

**FLUOROSCOPIC NAVIGATION FOR ROBOT-ASSISTED  
ORTHOPEDIC SURGERY**

by

Cong Gao

A dissertation submitted to The Johns Hopkins University in conformity with  
the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

August, 2022

© 2022 Cong Gao

All rights reserved

# Abstract

Robot-assisted orthopedic surgery has gained increasing attention due to its improved accuracy and stability in minimally-invasive interventions compared to a surgeon's manual operation. An effective navigation system is critical, which estimates the intra-operative tool-to-tissue pose relationship to guide the robotic surgical device. However, most existing navigation systems use fiducial markers, such as bone pin markers, to close the calibration loop, which requires a clear line of sight and is not ideal for patients. This dissertation presents fiducial-free, fluoroscopic image-based navigation pipelines for three robot-assisted orthopedic applications: femoroplasty, core decompression of the hip, and transforaminal lumbar epidural injections. We propose custom-designed image intensity-based 2D/3D registration algorithms for pose estimation of bone anatomies, including femur and spine, and pose estimation of a rigid surgical tool and a flexible continuum manipulator. We performed system calibration and integration into a surgical robotic platform. We validated the navigation system's performance in comprehensive simulation and *ex vivo* cadaveric

## ABSTRACT

experiments. Our results suggest the feasibility of applying our proposed navigation methods for robot-assisted orthopedic applications.

We also investigated machine learning approaches that can benefit the medical imaging analysis, automate the navigation component or address the registration challenges. We present a synthetic X-ray data generation pipeline called SyntheX, which enables large-scale machine learning model training. SyntheX was used to train feature detection tasks of the pelvis anatomy and the continuum manipulator, which were used to initialize the registration pipelines. Last but not least, we propose a projective spatial transformer module that learns a convex shape similarity function and extends the registration capture range.

We believe that our image-based navigation solutions can benefit and inspire related orthopedic robot-assisted system designs and eventually be used in the operating rooms to improve patient outcomes.

**Primary Reader and Advisor:** Mehran Armand

**Secondary Readers:** Mathias Unberath, Russell H. Taylor, and Amit Jain

# Acknowledgments

My five years at Hopkins to pursue my Ph.D. is an exceptionally unique, rewarding, and memorable experience in my life. I am incredibly grateful to the people who have carried, supported, and inspired me over the past five years.

First and foremost, I am deeply indebted to my advisor, Dr. Mehran Armand. I would not have had all these experiences without Dr. Armand having me on board his team when I was a visiting undergraduate student in 2016. Mehran is always kind and open to my research interests and offers me the flexibility to try and explore. I learned from him not only the technical skills but also a much bigger picture of the fundamental clinical motivations. I truly enjoyed working with Mehran to solve real clinical problems and investigate the advanced robotic system.

I also would like to express my greatest appreciation to my co-advisor, Dr. Mathias Unberath. I was very lucky to work with Mathias when he just started at Hopkins as a postdoctoral researcher. When I was going through a rough

## ACKNOWLEDGMENTS

patch in the early stage, he supported and encouraged me with his extraordinary research skills and pushed me up to speed. He always has creative research ideas and holds strong passion to make breakthroughs. The work of Chapter 3 and 5 would not have happened without Mathias' vision and support.

I could not have undertaken this journey without the advise and encouragement from Dr. Russell Taylor. A great amount of my knowledge growth was obtained from attending the weekly Friday meetings held by Dr. Taylor. He is always sharp on challenging technical problems and comes up with bright ideas to resolve them. I also gained tremendous inspiration from his in-depth mindset and diligent research habits. I sincerely thank Dr. Taylor's leadership on LCSR and management in such a fantastic research environment.

I have also had the pleasure of working with Dr. Amit Jain. Although there was not much in person interaction with Dr. Jain due to the remote collaboration nature during the COVID pandemic, I still gained extraordinary advise and feedback from him as an expert clinician. I am grateful for his valuable inputs on building my engineering solutions.

I am very grateful to Dr. Austin Reiter, who admitted me to the Ph.D. program and served as my primary advisor during my first year. Although we could not work longer due to his steered career path, I do appreciate his advise on my fresh deep learning project and all his encouragements.

## ACKNOWLEDGMENTS

The work in this dissertation would not have been possible without the collaborations of expert clinicians and medical school residents at Johns Hopkins. I gratefully thank Dr. Adam Margalit for his leadership in the transforaminal lumbar epidural injection experiments. Special thanks to Dr. David Cohen and Dr. Akhil Chhatre for performing manual cadaveric epidural lumbar injections. I am also thankful to Dr. Julius Oni and Dr. Carlos Ortiz Babilonia for their valuable clinical advice. I express my grateful thanks to Mr. Demetries Boston for his assistance in the numerous cadaver experiments with his expertise.

I would like to extend my sincere thanks to my lab members belonging to BIGSS, ARCADE and LCSR, fellow graduate students, visiting students and friends during my stay here: Adam Margalit, Alejandro Martin-Gomez, Ali Ebrahimi, Amir Farvardin, Anqi Feng, Ayushi Sinha, Baichuan Jiang, Benjamin Killeen, Catalina Gomez, Changyan He, Daniil Pakhomov, Farshid Alameigi, Hao Ding, Haomin Chen, Henry Phalen, Jasmine Cho, Jan-Nico Zaech, Javad Fotouhi, Jie Ying Wu, Joshua Liu, Justin Ma, Lydia Al-Zogbi, Liuhong Ma, Rachel Hegeman, Robert Grupp, Mahsan Bakhtiarinejad, Max Zhaoshuo Li, Manish Sahu, Mareike Thies, Molly O' Brien, Paul Wilkening, Ping-Cheng Ku, Shahriar Sefati, Xingtong Liu, Wei-Lun Huang, Weiyao Wang, Wenhao Gu, Yicheng Hu, and Yihao Liu.

I am also thankful to the administrative staff members of CS Department and LCSR, who provided prompt and unlimited support for managing the fan-

## ACKNOWLEDGMENTS

tastic environment: Ashley Moriarty, Alison Morrow, Lorrie Dodd, Jordan Card, Kim Franklin, Meagan Wade, Zack Burwell, Laura Graham, and Tracy Marshall.

I am fortunate to be supported by the U.S. National Institutes of Health under Grants R21EB028505, R01EB016703, R01EB023939, and Johns Hopkins internal funds. I also thank the computational resources at the Maryland Advanced Research Computing Center (MARCC).

Last but not least, I express my deepest gratitude to my family. My grandfather brought me up when I was a kid. He takes deep care of our family with love and sets a role model for all of us. My parents, Fubin Gao and Yue Gao, raised me up with encouragements and pushed me hard to pursue my dreams. My uncle and aunt always cheer for my achievements and share with their life experiences once I meet difficulties. I am extremely fortunate to get in love with my girl friend, Daisy, who has always supported me through the hard times. I am very grateful to have you all in my life.

# **Dedication**

*This thesis is dedicated to my loving parents, Fubin Gao and Yue Gao.*

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgments</b>	<b>iv</b>
<b>List of Tables</b>	<b>xvii</b>
<b>List of Figures</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Robot-Assisted Orthopedic Surgery . . . . .	2
1.2 Fluoroscopic Navigation . . . . .	3
1.3 2D/3D Registration Fundamentals . . . . .	8
1.3.1 Problem Formulation . . . . .	9
1.3.2 Challenges . . . . .	10
1.4 Thesis Overview . . . . .	12
1.4.1 Outline . . . . .	12
1.4.2 Contributions . . . . .	13

## CONTENTS

1.4.3 Publications . . . . .	14
<b>2 Orthopedic Application Background</b>	<b>17</b>
2.1 Femoroplasty . . . . .	18
2.2 Core Decompression of the Hip . . . . .	19
2.3 Transforaminal Lumbar Epidural Injection . . . . .	21
<b>3 SyntheX: Realistic Synthesis for X-ray Image Analysis</b>	<b>24</b>
3.1 Introduction . . . . .	24
3.2 Contributions . . . . .	26
3.3 Background . . . . .	27
3.4 Down-Stream Clinical Tasks . . . . .	30
3.4.1 Hip Imaging . . . . .	32
3.4.2 Surgical Robotic Tool Detection . . . . .	33
3.5 Precisely Controlled Investigations on SyntheX’ Sim2Real Performance Using the Hip Imaging Usecase . . . . .	34
3.5.1 Precisely Matched Hip Dataset . . . . .	36
3.5.2 Domain Randomization and Adaptation . . . . .	37
3.6 Dataset and Experimental Details . . . . .	38
3.6.1 Hip Imaging . . . . .	38
3.6.2 Robotic Surgical Tool Detection . . . . .	39
3.6.3 Benchmark Hip Imaging Investigation . . . . .	40

