^2(x, y)}} \tag{3.10}$$

In the above equations, sc represents the number of scales. Then the $LPE(x, y)$ a feature map image that obtained the underlying shape of the bone boundaries while $L\omega PA(x, y)$

preserves all the details in structure for both soft tissue and bone boundaries. Thus, the final local phase features image $LP(x, y)$ with a combination of these three different features could be represented as [1]:

$$LP(x, y) = LPT(x, y) \times LPE(x, y) \times L\omega PA(x, y) \qquad (3.11)$$

The combined image could not only suppress the soft tissue interface but make the bone surface more compact and localized, which could be used for the further enhancement and bone surface segmentation. The different local phases image and the combined image were illustrated in Fig 3.6.
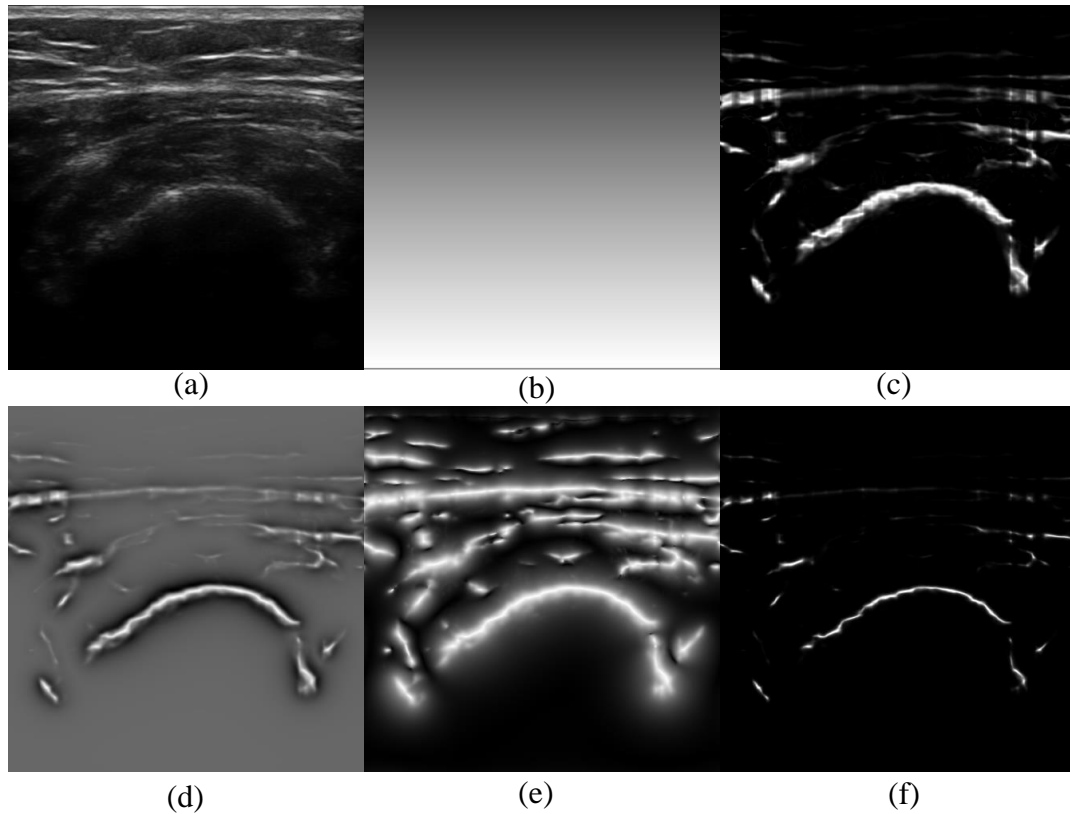
Figure 3.6 (a): B-mode US image of femur obtained in vivo. (b): Distance map used during LPT(x,y) image calculation. (c): LPT(x,y) image from (a). (d): LPE(x,y) image from (a). (e): LwPA(x,y) image from (a). (f): LP(x,y) image obtained with the combination of c-e.

The bone shadow region enhancement is based on a confidence map (CM) approach using $LP(x, y)$ image [7]. And with the consideration of US signal scattering and attenuation, the CM could be represented as:

$$CM_{LP}(x, y) = US_A(x, y)BSE(x, y) + (1 - US_A(x, y))\rho \qquad (3.12)$$

Here $CM_{LP}(x, y)$ represents the CM image using [11], and $US_A(x, y)$ denotes as the US transmission map [11], $BSE(x, y)$ is the enhanced bone shadow image and ρ is a constant value representative of echogenicity of tissue that nears the bone. In order to get the $BSE(x, y)$, we could estimate the transmission by the following rules. The $US_A(x, y)$ could be minimized by the following function:

$$\frac{\lambda}{2}\|US_A(x, y) - CM_{LP}(x, y)\|_2^2 + \sum_{J \in x}\left\|W_j \circ \left(D_j * US_A(x, y)\right)\right\|_1 \qquad (3.13)$$

In the above objective function, x is an index set and ∘ is the element-wise multiplication operator. W_j is a weighting matrix which is calculated as follow:

$$W_j(x, y) = exp\left(-|D_j * CM_{LP}(x, y)|^2\right) \qquad (3.14)$$

And D_j is obtained in a higher-order differential filter that enhances the bone surface feature as well as suppresses image noise [12].

After estimating the $US_A(x, y)$, we could get $BSE(x, y)$ as:

$$BSE(x, y) = \left[(CM_{LP}(x, y) - \rho)/[max(US_A(x, y), \epsilon)]^\delta\right] + \rho \qquad (3.15)$$

Where δ is the attenuation coefficient and ϵ is a constant that ensures the division of zero won't happen. In this work, $CM_{LP}(x, y)$ images were obtained with the following constant values: $\gamma = 0.03, \beta = 90, \eta = 2$. And for Eq. 3.12, the tissue echogenicity constant ρ was chosen as 90% of the maximum intensity value of $CM_{LP}(x, y)$. In the next step, $BSE(x, y)$

images that illustrated in Fig 3.7 are used to extract bone surface information and the simulation of point-cloud data, which we will explain in the next section.

| (a1) | (b1) | (c1) |

| (a2) | (b2) | (c2) |
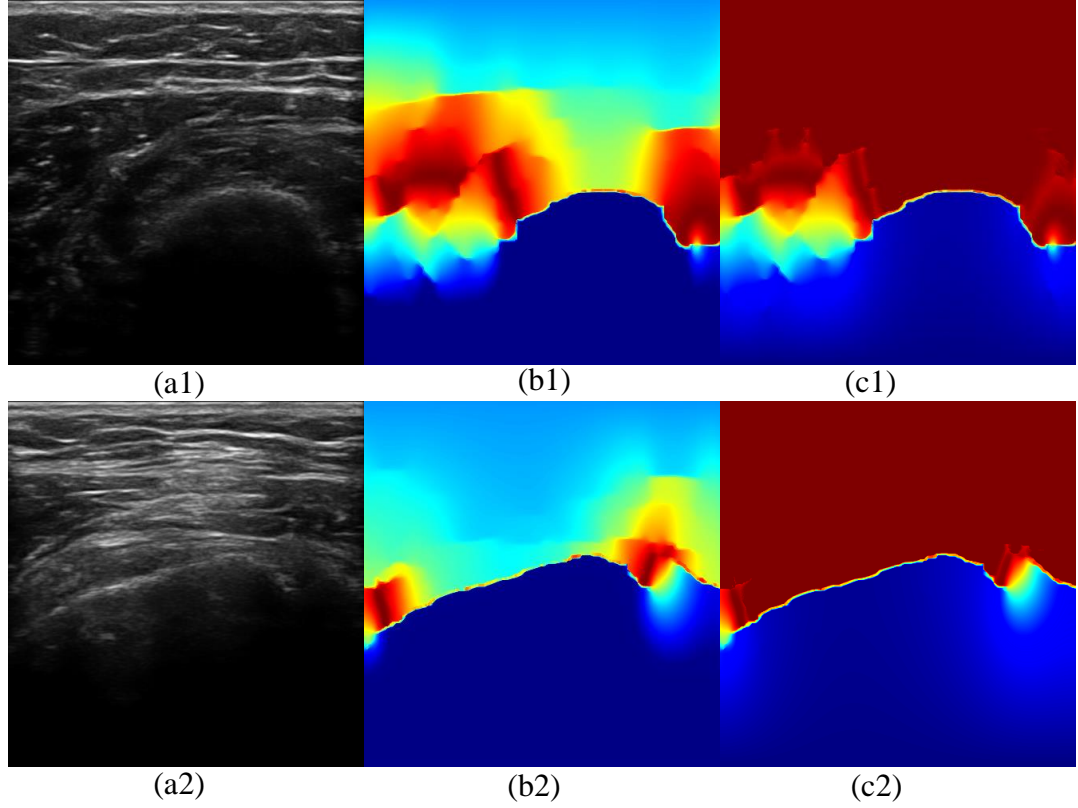
Figure 3.7 (a): B-mode US image of the femur with different rotation and position in

vivo; (b): $CM_{LP}(x, y)$ image from (a); (c): $BSE(x, y)$ image from (a) in jet color.

3.4 Point cloud modeling from bone shadow region images

Bottom up ray casting segmentation was used to obtain the bone surface information from $BSE(x, y)$ as the bone surface has much a higher intensity than its neighbor in the bottom. The 2-D bone surface information in one image could be defined as:

$$P_{2d}(x, y) = \begin{cases} (x, y); & if \ I(x, y) - I(x - 1, y) \geq \omega \\ None; & otherwise \end{cases} \qquad (3.16)$$

Where the $I(x, y)$ means the intensity of (x, y) in the $BSE(x, y)$ image. And ω is the threshold value for bone surface segmentation, which is set as 80 in this work after numerous trails.

After getting the bone surface point from $BSE(x, y)$ images, 2-D to 3-D transformation was needed to simulate the actual 3-D data. In the consecutive sequence of US images, the index of each image could be considered as the z-direction in 3D. Thus, we proposed to transform the simulated US images, which were in the US probe coordination into 3D CT femur model coordination with the help of the index information as illustrated in Fig 3.8. As those two data were in the scene of 3D slicer at the same time, the relative position between the 3-D CT femur model and simulated US images was known. The 3-D bone surface information from a consecutive sequence of the US images are calculated as follows:

$$P_{3d}(x, y, z) = \left[P_{2d}(x) * r_x + R_x, P_{2d}(y) * r_y + R_y, \mathbb{Z} * r_z + R_z\right] \qquad (3.17)$$

Where $P_{2d}(x)$ and $P_{2d}(y)$ are the 2-D position in one image, and r_x and r_y are the resolution of the US images in x-, y-direction, respectively. $\mathbb{Z}$ is the index of the consecutive sequence of the US image and r_z is the spacing between two consecutive simulated US images. After transforming the 2-D data into 3-D data, the relative position $R_x, R_x \ and \ R_x$ are added to it. Then those simulated point-based 3D bone surface data is in the femur model coordination, which means they're well registered and paired.

In the end, down-sampling is necessary as the original point-base bone surface have different size and different original length of US sequences. Down-sampling them into one fixed size makes the different size of the point-based bone surface could be represented in one size. As well as it's more inconvenience as the input of the CNN in the next steps. And

in the next section, we'll discuss the framework of the CNN and how the Q-learning works for hierarchical registration problem.
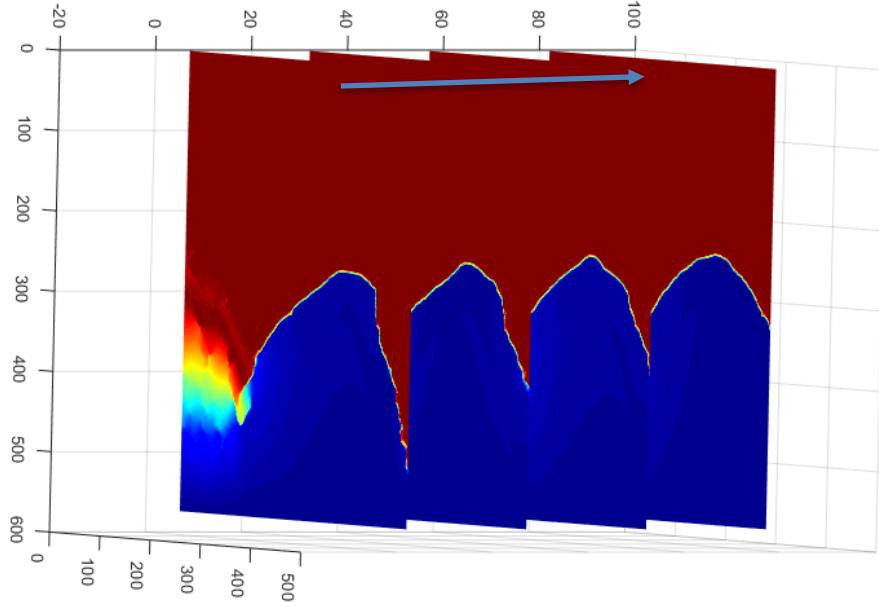
Figure 3.8 A sparse consecutive sequences of $BSE(x, y)$ images from femur US images obtained in vivo. The blue arrow indicates the z-direction being used in the point cloud modeling.

3.5 Hierarchical 3-D registration of the point cloud-based registration with RL

3.5.1 Problem formulation

As we mentioned above, the registration learning process could be treated as an optimal strategy learning process that aims to find the optimal sequence of motions to make the two point-based volumes aligned. Within the definition of the registration learning process, the agent is modeled using a deep convolution neural network, with the point-based data as the input and the optimal motion as the output.

Let P_r be a reference point-based bone surface volume and P_f be the floating point-based volume that needs to be registered to P_r. And the goal in the rigid-body registration is to estimate the transformation matrix T_g, which is a column-wise 4x4 homogeneous transformation matrix. Then for T_g, T_g is the matrix that makes the following equation workable $P_r = P_f \circ T_g$, which means the P_f and P_r are aligned with T_g. T_g can be represented in six parameters with three in translation $[t_x, t_y, t_z]$ and three in rotations $[\theta_x, \theta_y, \theta_z]$ as:

$$T_g(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z) = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & \cos\theta_x & -\sin\theta_x & t_y \\ 0 & \sin\theta_x & \cos\theta_x & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos\theta_x & -\sin\theta_x & 0 & 0 \\ \sin\theta_x & \cos\theta_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.18)$$

Then the process of finding T_g can be completed using Markov Decision Process (MDP) which defined by $\{S, A, \tau, r, \gamma\}$ [13]. S are the set of states that the agent can be reach, A is a set of action that the agent can take at a state, τ is the transition function that indicates the probability of each action to take at a state, r is the reward the agent can receive when reaches at a state and γ is the discounted factor that affects the long term reward for the agent which is set as 0.9 in this work.

At each time t, the current s_t is defined by the current transformation T_t. And the associate observation of the whole system at time t according to T_t is the difference of $P_r -$ $P_f \circ T_t$. The details of the computing the s_t is described later. Then the agent chooses an action a_t from the action set A to update the state to improve the alignment result by $T_{t+1} = a_t \circ T_t$. The action set A consists of twelve possible actions the agent can take that

leads to ± 1 changes in one element of $[t_x, t_y, t_z, \theta_x, \theta_y, \theta_z]$, such as ± 1mm in translations and $\pm 1^\circ$ in rotations. The transition function τ is defined that all available actions for the next state have an equal possibility at time t. The reward function is described explicitly below.

In the training process, the agent learns a 3-D registration policy that makes the two point-based data aligned with a sequential of optimal actions a_t^*, which is the best action at time t to improve the alignment result. In the testing process, the output of the agent system is a sequence of N consecutive actions $\{a_1^* \ldots a_N^*\}$ with the learned registration policy to achieve the correct alignment.

For this agent, the details of the MDP process (i.e. input, reward, etc.) and the deep neural network for the MDP based reinforcement learning are described in detail in the following sections.

3.5.2 <u>The supervised Target DNN-based Q-learning</u>

3.5.2.1 <u>The input of the agent system</u>

As we have registered pairs of US and CT point cloud models in the data acquisition procedure, then misalignment can be introduced to the simulated point cloud model that considered as P_r and the CT point cloud model was treated as P_f before using them in one registration training process. Moreover, the introduced transformation should be T_g as the gold standard in the training process.

However, as the point-based data is unordered and the size of P_f and P_r are different, the difference between two point-based data that can be represented the current registration situation cannot be computed by a simply minus operation, like the image. For that, planar

projection and k-nearest neighbors (k-NN) are used to compute the corresponding point of P_f in P_r which should be the s_t.