## CONTENTS

3.7 Model and Evaluation Paradigm . . . . .	41
3.8 Experiments and Results . . . . .	45
3.8.1 Hip Imaging . . . . .	45
3.8.2 Surgical Robotic Tool Detection . . . . .	46
3.8.3 Sim2Real Benchmark Findings . . . . .	47
3.8.3.1 The Effect of Domain Randomization . . . . .	48
3.8.3.2 The Effect of Domain Adaptation . . . . .	49
3.8.3.3 Scaling Up the Training Data . . . . .	50
3.9 Discussion . . . . .	54
3.10 Conclusion . . . . .	59
3.11 Acknowledgement . . . . .	60
<b>4 Pose Estimation of Bone Anatomy</b>	<b>61</b>
4.1 Introduction . . . . .	61
4.2 Contributions . . . . .	63
4.3 Related Work . . . . .	64
4.4 Methods . . . . .	67
4.4.1 Data Acquisition and Preprocessing . . . . .	67
4.4.2 Femur Registration . . . . .	68
4.4.2.1 Automatic Pelvis Initialization . . . . .	70
4.4.2.2 Intensity-based Pelvis Registration . . . . .	71
4.4.2.3 Multi-view Femur Registration . . . . .	73

## CONTENTS

4.4.3	Spine Vertebrae Registration . . . . .	74
4.4.3.1	Single-view Spine Registration . . . . .	76
4.4.3.2	Vertebra-by-Vertebra Registration . . . . .	77
4.5	Experiments and Results . . . . .	78
4.5.1	Femur Registration . . . . .	78
4.5.1.1	Simulation Setup and Evaluation Metrics . . . .	78
4.5.1.2	Simulation Results . . . . .	80
4.5.1.3	Cadaver Experiments and Results . . . . .	81
4.5.2	Spine Vertebrae Registration . . . . .	84
4.5.2.1	Simulation Study Setup . . . . .	84
4.5.2.2	Results . . . . .	85
4.6	Discussion . . . . .	86
4.7	Conclusion . . . . .	91
4.8	Acknowledgement . . . . .	91
<b>5</b>	<b>Projective Spatial Transformers for 2D/3D Registration</b>	<b>92</b>
5.1	Introduction . . . . .	92
5.2	Contributions . . . . .	93
5.3	Related Work . . . . .	94
5.4	Methods . . . . .	98
5.4.1	Projective Spatial Transformer (ProST) . . . . .	98
5.4.1.1	Canonical Projection Geometry . . . . .	99

## CONTENTS

5.4.1.2 Grid Sampling Transformer . . . . .	100
5.4.2 Approximating Convex Image Similarity . . . . .	101
5.4.3 Double Backward Training Loss . . . . .	103
5.4.4 Domain Randomization . . . . .	105
5.4.5 Application to 2D/3D Registration . . . . .	106
5.5 Experiments . . . . .	107
5.5.1 Simulation Environment and Dataset . . . . .	108
5.5.2 Ablation Study Design . . . . .	109
5.5.2.1 Architecture . . . . .	109
5.5.2.2 Image with Overlays . . . . .	111
5.5.2.3 Anatomy . . . . .	112
5.5.3 Pose Sampling Strategies and Training Parameters . . . .	112
5.5.4 Registration Testing and Evaluation Metrics . . . . .	114
5.6 Results . . . . .	116
5.7 Discussion . . . . .	123
5.8 Conclusion . . . . .	129
5.9 Acknowledgement . . . . .	129
<b>6 Fluoroscopic Navigation of Rigid and Flexible Surgical Tools and Manipulators</b>	<b>131</b>
6.1 Introduction . . . . .	131
6.2 Contributions . . . . .	133

## CONTENTS

6.3 Rigid Drilling/Injection Device . . . . .	134
6.3.1 Injection Device Registration Methods . . . . .	137
6.3.1.1 Joint Injection Device Registration . . . . .	138
6.3.1.2 Multi-view Injection Device Registration . . . . .	139
6.3.2 Experiments and Results . . . . .	140
6.3.2.1 Image-based Registration Simulation Study . . .	140
6.3.3 Discussion . . . . .	143
6.4 Flexible Continuum Manipulator . . . . .	146
6.4.1 Continuum Manipulator Model . . . . .	148
6.4.2 Navigation Methods . . . . .	149
6.4.2.1 CM Detection and Pose Initialization . . . . .	150
6.4.2.2 Image-based 2D/3D Registration . . . . .	152
6.4.3 Experiments and Results . . . . .	157
6.4.3.1 Image-based Registration Simulation Study . . .	158
6.4.4 Discussion . . . . .	164
6.4.5 Conclusion . . . . .	167
6.5 Acknowledgement . . . . .	167
<b>7 System-Level Validation and Clinical Applications</b>	<b>168</b>
7.1 Introduction . . . . .	168
7.2 Contributions . . . . .	169
7.3 Femoroplasty . . . . .	170

## CONTENTS

7.3.1	Injection Device Calibration . . . . .	173
7.3.1.1	Hand-eye Calibration of the Device Frame . . . . .	173
7.3.1.2	Hand-eye Calibration of the Injection Device . . . . .	174
7.3.2	Cadaveric Experiments and Results . . . . .	176
7.3.3	Discussion . . . . .	179
7.4	Core Decompression of the Hip . . . . .	181
7.4.1	CM Hand-eye Calibration . . . . .	182
7.4.2	Cadaveric Experiments and Results . . . . .	186
7.4.3	Discussion . . . . .	190
7.5	Transforaminal Lumbar Epidural Injections . . . . .	192
7.5.1	Pre-operative Injection Planning . . . . .	194
7.5.2	Post-op Evaluation . . . . .	196
7.5.3	System Setup and Calibration Method . . . . .	197
7.5.4	System Calibration Results . . . . .	198
7.5.5	Cadaveric Experiments and Results . . . . .	201
7.5.6	Discussion . . . . .	203
7.6	Conclusion . . . . .	208
7.7	Acknowledgement . . . . .	208
<b>8</b>	<b>Conclusion and Future Work</b>	<b>209</b>
8.1	Conclusion . . . . .	209
8.2	Future Work . . . . .	211

## CONTENTS

<b>A Supplementary Material for Chapter 3</b>	<b>214</b>
A.0.1 Domain Randomization . . . . .	214
A.0.2 Domain Adaptation . . . . .	216
A.0.2.1 CycleGAN . . . . .	217
A.0.2.2 Adversarial Discriminative Domain Adaptation .	219
A.0.3 Network Training Details . . . . .	221
<b>B Supplementary Materials for Chapter 6</b>	<b>222</b>
B.1 Centerline-based 2D/3D Registration . . . . .	223
B.1.1 Centerline Extraction . . . . .	223
B.1.2 Centerline-based 2D/3D Registration . . . . .	224
<b>Bibliography</b>	<b>227</b>
<b>Vita</b>	<b>259</b>

# List of Tables

3.1	Average performance metrics for hip imaging as a mean of 5-fold individual testing on 366 real hip X-ray images. . . . .	46
3.2	Average performance metrics for surgical tool detection as a mean of 5-fold individual testing on 264 real X-ray images of the continuum manipulator. . . . .	47
3.3	Segmentation Dice Score. The Dice score ranges from 0 to 1, with larger values corresponding to better segmentation performance.	52
3.4	Landmark Detection Errors (mm) at 90% Activation Percentage. Lower values are better. . . . .	53
4.1	Simulation Results of Pelvis and Femur Registration Errors . . .	81
4.2	Cadaver Study Results of Femur Pose Error Metrics . . . . .	83
4.3	Mean Registration Error in Spine Vertebrae Registration Simulation Study . . . . .	88
5.1	Registration Pose Error of Pelvis . . . . .	119
5.2	Target Registration Error (TRE) and Success Rate of Pelvis . . .	121
5.3	Target Registration Error (TRE) and Success Rate of Spine . . .	121
5.4	Target Registration Error (TRE) and Success Rate of Pelvis . . .	127
6.1	Mean Registration Error in Injection Device Registration Simulation Study. . . . .	143
6.2	Simulation Results of Registration Errors . . . . .	160
7.1	Cadaver Specimen1 Results of Error Metrics . . . . .	186
7.2	Cadaveric Needle Injection Accuracy . . . . .	203