The rotation difference between the P_r and P_f may cause some problems in finding the nearest neighbor points as the nearest neighbor point is defined in the spherical coordination. We should specify the direction of it using planar projection as illustrated in Fig 3.9. The direction of the projective plane is defined as:

$$\theta = arctan\left(\frac{\sqrt{a^2+b^2}}{c}\right) = arccos\left(\frac{c}{\sqrt{a^2+c^2+c^2}}\right); \; \varphi = arctan\left(\frac{b}{a}\right) \qquad (3.19)$$

$$U = \{-\sin\theta, \; \cos\theta, \; 0\}; V = \{\cos\theta * \sin\varphi, \; \sin\theta \sin\varphi, \cos\varphi\} \qquad (3.20)$$

Where $\{a, b, c\}$ is the normal vector of the projected direction, θ and, φ are the rotation angle in x -y plane and z-[x-y] plane of $\{a, b, c\}$. U and V are the normal vectors that represent the horizontal and vertical direction of the projective plane, respectively.
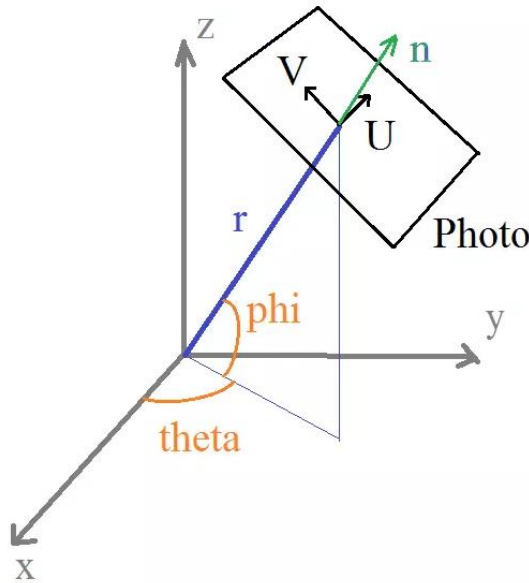
Figure 3.9 Illustration of the planar projection method. It shows the relationship of the projective plane and its relative parameters with the normal vector of the perspective direction.

And then, the points in the projective plane is calculated as:

$$center = \{\cos\theta * \cos\varphi, \ \sin\theta\cos\varphi, \sin\varphi\}$$

$$PP(x', y') = [(P(x, y, z) - center) \cdot U, \ (P(x, y, z) - center) \cdot V] \qquad (3.21)$$

Where the center is the center point of the projective plane and $PP(x', y')$ is the projected 2-D point of $P(x, y, z)$ in the projective plane.

K-NN is a non-parametric method used for classification and regression [15]. For the classification k-NN, if k = 1, the object is assigned to its nearest neighbor which means we can use it to find the nearest point of each point. The nearest neighbor points P_n can be

(a)   (b)
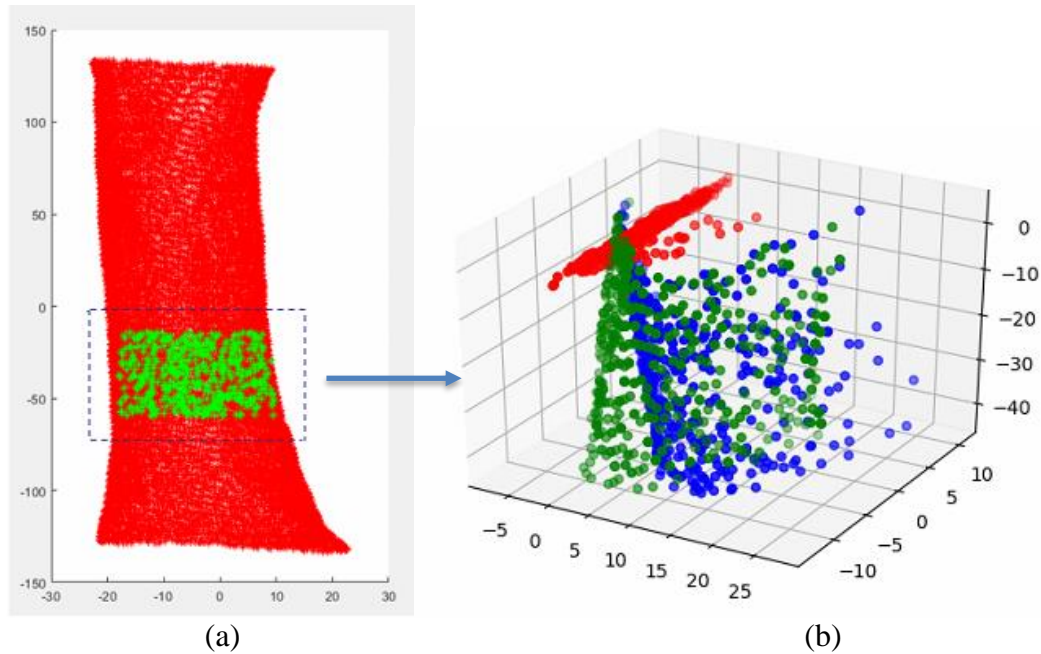
Figure 3.10 (a) The CT and US point cloud models in the projective plane. The red point set is CT model while the green is US model. (b) The KNN results for the US model in a. The green point set is the US model, and the green point set is the neighbor point in CT model for that US model. And the red point set is the difference between those two point sets with the same size.

treated as the 3-D corresponding point of P_f in P_r. And the input of the system s_t is defined as:

$$s_t = P_c(x, y, z) = P_n - P_f \circ T_g;$$

$$P_n = \left[knn\left(PP_r(x, y), PP_f(x, y) \right) \right]_{3d} \tag{3.22}$$

s_t is the difference of P_r and P_f which has the same size as P_f. P_n is 3-D nearest points between P_f and P_r which transformed from the 2-D nearest points. We assumed that if P_n and $P_f \circ T_g$ are aligned, s_t should be a set of nearly zero points as illustrated in Fig 3.10. With respect to that, the learning agent is learning how to make the s_t become a set of nearly zero points.

3.5.2.2 The supervised Q-learning

For the agent, the policy learning process can be formulated as a reinforcement learning problem [14, 16-17]. And we mentioned above, Q-learning is an MDP-based learning agent to learn a policy to under what circumstances [18]. And the action-value function in Q-learning is defined as:

$$Q_t(s_t, a_t) = max_\tau \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+1} + \cdots | s_t, a_t, \tau] \tag{3.23}$$

Where the actions-value function is estimated by a DNN and iteratively updated by the Bellman Equation [19]. Also, the action-value function using iterative value update can be represented as:

$$Q_t(s_t, a_t) \leftarrow (1 - \alpha)Q_t(s_t, a_t) + \alpha \cdot \left(r_t + \gamma \cdot \max_a Q(s_{t+1}, a) \right) \tag{3.24}$$

Where α is the learning rate ($0 < \alpha \leq 1$), r_t is the reward the agent achieved when moving from the state s_t to the next state s_{t+1} after taking action a_t, $\max_a Q(s_{t+1}, a)$ is the estimate of optimal future value.

In the traditional Q-learning algorithm, the iterative value update method for action-value function using DNN undergoes an unguided exploration of the agent which can result in low training efficiency. Because the agent must try a great number of combinations of different actions to achieve the optimal registration path. Then, a supervised training path is used here to instruct the agent to follow a greedy registration path which mimics how human register two objects in a most efficient manner [20]. Thus, the optimal action a_t^* at time t along the supervised registration path is computed as:

$$a_t^* = \min_{a_t \in A} D\left(T_g, a_t \circ T_t\right) \tag{3.25}$$

Where a_t^* aims to find the minimum difference or distance between the new and the ground truth transformation matrix. $D\left(T_g, a_t \circ T_t\right)$ is the distance between two transformation matrices using the Euclidean norm of the 6-D parameters of $T_g \circ (a_t \circ T_t)^{-1}$ with the Eqn. 3.18. And the Euclidean norm of the 6-D parameter v_t is captured by the formula:

$$\left\|v_t\left(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z\right)\right\|_2 = \sqrt{t_x{}^2 + t_y{}^2 + t_z{}^2 + \theta_x{}^2 + \theta_y{}^2 + \theta_z{}^2} \tag{3.26}$$

If there are more than one action leads to the optimal action at time t, those actions are taken with equal probabilities. Without loss of generality, the exploration limitation in transformation parameter space is within ± 20mm translation for x, y-direction, ± 30mm translation for z-direction and $\pm 15°$ rotation for x-, y-, z-direction. As we have pre-alignment before using this agent, the limitation of these parameters is comparatively small

but corresponding to the maximum possible misalignment of the two volumes to be registered.

Therefore, the action-value Q-function under a supervised greedy path can be computed as in a recursive manner, and we assumed the agent could run sufficient steps to reach the optimal alignment position:

$$Q(s_t, a_t) = \begin{cases} r(s_t, a_t) + \gamma Q(s_{t+1}, a_t^*), & if \ D(T_g, a_t \circ T_t) > \epsilon \\ r(s_t, a_t) + R, & o.w. \end{cases} \quad (3.27)$$

Where ϵ is the distance tolerance, R is a bonus reward and $r(s_t, a_t)$ is the immediate reward function. And the immediate reward $r(s_t, a_t)$ for the agent when taking a_t at s_t is formulated as:

$$r(s_t, a_t) = D(T_g, T_t) - D(T_g, a_t \circ T_t) \quad (3.28)$$

The agent is considered as successful when the distance between the current T_t and the ground truth T_g transformation is not larger than the tolerance $\epsilon = 0.5$. When the agent achieved success, it received a bonus reward R. The maximum immediate reward of the system is 1 with the continuous steps size 1, with the proof below:

$$r(s_t, a_t) = D(T_g, T_t) - D(T_g, a_t \circ T_t)$$

$$= \|v_t\|_2 - \|v_{t+1}\|_2 \leq \|v_t - v_{t+1}\|_2 = \|v_{t+1} - v_t\|_2 = 1 \quad (3.29)$$

Then, with the Bellman Equation, the maximum action-value function should be nearly 10. Thus, the success bonus reward R is 10 which is estimate Q-value for the whole system. From the Eqn.3.27, we know that the agent can perform the registration process by choosing the action with maximum Q-value during testing.

Based on the structure of the agent, with the difference point-based volume P_n as the input and 12 nodes as the outputs which corresponds the action in action set *A*. For that, mean square root error (RMSE) as the loss function:

$$Loss = \sum_{k=1}^{M} \sum_{a_i \in A} \|y_i(P_k) - Q(s_k, a_i)\|_2 \qquad (3.30)$$

Where $y_i(P_k)$ is the *i*-th output of the DNN for the *k*-th sample among the M training samples. In addition, the learning rate for the agent was 0.000006 with a decay of 0.65 every 5,000 mini-batch based back propagations.

3.5.2.3 PointNet++ based DNN

Due to the irregular format of the point cloud, PointNet++ is a novel deep neural network that can directly process the point cloud data instead of rendering data voluminously [21, 22]. PointNet++ is able to learn the deep local features of the point set from multiple scales by exploiting metric space distance.
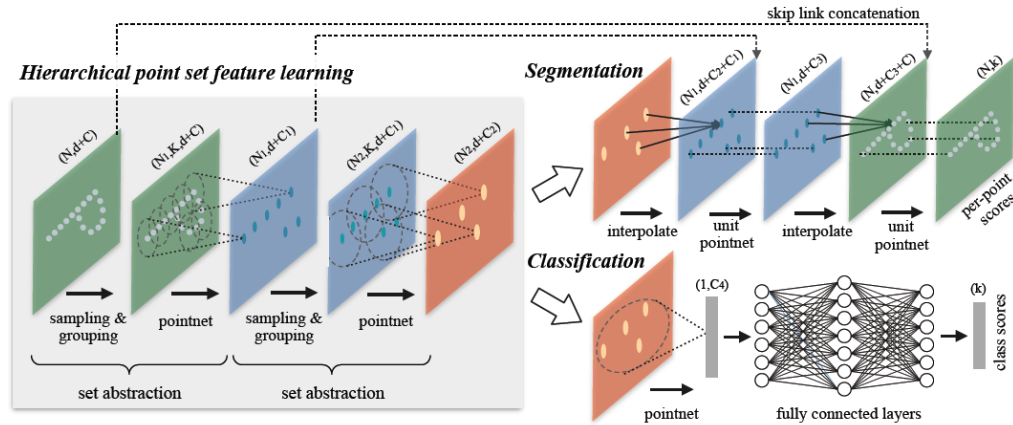
Figure 3.11 Illustration of the architecture of the classical hierarchical feature learning from PointNet++, which contains three different layers. It also showed the application of PointNet++ in set segmentation and classification using points after using the result of the hierarchical point set feature learning from [22].

And from Fig 3.11 we can see that the architecture of PointNet++ consists of three parts, hierarchical point set feature learning, feature learning under non-uniform sampling density, and point feature propagation for segmentation. Here, we used the architecture of PointNet++ to build our own DNN that learns the feature of out inputs for the agent which is identified as the Q-function for the system.

For the hierarchical point set feature learning, it is composed of a number of set abstraction levels, which consists of sampling layer, grouping layer, and PointNet layer. At each level, the new few feature points were processed and abstracted. The input of a set abstraction level is $N \times (d + C)$ matrix, where N is the number of the matrix, d is the coordination dimension, and C is the dimension of point features. And it outputs $N' \times (d + C')$ number of points with new total number and point feature dimension. For the sampling layer, the iterative farthest point sampling (FPS) is used to choose a subset of points, which defines the centroids of local regions [23]. And the output of sampling layer is $N' \times d$, a point set of the coordinates of centroids. Then, the grouping layer takes $N \times (d + C)$ matrix and $N' \times d$ as the input and outputs groups of point set $N' \times K \times (d + C)$, where K is the number of points in the neighborhood of centroid points. At the last layer, PointNet layer, of the set abstraction, it takes $N' \times K \times (d + C)$ as input. Each local point features for each region is abstracted from its centroid point with the output data size is $N' \times (d + C')$. For the local patterns learning process, we used the following function in PointNet [21]:

$$f\{x_1, x_2, \dots, x_n\} = \gamma \left(\max_{i=1,\dots,n} \{h(x_i)\} \right) \qquad (3.31)$$

Where $\{x_1, x_2, \dots, x_n\}$ is a set of points with $x_i \in \mathbb{R}^6$, and γ and h are multi-layer perception networks (MLP).

As the varying density among the different regions of one single point cloud, feature grouping from regions across different scales is necessary for the further global feature learning process. Two density adaptive layers, multi-scale grouping, and multi-resolution grouping, are used to solve this problem whose arichitecure were shown in Fig 3.12. Multi-scale grouping is deployed by randomly dropping out input points with a randomized probabilities ϑ, which we set as 0.95 here. It means each input can be dropped with probability $(1 - \vartheta)$. And the multi-resolution grouping layer, the weighted local regions at each level are generated according to its density. The region with sparse density is weighted lower and considered less reliable as it suffers more from sampling deficiency. When the density of a local region is high, the local features have more fine details information because it can inspect at higher resolution recursively in lower levels.

The architecture of DNN for the agent is based on those layers we mentioned above. It consists of three set abstraction levels, one PointNet layer, and three fully connected layers. The input size of those set abstraction levels is 512, 256 and 128. And the input and output data size of PointNet layer is 160 and 256. After taking the output of a single PointNet layer as the input, the three fully connected layers output 12 nodes, which denotes as the estimation Q-value of each a_t in action set A in s_t.
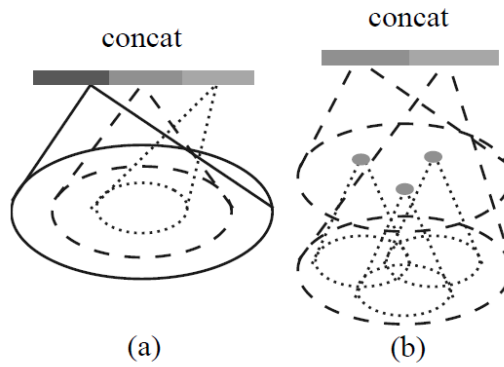
Figure 3.12 (a) Multi-scale grouping (b) Multiresolution grouping from [22]

3.6 <u>Data augmentation</u>

3.6.1 <u>Data Augment in registered models</u>

As we mentioned above, training a DNN requires a high number of labeled pairs with the known transformation matrix. Thus, we proposed two different ways to solve this problem. The first is to introduce randomly rigid-body motions to the femur model which could augment the available labeled data.

Each femur models are artificially introduced a rigid-body transformation P_r, where P_r is a 4x4 rigid-body rotation matrix as:

$$P_r = R_x(\alpha)R_y(\beta)R_z(\gamma) \tag{3.32}$$

The angle α, β, γ are the rotation angle in x-, y- and z-direction respectively. They are randomly and independently generated from $[-10^{\circ}, 10^{\circ}]$. Five different rigid-body transformations were made towards the CT femur models.

The second method is to segment the whole simulated femur US images based on different positions and different lengths. For each femur model, we divided it into eight portions in the same length. Then the length of the portion was changed and repeated the division operation to obtain the new dataset. And the length of the portion was set as [45mm 50mm 55mm 60mm]. These operations result in 10x5x8x4 = 1600 labeled pair test cases. And the testing and training dataset were separately randomly with an 8:2 ratio and came from different patients.