# List of Figures

1.1	Concept of a typical robot-assisted orthopedic surgical system. . . . .	2
1.2	Fluoroscopic navigation system workflow. . . . .	5
1.3	Surgeons performing total knee replacement using Stryker’s Mako robotic system. . . . .	7
2.1	(a) Illustration of femur fractures. (b) An example femoroplasty bone augmentation scheme [1]. . . . .	19
2.2	(a) Illustration of femur head osteonecrosis. (b) An example hip X-ray of patient with femoral head osteonecrosis. . . . .	20
2.3	(a) Illustration of spine injection. (b) An X-ray image taken after lumbar epidural injection with needle tip remaining. This image was reported in [2]. . . . .	22
3.1	Overall concept of SyntheX. Blue: Conventional approach for learning-based tasks on medical imaging. Curating a relevant database of real X-ray samples requires real data acquisition and costly annotation from domain experts. Orange: SyntheX enables simplified and scaled-up data curation because data generation is synthetic and synthesized data can be annotated automatically through propagation from the 3D model. SyntheX results in deep learning image analysis models that perform comparably to or better than real-data trained models. . . . .	30

## LIST OF FIGURES

3.2 (a) Hip imaging. The hip anatomical structures include left and right hemipelvis, lumbar vertebrae, upper sacrum, and left and right femurs. The anatomical landmarks consist of left and right anterior superior iliac spine (ASIS), center of femoral head (FH), superior pubic symphysis (SPS), inferior pubic symphysis (IPS), medial obturator foramen (MOF), inferior obturator foramen (IOF), and the greater sciatic notch (GSN). These landmarks are useful in identifying the anterior pelvic plane (APP) and initializing the 2D/3D registration of both pelvis and femur [3, 4]. (b) Surgical robotic tool detection. An illustration of the robotic surgical system including a continuum manipulator is shown on the top. An example real X-ray image and the corresponding segmentation and landmarks of the continuum manipulator is shown on the bottom. . . . .	31
3.3 (a) Generation of precisely matched synthetic and real X-ray database: Real X-rays and CT scans are acquired from cadaveric specimens and registered to obtain the relative camera poses. Using these poses, synthetic X-rays can be generated from the CTs that precisely match the real X-ray data in all aspects but appearance. (b) Changes in (synthetic) X-ray appearance based on simulation paradigm. . . . .	35
3.4 U-net based concurrent segmentation and landmark detection network architecture for multi-task learning. . . . .	42

## LIST OF FIGURES

3.5 Plots of average landmark detection error curves with respect to activated landmark percentage. <i>Real2Real</i> performance on the controlled dataset is shown in gold. An ideal curve should approach the bottom right corner: all landmarks detected with perfect localization. Each subplot compares the baseline <i>Real2Real</i> performance curve to varied <i>Sim2Real</i> methods that are evaluated on the same real data test set. The <i>Sim2Real</i> technique of the specific method is identified in the upper left corner of each subplot. Domain names followed by “CycleGAN” mean the training data are generated using CycleGAN trained between the specific image domain and the real image domain. “reg DR” and “str DR” refer to regular domain randomization and strong domain randomization, respectively. “ADDA” refers to adversarial discriminative domain adaptation. (a)-(c) present performance comparison of methods trained on precisely matched datasets. (d)-(f)(i) further evaluates the added effect of using domain adaptation techniques again using precisely matched datasets. (g),(h) demonstrate improvements in <i>Sim2Real</i> performance on the same real data test set when a larger, scaled-up synthetic training set is used. All the results correspond to an input image size of $360 \times 360$ px. . . . .	44
4.1 General workflow for pose estimation of bone anatomy. . . . .	67
4.2 Femur registration pipeline. Multiple view fluoroscopy projection geometries and simulation images with DRR-derived edges in green are presented. . . . .	70
4.3 Workflow of pelvis and femur registration. . . . .	74
4.4 Spine registration pipeline: Multi-view fluoroscopy projection geometry of a robotic injection device and spine vertebrae is presented. . . . .	75
4.5 Workflow of spine registration. . . . .	78
4.6 Normalized 2D histograms of (a) pelvis pose ( $\delta T_{pel}$ ), (b) femur pose ( $\delta T_{fem}$ ), (c) C-arm view2 ( $\delta T_{C_1}^{C_2}$ ), (d) C-arm view3 ( $\delta T_{C_1}^{C_3}$ ) error for the femur registration simulation studies. . . . .	82
4.7 Left: An example intraoperative fluoroscopic image is presented. The injected BBs are zoomed in. One example BB location is marked with a dotted circle. Right: Illustration of femur head injection path. The injection entry point ( $x_{EP}$ ) and femur head center point ( $x_{FH}$ ) are marked with red dots. The path is shown using a blue vector ( $p^{fem}$ ). . . . .	82

## LIST OF FIGURES

4.8	(a) Rendered vertebrae segmentation from pre-operative CT scans. (b) An example of a randomly simulated spine shape. (c) An example DRR image of the spine vertebrae. (d) An example simulation an X-ray image. . . . .	86
4.9	Normalized 2D histograms of spine registration pose error ( $\delta T_{Carm}^{ID}$ , $\delta T_{Carm}^v$ ) reported in joint magnitudes of translation and rotation. . . . .	87
4.10	Left: Scatter plot of correlation matrix between femoral head center translation error and pelvis translation error reported in the femoral head center. Right: Scatter plot of correlation matrix between femur entry point error and pelvis translation error reported in the femoral head center. Correlation coefficients are marked on the right bottom of each plot. . . . .	89
5.1	(a) Canonical projection geometry and a slice of cone-beam grid points are presented with key annotations. The blue fan covers the control points which are used for further reshaping. (b) Illustration of grid sampling transformer and projection. (c) Scheme of applying ProST to 2D/3D registration. . . . .	98
5.2	ProST DeepNet Architecture for 2D/3D registration. A pelvis segmentation $V_{seg}$ is illustrated as 3D input. The rigid pose parameter $\theta_m$ is shown using RGB cross arrows. An example target X-ray image $I_t$ and moving image $I_m$ generated from ProST are presented. Forward pass follows the black arrows. Backward pass follows orange arrows, where the gradient $\frac{\partial S_{net}}{\partial \theta_m}$ is computed. Detailed structures of CrossViT and 3D CNN are illustrated in blocks on the bottom. . . . .	102
5.3	Illustration of iterative 2D/3D registration using examples of (a) a real pelvis X-ray image with tool overlay and (b) a simulated spine X-ray image. Renderings of the initialization, ProST estimation, and ground truth pose of the anatomies are illustrated on the left. The black arrow shows the difference between initialization and ProST estimated poses. In each example, the target image, initial projection image, and the failed CMAES registration image from the initial are shown in the first row. The ProST registration image, successful CMAES registration image from ProST estimation, and an overlay image with DRR-derived edge in green are shown in the second row. . . . .	117
5.4	Histogram of registration pose errors in translation and rotation for pelvis simulation and standard AP view real X-ray study, respectively, from ProST baseline model registrations. . . . .	120