3.6.2   Generation of Input pairs

However, the generated pairs from the previous operations are still not enough to secure a reliable dataset for training. The different transformations should be introduced to enlarge the dataset.

Due to the complex anatomy structure and non-convex nature in-vivo, the initial alignment should be made before fine-registration in the US-based CAOS system. Here, we degraded the full length of P_r into a slightly shorter model. The truncated model which was 30 mm longer than P_f in both side of z-direction can be considered as the initial estimation position for P_f in the training process, which could also improve the time-efficiency for the convergence of the DNN.

(a)   (b)   (c)
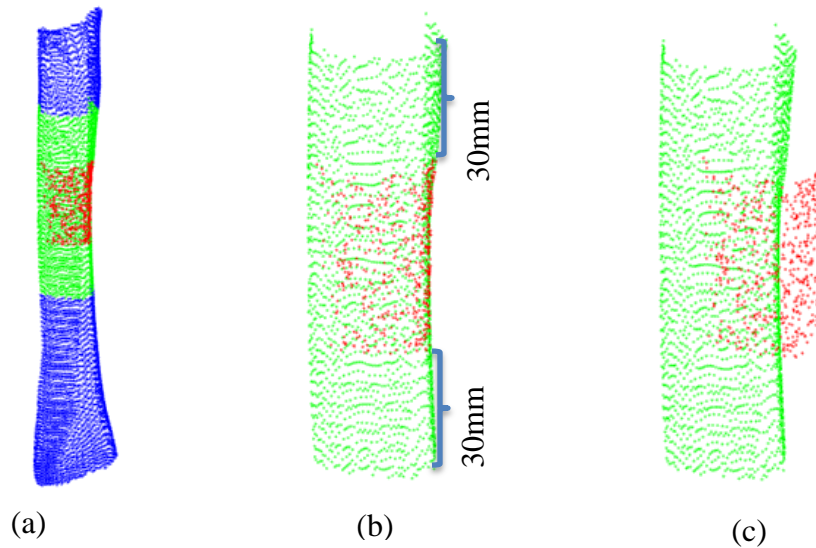
Figure 3.13 (a) The point cloud models from CT and US showing in blue and red, respectively. And the green point set represents the truncated CT model from the CT model. (b) The registered point cloud models of the truncated CT and US were shown in green and red from a. (c) The manual rotation and translation were introduced from b randomly, which is being used in the training process.

The size of the truncated model responses with the workable range of the agent. In addition, only the top surface that is visible in the US data of the femur bone from CT was used in the generation of input. Thus, the truncated P_r and P_f with known T_g from the randomly rigid-body permutation are the inputs for the agents in each training progress. And these models were shown in Fig 3.13.

As we mentioned above, the input pairs were randomly selected from the registered pairs and then generated transformation to the simulated US point set before each registration process. Then each aligned pair was randomly de-aligned using rigid-body permutations within the same range of workable range for the agent, which is $[\pm20\text{mm}, \pm20\text{mm}, \pm30\text{mm}, \pm15°, \pm15°, \pm15°]$. Since the goal of registration is to estimate the transformation between each pair of the misaligned models, multiple transformations should be applied in the registered models to fully cover the parameter space. Thus, during training, each registered US-CT pair was augmented 3200 times by introducing random transformation during training process, resulting in more than 5M training data, which should be sufficient and unique enough for the reliable outcomes of the system.

3.7 Quantitative analysis on point cloud modeling and registration results

The evaluation of the accuracy of simulated point bone surface cloud and alignment results between the US-CT can be interpreted using quantitative measurement between two point clouds. Two types of quantitative measurements are used here, target registration error (TRE) and Fiducial registration error (FRE).

3.7.1 <u>Fiducial registration error (FRE)</u>

The FRE for 3-D registration is the error in finding the correct position of fiducial

markers, which equals the root mean square distance in fiducial alignment between two

3-D models after registration [30]. And the FRE is computed as:

$$FRE^2 = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|^2 \tag{3.33}$$

Where N is the total fiducial number, and x_i and y_i is the corresponding fiducial marker

points in those two point clouds. If the alignment result is good, FRE should be nearly

zero. In our analysis, six fiducial markers were selected manually in US point clouds.

And the corresponding markers were automatically selected from the neighbor models

for CT which have the same index with the US point clouds.

3.7.2 <u>Target registration error (TRE)</u>

The TRE is the distance between corresponding points other than fiducial points after

registration, which does not need manual fiducial points selection [31]. TRE is also

interpreted as a root-mean-square error but is uncorrelated with FRE [32]. The term target

is used to suggest that the corresponding point with the same index in the point cloud pair

in our system. And TRE is computed as:

$$TRE^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{3}|x_{ij} - y_{ij}|^2 \tag{3.34}$$

Where j in the dimension number, ranging from 1-3, and x_{ij} and y_{ij} are the j-dim

coordination value of corresponding points for point cloud pair. And for our system, we

chose the total 512 number of points in the floating-point cloud as the interest points here.

Generally, TRE is smaller than FRE that can be used as the lower bound of the registration

evaluation. And FRE can be used as the upper bound of the quantitative analysis.

Reference

[1] Hacihaliloglu, Ilker. "Enhancement of bone shadow region using local phase-based ultrasound transmission maps." International journal of computer assisted radiology and surgery 12, no. 6 (2017): 951-960.

[2] Rupin Dalvi, Rafeef Abugharbieh, Mark Pickering, Jennie Scarvell, and Paul Smith, "Registration of 2d to 3d joint images using phase-based mutual information," in Medical Imaging 2007: Image Processing. International Society for Optics and Photonics, 2007, vol. 6512, p. 651209.

[3] Magnus Hemmendor, Mats T Andersson, Torbjorn Kronander, and Hans Knutsson, "Phase-based multidimensional volume registration," IEEE Transactions on Medical Imaging, vol. 21, no. 12, pp. 1536-1543, 2002.

[4] Matthew Mellor and Michael Brady, "Phase mutual information as a similarity measure for registration," Medical image analysis, vol. 9, no. 4, pp. 330-343, 2005.

[5] Bull, David R. Communicating pictures: A course in Image and Video Coding. Academic Press, 2014.p45

[6] Felsberg, Michael, and Ullrich Köthe. "Get: The connection between monogenic scale-space and gaussian derivatives." In International Conference on Scale-Space Theories in Computer Vision, pp. 192-203. Springer, Berlin, Heidelberg, 2005.

[7] Hacihaliloglu, Ilker, Abtin Rasoulian, Robert N. Rohling, and Purang Abolmaesumi. "Local phase tensor features for 3-D ultrasound to statistical shape+ pose spine model registration." IEEE transactions on Medical Imaging 33, no. 11 (2014): 2167-2179.

[8] Felsberg, Michael, and Gerald Sommer. "A new extension of linear signal processing for estimating local properties and detecting features." In Mustererkennung 2000, pp. 195-202. Springer, Berlin, Heidelberg, 2000.

[9] Felsberg M, Sommer G (2001) The monogenic signal. IEEE Trans Signal Process 49(12):3136–3144

[10] Ahror Belaid and Djamal Boukerroui, "scale spaces filters for phase based edge detection in ultrasound images," in Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on. IEEE, 2014, pp. 1247-1250.

[11] Athanasios Karamalis, Wolfgang Wein, Tassilo Klein, and Nassir Navab, "Ultrasound confidence maps using random walks," Medical image analysis, vol. 16, no. 6, pp. 1101-1112, 2012.

[12] Meng, Gaofeng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. "Efficient image dehazing with boundary constraint and contextual regularization." In Proceedings of the IEEE international conference on computer vision, pp. 617-624. 2013.

[13] Bellman, Richard. "A Markovian decision process." Journal of mathematics and mechanics (1957): 679-684.

[14] Hasselt, Hado V. "Double Q-learning." In Advances in Neural Information Processing Systems, pp. 2613-2621. 2010.

[15] Altman, Naomi S. "An introduction to kernel and nearest-neighbor nonparametric regression." The American Statistician 46, no. 3 (1992): 175-185.

[16] Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves et al. "Human-level control through deep reinforcement learning." Nature 518, no. 7540 (2015): 529.

[17] Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser et al. "Mastering the game of Go with deep neural networks and tree search." nature 529, no. 7587 (2016): 484.

[18] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." Machine learning 8, no. 3-4 (1992): 279-292.

[19] Dixit, Avinash K., and John JF Sherrerd. Optimization in economic theory. Oxford University Press on Demand, 1990.

[20] Liao, Rui, Shun Miao, Pierre de Tournemire, Sasa Grbic, Ali Kamen, Tommaso Mansi, and Dorin Comaniciu. "An artificial agent for robust image registration." In Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[21] Qi, Charles R., Hao Su, Kaichun Mo, and Leonidas J. Guibas. "Pointnet: Deep learning on point sets for 3d classification and segmentation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652-660. 2017.

[22] Qi, Charles Ruizhongtai, Li Yi, Hao Su, and Leonidas J. Guibas. "Pointnet++: Deep hierarchical feature learning on point sets in a metric space." In Advances in neural information processing systems, pp. 5099-5108. 2017.

[23] Eldar, Yuval, Michael Lindenbaum, Moshe Porat, and Yehoshua Y. Zeevi. "The farthest point strategy for progressive image sampling." IEEE Transactions on Image Processing 6, no. 9 (1997): 1305-1315.

[24] Pal, Sankar K., and Sushmita Mitra. "Multilayer perceptron, fuzzy sets, and classification." IEEE Transactions on neural networks 3, no. 5 (1992): 683-697.

[25] Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. Journal of Digital Imaging. 2013; 26(6): 1045-1057. doi: 10.1007/s10278-013-9622-7

[26] Kikinis, Ron, Steve D. Pieper, and Kirby G. Vosburgh. "3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support." In Intra-operative imaging and image-guided therapy, pp. 277-289. Springer, New York, NY, 2014.

[27]https://www.slicer.org/wiki/Documentation/Nightly/Modules/Segmentations

[28] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi and G. Fichtinger, "PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems," in IEEE Transactions on Biomedical Engineering, vol. 61, no. 10, pp. 2527-2537, Oct. 2014. doi: 10.1109/TBME.2014.2322864

[29] Ungi, Tamas, Andras Lasso, and Gabor Fichtinger. "Open-source platforms for navigated image-guided interventions." (2016): 181-186.

[30] Fitzpatrick, J. Michael, Jay B. West, and Calvin R. Maurer. "Predicting error in rigid-body point-based registration." IEEE transactions on medical imaging 17, no. 5 (1998): 694-702.

[31] Fitzpatrick, J. Michael, and Jay B. West. "The distribution of target registration error in rigid-body point-based registration." IEEE transactions on medical imaging 20, no. 9 (2001): 917-927.

[32] Datteri, Ryan D., and Benoît M. Dawant. "Estimation and reduction of target registration error." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 139-146. Springer, Berlin, Heidelberg, 2012.

[33] Bartha, Laura, Andras Lasso, Csaba Pinter, Tamas Ungi, Zsuzsanna Keri, and Gabor Fichtinger. "Open-source surface mesh-based ultrasound-guided spinal intervention simulator." International journal of computer assisted radiology and surgery 8, no. 6 (2013): 1043-1051.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 <u>Overview</u>

In this chapter, the qualitative and quantitative analysis of the point-cloud models from the bone surface segmentation and US-CT registration results are presented. In section 4.1, the results in point-set modeling are evaluated using TRE and FRE for both a single US image and a set of US images.

  In section 4.2, the evaluation US-CT registration results are performed using TRE and FRE to compare the other traditional point cloud-based registration and the ground truth results.

4.2 <u>Femoral point cloud modeling from US images</u>

The accuracy of point cloud modeling is crucial for the following registration process which requires it reflects the distinct anatomy structure of the femur bone surface. For the bone segmentation results in the single US image, they were compared with the manual segmentation results and expertise segmentation results that are considered as the gold standard. Moreover, the point cloud modeling results from a set of US images were compared with the CT model that has delicate quality.

4.2.1   <u>Point-cloud modeling from a single US image</u>

From the segmentation results in Fig 4.1, the tissue/bone surfaces were clear enhanced and visualized in the enhanced bone surface images. Although some virtual artifacts appeared

in the enhanced US images, the effects of them were unknown to the final registration results and would be evaluated in the next steps in point cloud modeling.

The bone segmentation results using the enhanced bone surface images as the input is shown in Fig 4.1. From Fig 4.1, the segmentation results indicate that the proposed method yields a reliable segmentation that representing the actual anatomy structure.

(a)   (b)

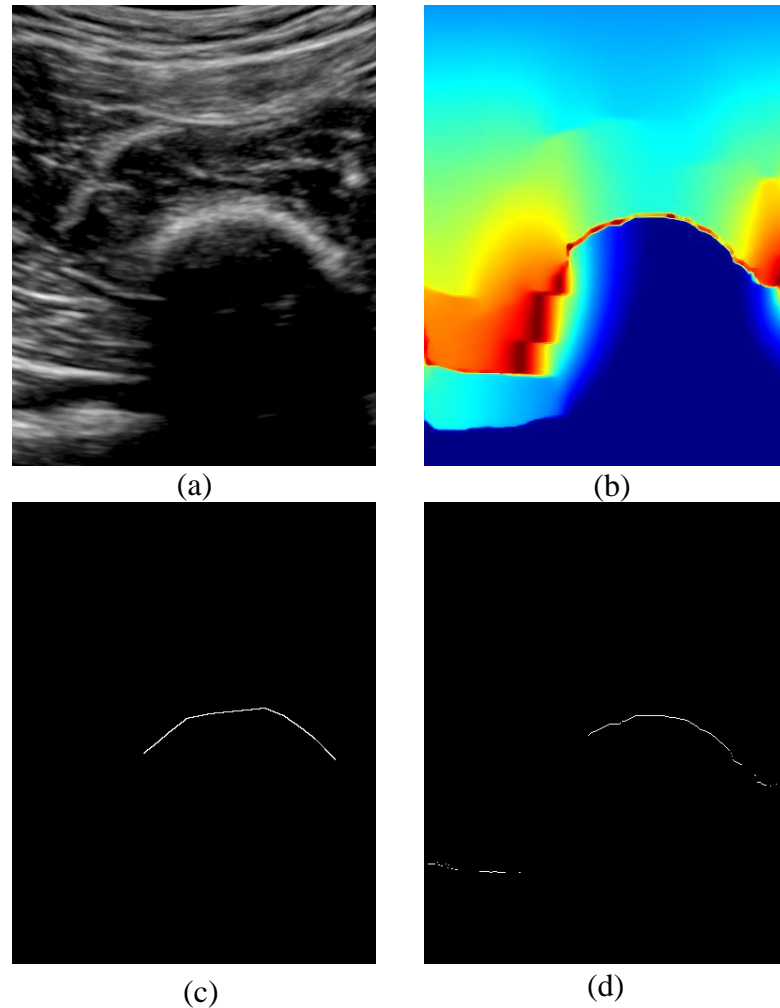(c)   (d)

Figure 4.1 (a) B-mode US image of femur obtained in vivo. (b) BSE(x,y) image from a. (c) The manual bone surface segmentation result by expertise that used as the gold standard for evaluation. (d) The automatic bone surface segmentation using the proposed thresholding method.

In the quantitative analysis, the median FRE error of the proposed segmentation method is 0.51 mm (SD: 0.08 mm) while the minimum and maximum error in FRE is 0.26 mm and 0.93 mm respectively on 100 images. The quantitative results demonstrate the potential of applying the segmentation results from the US image in the future registration method as it has a comparable result with the expertise segmentation results.

4.2.2   Point-cloud modeling from a set of US images

The point-cloud modeling method takes the bone surface information as the input and integrates with the index of a set of US images as the depth information to reconstruct the femur point-cloud models. The qualitative result of modeling is shown in Fig 4.2. Comparing with the BSE(x,y) images, the point cloud modeling result was less noisy. And the result showed that the point-cloud models can represent the anatomy structure of the femur bone surface and has proper alignment with the CT models. Moreover, the virtual artifact appears in the enhanced US images cause some noisy points in the simulated point cloud models. However, those noise points can help to improve the robustness of the proposed DNN-based registration method.