## LIST OF FIGURES

5.5	Upper: Single Degrees-of-Freedom similarity loss shape comparison between network similarity and Grad-NCC. An example of projection geometry with axis directions is illustrated on the left. Bottom: Box plot of registration pose errors of baseline model in simulation decoupled to the individual degree of freedom. . . . .	122
5.6	Correlation of similarity and pose Riemannian distance. . . . .	124
5.7	Examples of challenging view pelvis X-rays. . . . .	128
6.1	(a) Models of the integrated drilling and injection component. (b) Cement injection with an inserted syringe. (c) An example picture of the actual integrated robotic system for cement injection. (d) Picture of the metallic BBs glued on the surface of the injection device. Three example BBs are marked with red circles. (e) Illustrations of transformations used for injection device registration. Key frames are marked with red arrows. . . . .	135
6.2	Injection device registration pipeline: Multiple view fluoroscopy projection geometries and real injection device X-ray images with DRR-derived edges in green are presented. The rotation center frame is set at the center of the guide tube, which is illustrated using red cross arrows at the bottom left corner. . . . .	138
6.3	Workflow for injection device registration. . . . .	141
6.4	Normalized 2D histograms of injection device registration pose error ( $\delta T_{Carm}^{ID}$ ) reported in joint magnitudes of translation and rotation. . . . .	144
6.5	(a) Figure of the continuum manipulator with flexible cutting/debriding tool [5]. The actuation cables and the FBG sensor are demonstrated. (b)-(e) CM Model. (b) Basic model components including base, notch and top segments. (c) An example CM configuration. The notch joint angle is illustrated as $\tau_i$ . The cubic spline control points are shown aside as $c_1, c_2, \dots, c_5$ . The CM base reference frame is shown in RGB cross arrows. (d) An example flexible tool model. (e) Integrated CM with flexible tool inside. . . . .	148
6.6	Illustration of the proposed image-based navigation for the CM. Key frames are shown with red cross arrows. (a) Pre-operative hand-eye calibration of the CM. The hand-eye matrix $T_{CM}^{ee}$ is represented by the curved arrow. (b) CM feature detection and centerline-based initialization. 2D landmark heatmaps and corresponding 3D landmarks are shown in red and green. 2D and 3D centerlines are shown in yellow and pink, respectively. The CM initial pose estimation ( $T_{CM}^{Carm}$ ) <sub>ini</sub> is marked with a black arrow. (c) Scheme of intensity-based 2D/3D registration of the CM. The registration transformations are described in Section II-C in detail. . .	150

## LIST OF FIGURES

6.7 (a) 3D segmentation label and landmark positions on an example CM model. (b) An example simulation image using DeepDRR. (c) Overlay of the 2D segmentation mask and landmark heatmaps on the cropped image in (b). (d) Extraction of the 2D centerline from a segmentation mask consists of dilation, edge, and distance map computation, and finally, shortest path extraction. (e) A simplified illustration of the centerline-based 2D/3D registration geometry. . . . .	151
6.8 (a): Illustration of multi-view registration. Three example C-arm view source frames are noted with red cross arrows. (b): Illustration of multiple CM registration. The transformations are marked separately on the right and described in Section II-C. . . . .	153
6.9 Workflow of continuum manipulator registration. . . . .	157
6.10 CM registration simulation results. Top: Histogram plot of CM tip error (in orange) and CM tip to femur entry point relative error (in blue) in the AP view C-arm frame. Bottom: Box plot of CM base registration error in translation and rotation of all axes. . . . .	161
6.11 Scatter plots of the correlation matrix between CM tip position error and CM base TransX error of three registration methods. Correlation coefficients are marked on the upper left of each plot. . . . .	165
7.1 Fluoroscopic navigation pipeline for robot-assisted femoroplasty. . . . .	171
7.2 Workflow for femoroplasty. . . . .	172
7.3 Hand-eye calibration scheme. Coordinate frames are marked as red cross arrows. Key transformations are shown in black arrows. The hand-eye calibration transformations are highlighted. The 3D injection device model for 2D/3D registration is illustrated on top. An example X-ray image used for hand-eye calibration is shown on the right. Example BBs are marked in a red circle. . . . .	174
7.4 (a) Femoroplasty cadaver experiment setup. The C-arm, optical tracker and lower torso specimen are marked in the image. (b) Picture of guided drilling using robotic injection device after fluoroscopic navigation. . . . .	177
7.5 (a) Measured and planned trajectory on the cadaver femur model. (b) X-ray image of the guided drilling using an injection device in the cadaveric experiment. (c) Illustrations of registration convergence in the cadaveric study. Multi-view cadaver X-ray images are placed as background and DRR-derived edges are overlaid in green. The top row shows injection device registration, and the bottom row shows the pelvis and femur registration. . . . .	178

## LIST OF FIGURES

7.6 (a) Illustration of the hand-eye calibration transformations. The image-based hand-eye calibration loop is shown in black. The image-based hand-eye transformation is $T_{CM}^{ee}$ . The optical tracker-based hand-eye calibration loop is shown in green, and $T_{O_s}^{ee}$ is the optical tracker-based hand-eye matrix. (b) Hand-eye calibration data collection setup with UR-10, C-arm Polaris tracker, and markers. . . . .	184
7.7 Workflow of continuum manipulator cadaveric experiment for Specimen1. . . . .	187
7.8 Workflow of continuum manipulator cadaveric experiment for Specimen2. . . . .	187
7.9 Upper Left: Specimen1 experiment setup with C-arm, UR-10, Polaris tracker, and CM. Upper Right: Specimen2 experiment setup with UR10 and CM inserted into the femur head. Middle: Examples of BB locations. Bottom: Examples of Polaris fiducial detection results. . . . .	188
7.10 2D overlay examples of fluoroscopic images (background) and DRR-derived edges in green when registration is converged. Top: Cadaveric Specimen1 CM registration results and position overlay after reposition (in red). The original full-size image is placed in the background and the cropped CM registration image is placed in the foreground lower-left corner. Bottom: Cadaveric Specimen2 CM registration results of each C-arm view for the straight and curved pathway insertions. . . . .	188
7.11 Overall pipeline of our robotic needle injection system. Inputs include (a) patient-specific CT scan and spine vertebrae segmentation, and (b) an injection device model. The planning module shows (c) the surgeon's interface to annotate needle injection trajectories and an example display of the planned trajectories on the CT segmentation. Multi-view registration presents: (d) multi-view C-arm X-ray projection geometries. The source-to-detector center projection line is rendered in green and the detector planes are rendered as squares. The needle injector guide and the spine anatomy are rendered using the registration pose. (e)(f)(g) Registration overlay images of the needle injector guide. The outlines of the reprojected injection device are overlaid in green. (h)(i)(j) Registration overlay images of the cadaveric spine vertebrae. We present an actual cadaveric needle injection image in (k). . . . .	194
7.12 Workflow for transforaminal lumbar epidural injections. . . . .	195

## LIST OF FIGURES

7.13 (a) Picture of robotic injection system setup including C-arm, UR-10, optical tracker, injection device, and a cadaveric specimen. Bottom right: Picture of the syringe mount. (b) An example X-ray image used for hand-eye calibration. Example BBs are marked in a red circle. (c) An example X-ray image used for needle calibration. The needle tip and base points are marked with red circles. . . . .	199
7.14 (a) Screenshot of planning trajectories. (b) An example X-ray image was taken after the robotic needle injections. (c) Rendering of the Post-operative CT scans. (d) Illustration of the manually labeled safety zones. . . . .	201
7.15 Scatter violin plot of cadaver needle injection accuracy comparison between robot and surgeon. . . . .	204
B.1 Network architecture used for concurrent segmentation and landmark detection. . . . .	223
B.2 Centerline-based 2D/3D registration. (a): Extraction of the 2D centerline from a segmentation mask consists of dilation, edge and distance map computation, and finally, shortest path extraction. (b): Geometric relations of curve-based 2D/3D registration. (c): Coordinate transformation from world frame $F_M$ to camera frame $F_C$ . 3D positions of $A^C$ and $B^C$ in $F_C$ are defined as $\vec{p}_A = (x_A^C, y_A^C, d_A)$ and $\vec{p}_B = (x_B^C, y_B^C, d_B)$ . $\vec{p}_{AB} = \vec{p}_B - \vec{p}_A$ . In $F_M$ , we define vector $\vec{p}_2$ , which is parallel to $\vec{p}_{AB}$ and has the same length, starting at the origin, which defines as $A^W$ (the first landmark), ending at $B_2^W$ (the second landmark). . . . .	224

# **Chapter 1**

## **Introduction**

Robot-assisted surgical systems are dedicated to achieving minimally-invasive surgeries with improved patient outcomes. This dissertation documents my efforts in developing advanced fluoroscopic navigation systems for robot-assisted orthopedic applications. The goal is to navigate the surgical robotic device to planned positions using purely X-ray image information. This Chapter provides an overview of robot-assisted orthopedic surgery, fluoroscopic navigation, and the fundamentals of one key technique: 2D/3D registration. The dissertation outline, major contributions, and publications are listed at the end of this Chapter.