	Mean	Min	Max
FRE	0.21 mm	0.16 mm	0.34 mm
TRE	0.18 mm	0.14 mm	0.25 mm

Table 4.1 Point Cloud modeling results evaluation between the simulated US models and CT models on 50 3-D paired using TRE, FRE.
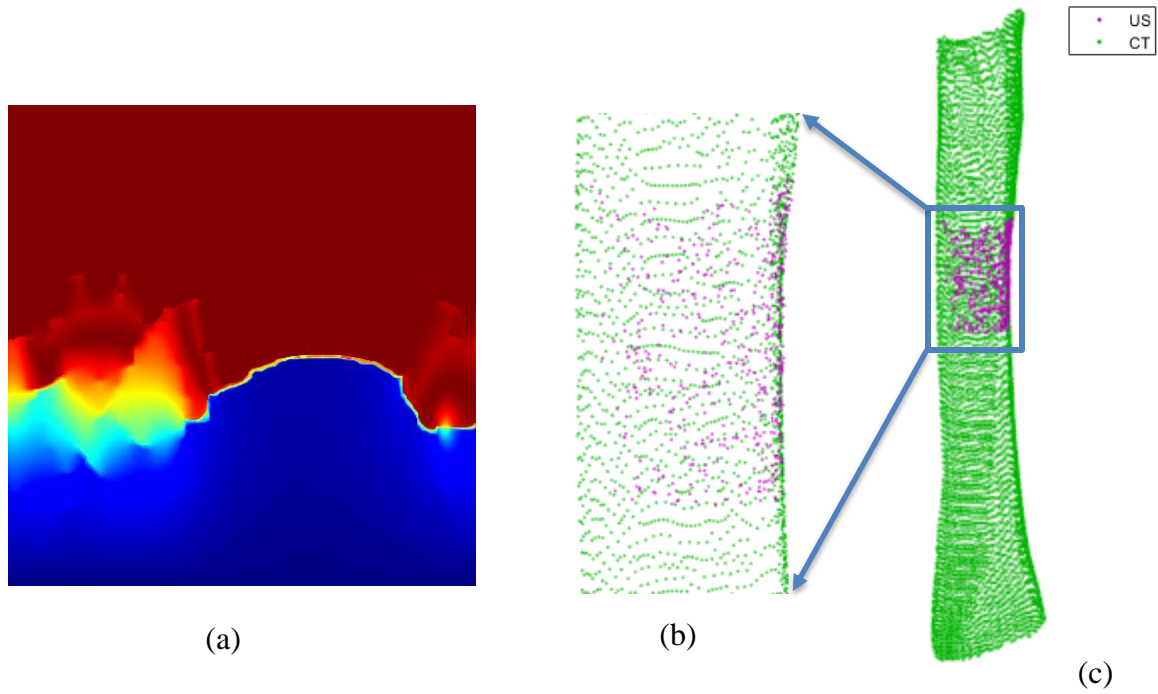
Figure 4.2 (a) A single BSE(x,y) image selected from a set of US images. (b) The zoom view of registerd CT and US point clouds from c. (c) The point cloud modeling result for the US from b and the CT model in a point cloud format. The green was the CT model while the purple was the simulated US model.

To quantify the point cloud modeling results, the root means square error (RMSE), FRE and TRE for the modeling are used because the reconstruction process can be considered as the atypical "alignment" process. As we already know the relative portion of the simulated point clouds inside the CT models, we can measure the difference between these pairs to evaluate the reconstruction results. The mean of RMSE, FRE and TRE are 0.62 mm, 0.21 mm and 0.18 mm on 20 models. And the maximum value of them is 0.78 mm, 0.34 mm, and 0.25 mm respectively. The quantitative results manifest the point cloud models is qualify for the following process.

4.3 <u>US-CT registration evaluation</u>

Our proposed registration method was quantitatively evaluated using TRE and FRE analysis with a comparison with three mainstream point cloud registration methods. **Iterative closet point** (ICP) registration is the most famous point cloud registration method by minimizing the difference between two point clouds, which is widely used in bone registration [1-3]. **Coherent point drift** (CPD) uses a probabilistic approach to optimally align two point set that maximizes its Gaussian mixture model (GMM) [4, 5]. **Normal distribution transformation** (NDT) divides the data into cells and assigns the cells in normal distributions. Then it matches the normal distribution into the cells using Newton's algorithm [6-8].

The experiment for the proposed method was conducted on a workstation with Intel Core i7 9400 @2.90 GHz CPU, 16GB RAM, and Nvidia GeForce RTX 2070 Super GPU in Python 3.7 while it took about 1 week and 2 days to converge for the proposed method. And that three mainstream point cloud registration methods were executed and evaluated in MATLAB 2019b.

4.3.1 <u>Evaluation of the proposed registration method</u>

The qualitative results for the proposed hierarchical registration method are shown in Fig 4.3. The difference between the registered models was nearly zeros after registration which was represented in blue. However, there're a few larger values in the difference value sets due to the difference of the density for the two models and the noises in the simulated bone surface point clouds. But we can still say the alignment results turn out good and can be considered as a potential method for future studies in practically.
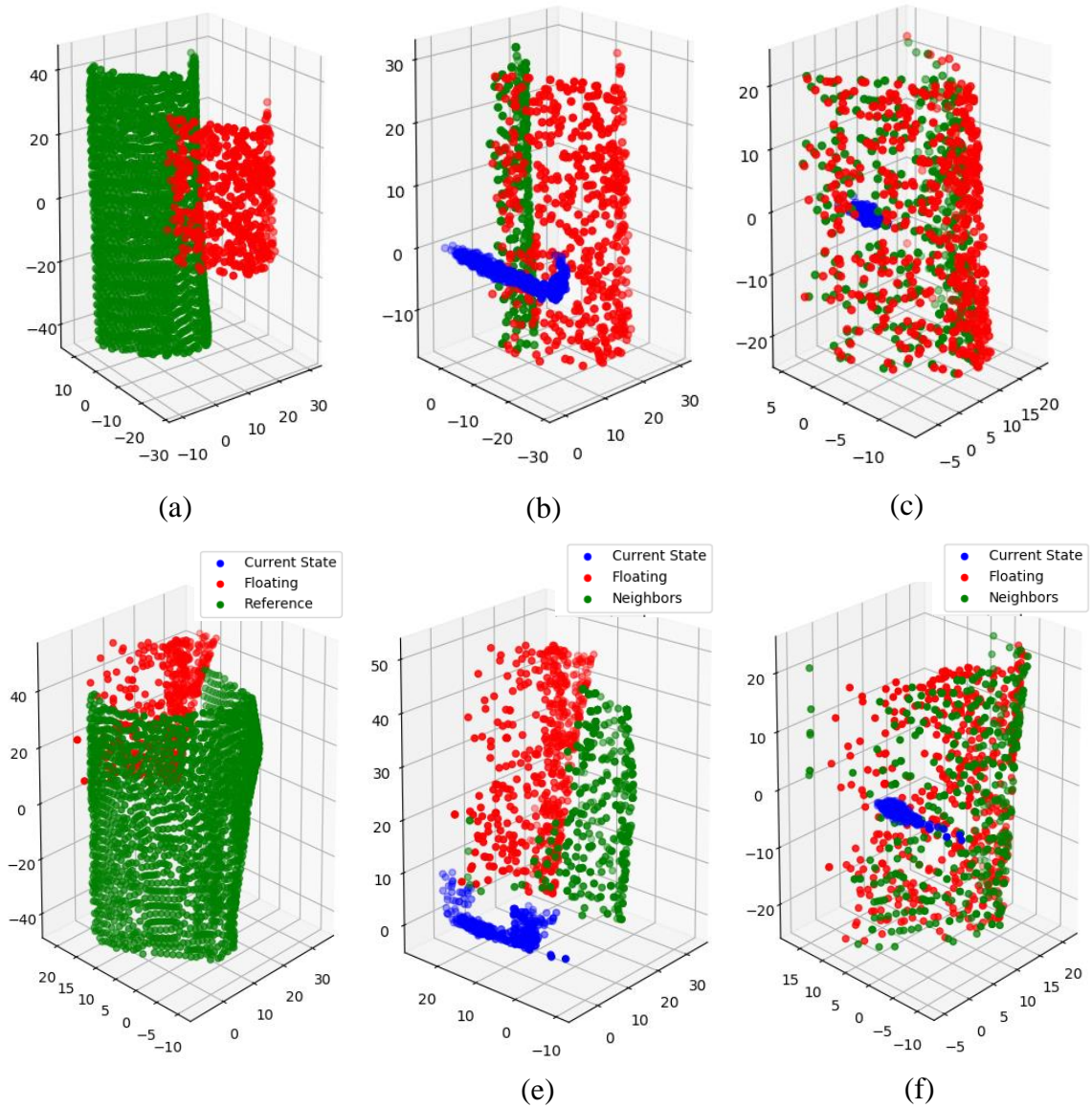
Figure 4.3 Examples of 3-D registration results using the proposed RL registration method. The images from left to right: (1) The pre-register pairs of CT and US point cloud models. The green is the truncated CT model, while red is the US model. (2) The initial inputs for the RL system. The red is the US model, and the green is the neighbor points of that US model from the CT model. The input of the system, which is the difference between that models represents in blue. (3) The registration result for the mis-register pairs.