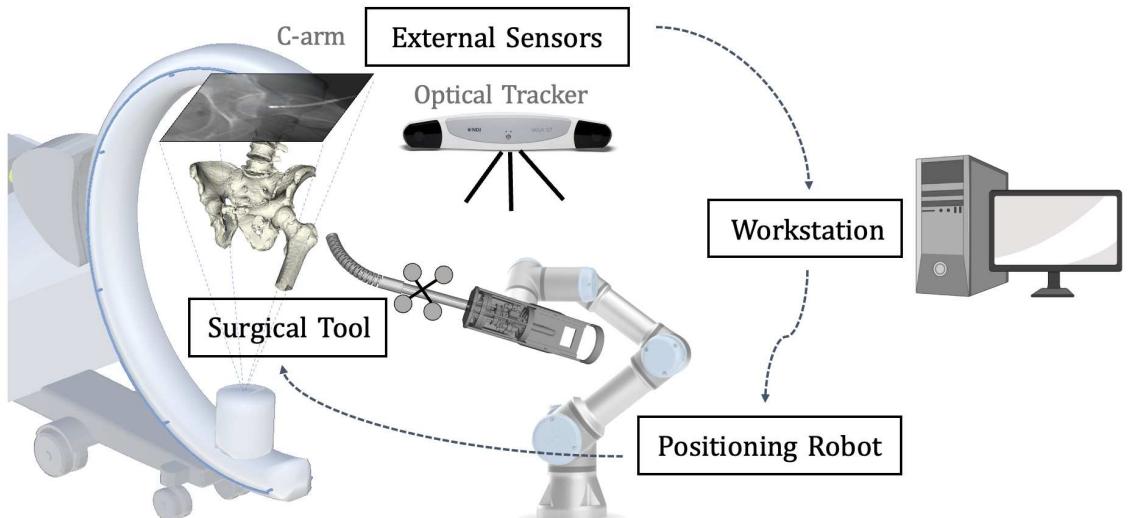
## CHAPTER 1. INTRODUCTION

### 1.1 Robot-Assisted Orthopedic Surgery

Since late 1980s, robots have been investigated to assist surgeries. ROBODOC was the first orthopedic robotic system developed for total hip replacement in 1992 [6–8]. Nowadays, there has been a vastly increased number of robotic systems designed for various orthopedic applications [9]. Robots are superior in precision, stability and repeatability than human's free hands, and, therefore, are more successful in achieving the goals of minimally-invasive surgeries and lead to improved patient outcomes [10].

A typical robot-assisted orthopedic surgical system has the following components (Fig. 1.1):

- Positioning Robot: A robot arm with high precision and stability that holds and positions surgical tools attached to the robot's end effector.



**Figure 1.1:** Concept of a typical robot-assisted orthopedic surgical system.

## CHAPTER 1. INTRODUCTION

- Surgical Tools: Instruments that are manipulated by the surgeon for passive systems or operated by the robot controller for active systems.
- External Sensors: Capture the intra-operative data of the tool and patient. Typical sensors include optical tracker, imaging unit such as fluoroscopy C-arm or CT machine, and force sensor, etc.
- Workstation: Collects the sensor's data, runs programs like image registrations, controls the robot, and displays results to the surgeons.

An effective navigation system is critical to the performance of a robotic surgical system. The navigation system aims to accurately estimate the intra-operative pose relationship of the surgical tool and patient anatomy, and guide the robotic surgical tool to planned positions. Depending on the imaging sensors used in the system, there are different navigation approaches, such as fluoroscopic, CT or MRI guided, and endoscopic, etc [11]. This dissertation focuses our efforts in developing fluoroscopic navigation for robot-assisted orthopedic surgical systems.

## 1.2 Fluoroscopic Navigation

X-ray imaging systems are one of the most important evaluation tools utilized in orthopedic surgeries, because X-ray imaging is fast, low-cost and supplies in-depth structures of the bony anatomy. While radiography provides

## CHAPTER 1. INTRODUCTION

static images, fluoroscopy uses X-rays to obtain real-time images and videos of internal structure and function of the patient, which enhances surgical decision-making and decreases the duration of surgery [12]. The C-arm, an X-ray machine with a half-moon frame, was developed to make the X-ray machine mobile in all directions [13]. It is now commonly used in most orthopedic operating rooms. The C-arm fluoroscopy imaging unit is one of the most valuable tools in the armamentarium of an orthopedic surgeon [12]. Examples of fluoroscopy use cases in regular orthopedic practice include diagnosis of arthrography [14], guidance for percutaneous biopsy of the vertebra lesion [15], and pedicle screw placement [16].

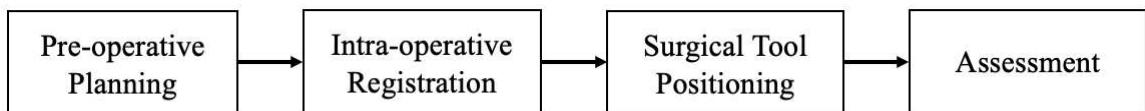
Fluoroscopic navigation is the process used to guide the surgical instruments to planned locations associated with the target anatomy structures using fluoroscopic images. In conventional procedures, fluoroscopic navigation is achieved through a “mental mapping” process by the surgeons themselves. Due to the natural projective geometry of X-ray imaging, it is very challenging to precisely estimate the 3D location of surgical tools with respect to patient anatomy. To address this challenge for accurate tool positioning, more C-arm X-ray images need to be acquired at the cost of added procedure time and radiation, increased size and number of incisions, risk of infection, and prolonged recovery time.

Unlike “mental mapping”, computer-assisted navigation computes the crit-

## CHAPTER 1. INTRODUCTION

ical surgical tool to anatomy relationship from data captured by sensing units, which guides surgical tool positioning. The workflow of computer-assisted fluoroscopic navigation can be summarized as follows:

- **Pre-operative planning:** A customized procedure plan to guide the surgical tool is made by expert clinicians from pre-operative imaging data of the patient, such as CT or MRI. The plan can be defined as entry or target points of the anatomy, a trajectory, or a series of tool motions.
- **Intra-operative registration:** The intra-operative spatial relationship of the surgical tool with respect to the patient anatomy is estimated using external sensors, such as optical tracking sensors. The sensors track the behavior of target objects and report their pose transformations. In the context of fluoroscopic navigation, the C-arm machine functions as an imaging sensor to monitor anatomical structures and tools. System calibration is commonly required to close the transformation chain of multiple sensing unit observations.
- **Surgical tool positioning:** By integrating the registration estimations to the plannings, the surgical tool is guided to the planned target. In



**Figure 1.2:** Fluoroscopic navigation system workflow.

## CHAPTER 1. INTRODUCTION

robot-assisted orthopedic surgeries, the surgical tool is operated by a robotic surgical system. This integration requires a pre-operative step of system calibration which produces a calibration matrix connecting the surgical tool frame to the robot end effector frame.

- **Assessment:** The navigation accuracy is assessed after the surgical tool is in position. This can be done by taking several verification fluoroscopic images and performing the registration again or post-operative image reconstruction scans, such as CT or MRI. If the accuracy is satisfactory, surgical procedures are performed afterward. Otherwise, we redo the registration and tool positioning steps until the performance is acceptable.

Intra-operative registration is the most important step affecting the navigation accuracy. Specifically, the spatial transformation is reported between 3D models of the surgical tool and patient anatomy. The patient model can be extracted from patient-specific 3D image reconstructions or built from statistical shape models [17]. The registration methods commonly utilize geometric features, such as landmarks or contours, to estimate 3D poses. Depending on whether physical landmark fiducials are used, the navigation approaches can be classified as fiducial-based and fiducial-free methods.