	TRE (mm)			FRE (mm)			Time (s)			Success	
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	1 step	
Proposed method	3.82	2.16	12.42	4.32	3.53	14.07	8.36	4.75	18.32	0.31	92.7%
ICP	8.42	2.63	17.92	9.19	3.15	18.47	0.31	0.24	0.86	-	35.6%
CPD	9.10	3.57	17.16	10.63	4.74	19.14	0.68	0.51	1.03	-	31.9%
NDT	13.51	6.18	20.18	14.91	8.04	22.01	0.83	0.73	1.26	-	24.3%

Table 2 Registration results evaluation on the proposed method and the other point cloud registration methods. The quantitative analysis was on TRE, FRE, Time, and success rate of each registration method.

4.3.2 <u>Comparison with state-of-art point-based registration methods</u>

The quantitative evaluation of registration accuracy was accessed with TRE and FRE, and the success of registration is defined as the TRE for the registered pairs is not larger than 10 mm. Contrary to the proposed method, ICP, CPD, and NDT easily failed in challenging cases especially for the cases with more significant perturbation in the z-direction, resulting in relatively low success rate. It should be noticed that the ICP registration, which was widely used in the previous studies, performed worse in both accuracy and success rate than the proposed method without initial registration.

The average error of FRE and TRE was 4.78 mm and 3.52 mm for the proposed method, and the success rate was over 92% that is much higher than those three methods. From the summary table, we can see that the success rate of the proposed method outperformed the other three mainstream registration methods. As the direct features like the similarity or normal distribution are quite similar in femur surfaces, the traditional registration methods

may not appropriate for it while the DNN-based method can learn the underlying contextual of the femur bone.

The average run time for each step in the proposed hierarchical registration in the testing cases was 0.31 s, and the total run time was 8.36 s, which cannot meet the real-time requirement. The run time for the RL-based registration method depends on the architecture of the computer, such as CPU and GPU, and the pipeline of the algorithm. And the average run time for the other three registration methods was less than 1 s, which were 0.31 s, 0.68 s, and 0.83 s. The improvement of the proposed method should be solved in the future, and we'll discuss it in the next section.

Reference

[1] Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." In Sensor fusion IV: control paradigms and data structures, vol. 1611, pp. 586-606. International Society for Optics and Photonics, 1992.

[2] Chen, Yang, and Gérard Medioni. "Object modeling by registration of multiple range images." Image and vision computing 10, no. 3 (1992): 145-155.

[3] Zhang, Zhengyou. "Iterative point matching for registration of free-form curves and surfaces." International journal of computer vision 13, no. 2 (1994): 119-152.

[4] Myronenko, Andriy, and Xubo Song. "Point set registration: Coherent point drift." IEEE transactions on pattern analysis and machine intelligence 32, no. 12 (2010): 2262-2275.

[5] Myronenko, Andriy, Xubo Song, and Miguel A. Carreira-Perpinán. "Non-rigid point set registration: Coherent point drift." In Advances in neural information processing systems, pp. 1009-1016. 2007.

[6] Biber, Peter, and Wolfgang Straßer. "The normal distributions transform: A new approach to laser scan matching." In Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453), vol. 3, pp. 2743-2748. IEEE, 2003.

[7] Magnusson, Martin, Achim Lilienthal, and Tom Duckett. "Scan registration for autonomous mining vehicles using 3D‐NDT." Journal of Field Robotics 24, no. 10 (2007): 803-827.

[8] Kohlbrecher, S., Von Stryk, O., Meyer, J. and Klingauf, U., 2011, November. A flexible and scalable slam system with full 3d motion estimation. In 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics (pp. 155-160). IEEE.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 <u>Conclusions</u>

The goal of this work was to investigate the use of reinforcement learning (RL) in point-based ultrasound (US) to computed tomography (CT) registration for the surgical guidance of femoral shaft fracture. For that, we addressed the main challenges in applying the RL-based method on US-CT registration for computer-assisted orthopedic surgery (CAOS) applications.

Firstly, we explore the utilization of US simulator in 3D slicer and PLUS toolkit to generate simulated US images from a synthetic femur phantom. The anatomical surface mesh data obtained from CT can be converted into simulated US volumetric images using a US physics-based model. Therefore, the simulation on any number of moving, intersecting objects secures the unique and sufficient dataset in training an RL model.

Secondly, we investigate how the enhancement of surface regions from ultrasound (US) data using local phase-based bone feature enhancement being applied in bone segmentation. Bone surfaces from US were segmented from the enhancement US images by bottom-up ray casting method. As applied the bone surface segmentation result in point cloud modeling, the anatomy structures were reflected accurately and optimally aligned with the CT models making US a valid modality for surgical guidance in CAOS.

Finally, we explore how to achieve automatic and flexible US-CT registration using Q-learning. We proposed a 3-D rigid registration method using Q-learning takes the point-

cloud models with known transformation matrix as the inputs and outputs a continuous motion of actions to achieve the alignment between the two un-aligned models. With the comparison with three traditional point-based registration method, the registration for Q-learning outperforms in the registration error and the registration success rate significantly. Through the results for different lengths of volumetric US images, the proposed method demonstrated its robustness and flexibility in dealing with different circumstances successfully. Accordingly, we believe the proposed registration method could be used practically for medical image registration in CAOS.

5.2 <u>Limitations</u>

The US dataset of this work comes from the US simulator in 3D slicer, while the in-vivo US images have more speckle noises and inconsistent anatomy structure for the US images. Therefore, the in-vivo experiments are needed for further studies to evaluate the robustness of the bone surface segmentation and modeling as well as the 3-D rigid-body registration.

Also, the range for the registration method is limited as the full cover of parameter space requires dramatic high expense in computation. For the full parameter space with ±30 perturbation in each parameter, more than 60^{12} cases should be generated for the agent which increases exponentially with the number of perturbations. Furthermore, the real-time of the proposed method could not be guaranteed under every circumstance as the total steps for each registration process vary greatly. Both of these limitations are actually the limitation on the architecture of the proposed method which indicates the optimization of the architecture should be investigated in the future.

Last but not least, while there is no theoretical guarantee that the agent could finally achieve correct registration with optimal motions, in practice the agent always converges

to one position that could not stop early when the correct registration position is achieved but never produce cyclical movements. This is presumably due to the fact that our supervised registration path is approximately a straight line in the registration parameter space.

5.3 <u>Future Works</u>

Overall, we believe our work for point cloud modeling on bone surfaces and the US-CT registration has shown that US can be applied in image guidance procedure for computer-assisted orthopedic surgery (CAOS) practically as it has adequate accuracy and time-efficiency. However, the presented architecture of the proposed method can still be improved and extended to various studies as described below:

5.3.1   <u>Improvement on Point cloud modeling for bone surfaces</u>

The proposed method in point cloud modeling has shown its time-efficiency to reconstruct the 3-D model from the bone surfaces in US images. However, the current bone surface sampling method for modeling is a simple random selection method. For further studies, we can use a more theoretical simulation method to provide more accurate simulation results for further investigation, such as particle simulation. As particle simulation can quickly simplify the bone surfaces with salient anatomic features.

In addition, as the different densites of US point clouds may lead to different results as the accuracy in finding neighbor points in CT. More investigation should be performed in computing the relationship between the number of US and CT points with the registration error.

5.3.2 <u>Improvement of the performance for the proposed method</u>

In the proposed work, the performance of the registration method depends on the architecture of the proposed RL-based registration method. In order to improve the workable range of the registration method and the time-consuming problems, the architecture can be improved by optimizing the PointNet++ based neural network that reduces the computation expense. Furthermore, training two different neural networks for initial alignment and fine alignment respectively, which can obviously reduce the computation expense in each network, can also be explored in improving the performance of the registration method.

5.3.3 <u>Assessment of US-CT registration framework in vivo</u>

The proposed method can be extended in-vivo as all the work in this thesis is based on the simulation phantom in the 3D slicer. However, the effects of the speckle noise in the in-vivo US for the proposed method are still unknown. Thus, the robustness and accuracy of the proposed registration method should be evaluated in-vivo in order to achieve practically in the future.

5.3.4 <u>Comparison of registration results with other Deep learning-based registration methods</u>

The proposed RL-based registration is evaluated and compared with three different traditional point-based registration methods. Nevertheless, the proper performance comparisons between the proposed method and the other DL-based registration methods are still under investigation. The comparison results, which include time-complexity,

accuracy and the workable range, can indicate the accuracy and reliability of the proposed

registration method