**Fiducial-based Navigation:** Optical tracking is the most popular approach used in fiducial-based orthopedic navigation systems in the literature. Fiducial markers composed of reflective balls attached to surgical tools are

## CHAPTER 1. INTRODUCTION



**Figure 1.3:** Surgeons performing total knee replacement using Stryker's Mako robotic system.

tracked by an infrared camera and determine the 3D orientation and location of the tools are determined. Optical tracking operates in real-time and achieves sub-millimeter positional accuracy. To close the registration loop, fiducial markers need to be attached to the patient anatomical structures. In Fig. 1.3, we present an example of Stryker's Mako robotic system using optical fiducial-based navigation for total knee replacement.

Fiducial-based navigation has the following limitations: First, the line of sight between the tracked fiducial markers and the camera must always be clear, limiting the surgeons' movement in the OR. Second, attaching bone pins introduces additional incision to the patient, which is invasive and prolongs recovery time. Although there are non-invasive markers designed for specific anatomical structures (e.g. Stryker's spine marker), using fiducial markers always complicate the instrument setup in the operating room.

**Fiducial-free Navigation:** Fiducial-free navigation approaches seek to re-

## CHAPTER 1. INTRODUCTION

trieve the intra-operative spatial relationship without using any external objects. Using C-arm fluoroscopy, the 3D object pose transformation is estimated using 2D X-ray image information. The process is known as 2D/3D registration. For example, Grupp et al. proposed a fiducial-free fluoroscopic navigation approach for pose estimation of periacetabular osteotomy fragments [18].

Fiducial-free navigation does not have limitations due to the line of sight issue and is less invasive to the patients. For this dissertation, we developed various novel 2D/3D registration techniques that supply fiducial-free navigation for multiple orthopedic applications. We introduce 2D/3D registration fundamentals in the next section.

### 1.3 2D/3D Registration Fundamentals

As introduced in the previous section, 2D/3D registration is the process that estimates the pose of the 3D objects, such as the CT, from 2D images, such as the X-ray. In this dissertation, image intensity-based 2D/3D registration is the basis of fluoroscopic navigation techniques. In this section, we introduce the fundamentals of 2D/3D registration, including problem formulation and challenges.

## CHAPTER 1. INTRODUCTION

### 1.3.1 Problem Formulation

In general, 2D/3D registration aims at finding the 3D pose of shape parameters of an object by optimizing a similarity function of the observed target image and digitally reconstructed radiograph (DRR) image obtained from the 3D object data. Given a set of 3D volumetric data  $V_m$  and 2D images  $I_n$ , where subscripts  $m, n$  suggest multiple 3D objects or 2D image views, respectively, object pose or shape parameter  $\theta_m$ , a DRR projection operator  $\mathcal{P}$ , a similarity function  $\mathcal{S}$ , a regularizer  $\mathcal{R}$  over parameters, the optimization process can be formulated as follows:

$$\{\theta_m, m \in \{0, \dots, M\}\} = \min_{\theta_m} \sum_{n=0}^N \mathcal{S}\left(I_n, \sum_{m=0}^M \mathcal{P}(V_m; \theta_m)\right) + \mathcal{R}(\theta_m) \quad (1.1)$$

In terms of rigid registration where the deformation is discarded, the parameters  $\theta_m \in SE(3)$  refer to rigid pose transformation, which have 6 degrees of freedom (DoF). In conventional image-based 2D/3D registration, the optimization of Eq. 1.1 is usually performed iteratively from an initial  $\theta_m$  estimation. The updates  $\delta\theta_m$  are derived from Eq. 1.1 using gradient-based or gradient-free methods, such as BFGS [19, 20] or CMA-ES [21] and BOBYQA [22], respectively.

In certain cases when 2D and 3D data representations are not pixels or voxel grid-based but sparse features, e.g., landmarks, analytic solutions to

## CHAPTER 1. INTRODUCTION

Eq. 1.1, such as the perspective n point (PnP) algorithm [23], exist, which is considered as feature-based 2D/3D registration.

### 1.3.2 Challenges

Unfortunately, despite substantial efforts over the past decades, 2D/3D registration is not yet widely used as a robust image navigation technique at the bedside, because of the existence of the following hard open challenges [24]:

- **Narrow capture range of similarity metrics:** Conventional intensity-based methods mostly use hand-crafted similarity metrics between  $I_n$  and the DRR projection  $\mathcal{P}(V_m; \theta_m)$  as loss function. Common choices for the similarity metric are Normalized Cross Correlation (NCC), gradient information [20], or Mutual Information (MI) [25]. These metrics are generally non-convex and fail to accurately represent pose offsets when perturbations are large. Thus, without proper initialization, the optimization is prone to get stuck in local minima, returning wrong registration results. Thus, the initial estimate of the target parameters  $\theta_m$  must be close enough to the true solution for the optimization to converge to the global minima. Estimating good initial parameters is commonly achieved using some manual interaction such as landmark annotations, which is cumbersome and time-consuming.

## CHAPTER 1. INTRODUCTION

- **Ambiguity:** The aforementioned complexity also leads to registration ambiguity, which is most pronounced in single-view registration [26, 27]. Because the spatial information along the projection line is collapsed onto the imaging plane, it is hard to precisely recover the information in the projective direction. A well-known example is the difficulty of accurately estimating the depth of the 3D scene from the camera center using a single 2D image. These challenges already exist for rigid 2D/3D registration and are further exacerbated in rigid plus deformable registration settings.
- **Verification and uncertainty:** As a central component of image-based surgical navigation platforms, 2D/3D registration supplies critical information to enable precise manipulation of anatomy. To enable users to assess risk and make better decisions, there is a strong desire for registration algorithms to verify the resulting geometric parameters or supply uncertainty estimates. Perhaps the most straightforward way of verifying a registration result is to visually inspect the 2D overlay of the projected 3D data – this approach, however, is neither quantitative nor does it scale since it is based on human intervention.

These open problems can be largely attributed to the variability in the problem settings (e. g., regarding image appearance and contrast mechanisms, pose variability, ...) that cannot easily be handled algorithmically because the desired properties cannot be formalized explicitly. Due to these open chal-

## CHAPTER 1. INTRODUCTION

lenges, image-based navigation solutions using 2D/3D registration are custom designed with certain assumptions or constraints for each clinical application, simply because the challenges of each application are unique. In this dissertation, we present our efforts in developing navigation solutions for the investigated orthopedic applications, which are based on the above registration fundamentals.

## 1.4 Thesis Overview

### 1.4.1 Outline

Chapter 2 introduces clinical background of the investigated orthopedic applications.

Chapter 3 presents a novel method for generating large-scale realistically synthetic X-ray images to train machine learning models. The downstream tasks of hip imaging and surgical robotic tool detection are used for initializing the 2D/3D registration of the pelvic bone (Chapter 4) and continuum manipulator (Chapter 6).

Chapters 4 and 6 present key registration methods for the relative pose estimation of the surgical tool and the patient anatomy.

Chapter 5 introduces a novel projective spatial transformer module that is

## CHAPTER 1. INTRODUCTION

able to compute differentiable gradients of projection operation. We demonstrate its benefit of largely extending the 2D/3D registration capture range by learning a convex shape similarity function.

Chapter 7 demonstrates the efforts of system integration and *ex vivo* cadaveric experiments on the above introduced orthopedic clinical applications.

Chapter 8 concludes the dissertation and provides future work outlooks.

### 1.4.2 Contributions

The contributions of this dissertation are as follows:

- An X-ray image analysis pipeline using realistically synthesized data and domain generalization techniques [28], which we refer to as SyntheX. This work demonstrates that deep network models trained using SyntheX perform on par or superior to the models trained using real data on variant clinical tasks, and from precisely controlled ablation experiments (Chapter 3).
- Fiducial-free image intensity-based methods for pose estimation of bony anatomy, including the pelvis, femur [3, 29] and spine vertebrae [30], using multi-view X-ray images (Chapter 4).
- A learning-based 2D/3D registration method using a novel projective spatial transformer module, which largely extends the initialization capture

## CHAPTER 1. INTRODUCTION

range by approximating a convex shape loss function. Superior performance is presented compared with conventional 2D/3D registration methods [31]. (Chapter 5)

- Image intensity-based methods for pose estimation of the rigid and flexible surgical tools, including the custom-designed injection device and flexible continuum manipulator [3, 32]. (Chapter 6)
- System-level verification including system calibration and testing, and efforts on clinical studies with *ex vivo* cadaveric experiments on the applications of femoroplasty [3], core decompression of the hip [33] and transforaminal lumbar epidural injections [30]. (Chapter 7)

### 1.4.3 Publications

The dissertation is written based on the following publications

- Gao, C.\*, Phalen, H.\*<sup>1</sup>, Margalit, A., Ma, J.H., Ku, P., Unberath, M., Taylor, R. H., Jain, A. and Armand, M. Fluoroscopy-Guided Robotic System for Transforaminal Lumbar Epidural Injections. *IEEE transactions on medical robotics and bionics*, 2022, doi: 10.1109/TMRB.2022.3196321. (Chapter 4, 6, and 7 in this dissertation) \* indicates joint first co-authors.
- Gao, C., Phalen, H., Sefati, S., Ma, J.H., Taylor, R. H., Unberath, M. and Armand, M. 2021. Fluoroscopic navigation for a surgical robotic system

## CHAPTER 1. INTRODUCTION

including a continuum manipulator. *IEEE Transactions on Biomedical Engineering*, 69(1), 453-464. (Chapters 6 and 7 in this dissertation)

- Gao, C., Liu, X., Gu, W., Killeen, B., Armand, M., Taylor, R.H. and Unberath, M. 2020, October. Generalizing spatial transformers to projective geometry with applications to 2D/3D registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 329-339). Springer, Cham. (Chapter 5 in this dissertation)
- Gao, C., Farvardin, A., Grupp, R.B., Bakhtiarinejad, M., Ma, L., Thies, M., Unberath, M., Taylor, R.H. and Armand, M., 2020. Fiducial-free 2D/3D registration for robot-assisted femoroplasty. *IEEE transactions on medical robotics and bionics*, 2(3), pp.437-446. (Chapter 4 and 6 in this dissertation)
- Gao, C., Grupp, R.B., Unberath, M., Taylor, R.H. and Armand, M., 2020, March. Fiducial-free 2D/3D registration of the proximal femur for robot-assisted femoroplasty. In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling* (Vol. 11315, p. 113151C). International Society for Optics and Photonics. (Chapter 4 in this dissertation)
- Gao, C., Unberath, M., Taylor, R. and Armand, M., 2019. Localizing dexterous surgical tools in X-ray for image-based navigation. arXiv preprint

## CHAPTER 1. INTRODUCTION

arXiv:1901.06672. (Chapter 6 in this dissertation)

and the following preprint manuscript that is under revision upon submission of the dissertation:

- Gao, C., Killeen, B.D., Hu, Y., Grupp, R.B., Taylor, R.H., Armand, M. and Unberath, M., 2022. SyntheX: Scaling Up Learning-based X-ray Image Analysis Through In Silico Experiments. arXiv preprint arXiv:2206.06127. *Nature Machine Intelligence*, under review (Chapter 3 in this dissertation)

# **Chapter 2**

## **Orthopedic Application**

### **Background**

In this Chapter, we provide clinical background introduction to three orthopedic applications that are studied in this dissertation work. They are femoro-plasty, core decompression of the hip, and transforaminal lumbar epidural injection. Since each application has its unique requirements and challenges, we hope that the readers can gain more contextual information of the clinical motivations before getting to the rest of this dissertation.

## CHAPTER 2. ORTHOPEDIC APPLICATION BACKGROUND

### 2.1 Femoroplasty

Osteoporosis, characterized by decreased bone mass and microarchitectural deterioration of bone tissue, is a global health problem. There is a wealth of research on the extent to which bone loss may impair strength and increase the risk of fracture. The rate of mortality one year after an osteoporotic hip fracture has been reported to be between 20% and 45% [34, 35], and the rate of a second fracture occurrence is 6–10 times higher in patients who have already suffered an osteoporotic hip fracture [36]. Fig. 2.1 (a) shows illustrations of femur fractures<sup>1</sup>. Current approaches to reducing the risk of hip fracture consider long-term preventive measures such as estrogens and selective estrogen receptor modulators, calcitonin, and bisphosphonates. These approaches, however, do not provide much-needed immediate prevention, especially for the elderly.

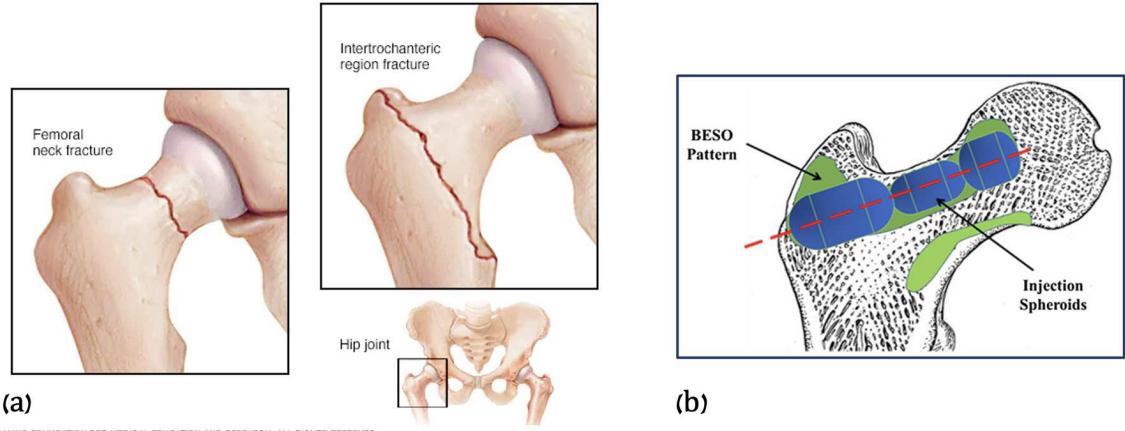
Femoroplasty is proposed as an alternative therapeutic treatment for patients with osteoporosis [37]. It aims to prevent potential osteoporotic hip fractures by augmenting the osteoporotic femoral neck and trochanter area by injecting bone cement such as Polymethylmethacrylate (PMMA) [38]. Fig. 2.1 (b) illustrates an example femoroplasty augmentation scheme. Recent studies suggest that femoroplasty is a short-term preventive approach and has the

---

<sup>1</sup>Image from Mayo Clinic Foundation: <https://www.mayoclinic.org/diseases-conditions/hip-fracture/multimedia/hip-fractures>

## CHAPTER 2. ORTHOPEDIC APPLICATION BACKGROUND

potential to reduce the risk of fracture in an osteoporotic hip [1, 38–40].



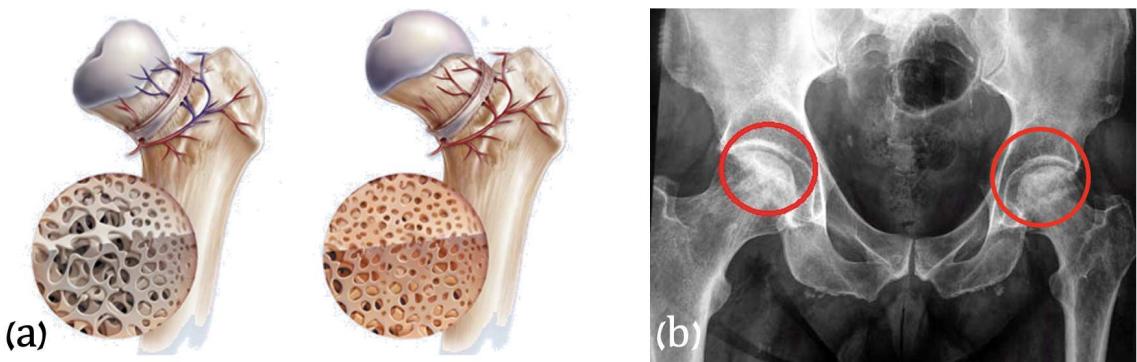
**Figure 2.1:** (a) Illustration of femur fractures. (b) An example femoroplasty bone augmentation scheme [1].

Patient-specific femoroplasty requires precise intraoperative navigation of the injection device according to a planned trajectory. In this dissertation, we will present a robotic system with a custom-designed injection device for robot-assisted femoroplasty. We propose a navigation pipeline using C-arm fluoroscopic images to estimate the pose of the injection device and the femur, which guides the device to the planning trajectory to perform femoroplasty.

## 2.2 Core Decompression of the Hip

Osteonecrosis of the hip is a disease that results from loss of blood supply to the bone. It is a devastating condition affecting patients between the ages of 30 - 50 that usually progresses to femoral head collapse, and eventually leads to

## CHAPTER 2. ORTHOPEDIC APPLICATION BACKGROUND



**Figure 2.2:** (a) Illustration of femur head osteonecrosis. (b) An example hip X-ray of patient with femoral head osteonecrosis.

total hip arthroplasty [41]. Illustrations of osteonecrosis and an example X-ray image of the osteonecrotic femoral heads is shown in Fig. 2.2<sup>2</sup>. To reduce the pressure in the femoral head, enhance vascular flow, and alleviate pain, core decompression was developed to treat osteonecrosis more than three decades ago. It is the most common early-stage treatment of osteonecrosis to preserve the femoral head from total hip replacement [42, 43].

Core decompression consists of using a drill to remove the 8-10 mm cylindrical core from an osteonecrotic lesion [44]. Complete removal of a lesion in the femoral head requires access by drilling through the narrow femoral neck and then debriding a larger volume of necrotic bone beyond this access point. Incomplete debridement of the entire lesion results in incomplete removal of the necrotic bone and potential collapse or damage to the cartilage overlying the femoral head [45]. High accuracy is needed to remove the minimal healthy

<sup>2</sup>Image from <https://www.spineorthocenter.com/services/orthopedic/joint-diseases/osteonecrosis> and <https://www.goudelis.gr/en/content/femoral-head-osteonecrosis-avascular-necrosis>

## CHAPTER 2. ORTHOPEDIC APPLICATION BACKGROUND

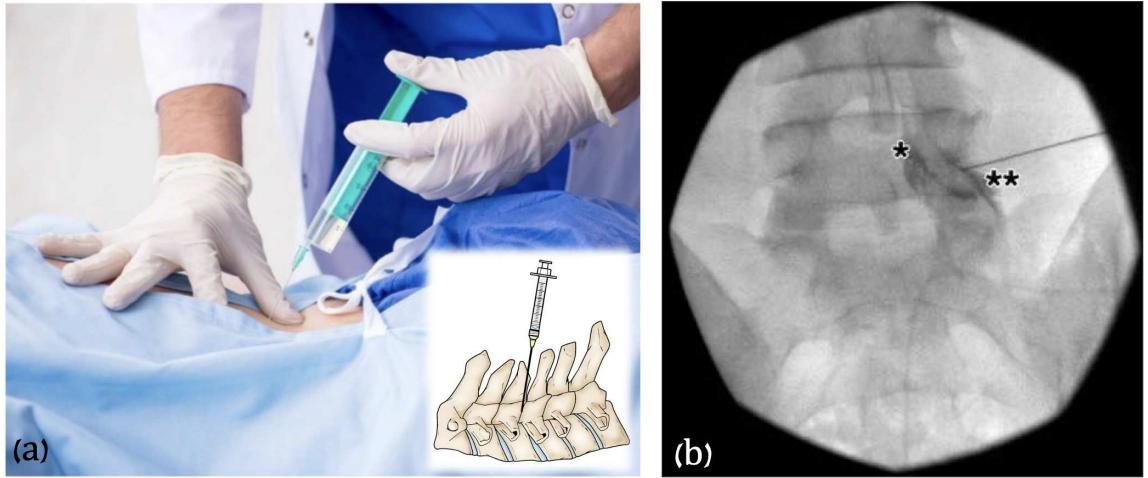
bone to maintain structural integrity and stability, especially in the narrowest regions of the femoral neck. However, to debride a larger volume once having passed through the narrow femoral neck, significant dexterity of the tool is required.

In this dissertation, we will present a surgical robotic system equipped with a flexible continuum manipulator with high dexterity, which is designed for the less-invasive treatment of osteonecrosis. We propose a fluoroscopic navigation pipeline to estimate its pose and shape with respect to the femoral bone, which is used to position the continuum manipulator and track its intra-operative updates.

## 2.3 Transforaminal Lumbar Epidural In- jection

Epidural steroid injection in the lumbar spine is a common non-surgical treatment for lower back pain or sciatica. Globally, between 60-80% of people are estimated to experience lower back pain in their lifetime and it is among the top causes of adult disability [46, 47]. Since the 1950s, epidural steroid injections have been utilized for conservative treatment of lower back pain [48]. Efficacy of treatment is reported as 84%, with adequate targeting of the injection site thought to be critical to successful treatment [49]. This is currently the

## CHAPTER 2. ORTHOPEDIC APPLICATION BACKGROUND



**Figure 2.3:** (a) Illustration of spine injection. (b) An X-ray image taken after lumbar epidural injection with needle tip remaining. This image was reported in [2].

most frequently performed procedure in pain medicine in the United States [2].

The 3 main approaches for administering these injections in the lumbar spine include transforaminal, interlaminar, and caudal approaches. The main advantage of the transforaminal approach is the presumed ability to deliver medications as close as possible to the lumbar nerve roots [2]. Illustration of the injection is shown in Fig. 2.3 (a)<sup>3</sup>.

These injections are frequently delivered under fluoroscopic, computed tomography (CT), or ultrasound guidance to increase the accuracy of needle placement. To our knowledge, there are no differences in outcomes between these modalities, although fluoroscopy is the most commonly utilized by clinician [50].

The clinician will acquire several images before and during the manual inser-

<sup>3</sup>Images from <https://www.spineuniverse.com/treatments/pain-management/thoracic-epidural-injection> and <https://orthoinfo.aaos.org/en/treatment/spinal-injections>

## CHAPTER 2. ORTHOPEDIC APPLICATION BACKGROUND

tion of the needle. When satisfied with needle placement, the clinician will inject a steroid and remove the needle. An example X-ray image was taken after epidural injection is shown in Fig. 2.3 (b). Several injections at different levels of the spine may be performed in sequence. The clinical effect of TLESI is related to the diffused pattern of the injected steroid agent. However, failure injections may cause damage to the vessels or nerves, which can result in symptoms of spinal nerve pricking or more complicated issues [51, 52].

In this dissertation, we will present an autonomous robotic lumbar epidural steroid injection system using fluoroscopic navigation, which suggests improved injection accuracy compared to an experienced clinician's manual injection.

# **Chapter 3**

## **SyntheX: Realistic Synthesis for X-ray Image Analysis**

### **3.1 Introduction**

In this Chapter, we present our efforts in enabling learning-based X-ray image analysis through in silico experiments demonstrated on two downstream clinical tasks. The selected learning-based X-ray imaging analysis tasks will be used for 2D/3D registration initialization as part of the proposed fluoroscopic navigation pipelines. Specifically, the pelvic anatomical landmark detection method in the hip imaging task is used to initialize the pelvis 2D/3D registration in Chapter. 4.4.2.1. In addition, the technique of continuum manipulator detection in the robotic surgical tool detection task is used in Chapter. 6.4.2.1.

## CHAPTER 3. SYNTHEX

We introduce details of each use case in the following associated Chapters.

The envisioned learning-based methods for both clinical tasks require a labeled image dataset for network training. However, acquiring large-scale real clinical X-ray data with expert annotation is particularly challenging due to its incompatibility with the current clinical routine and the time-consuming annotation process. This Chapter demonstrates that realistic simulation of image formation from human models is a viable alternative to large-scale real data collection. Combined with contemporary domain randomization or domain adaptation techniques, training on realistically synthesized data results in machine learning models that – on real data – perform comparably to models trained on the *precisely matched* hip imaging real data training set. Furthermore, because the synthetic generation of training data from human-based models can easily scale up the number of individual samples available for training, we find that the proposed model transfer paradigm for X-ray image analysis, which we refer to as SyntheX, can even outperform real data-trained models due to the effectiveness of training on larger, more varied data. Finally, we show the superiority of SyntheX across two aforementioned different X-ray imaging clinical tasks, demonstrating its general usage in X-ray image machine learning analysis applications.

## 3.2 Contributions

The contributions of this Chapter are as follows:

- We propose a synthetic training data generation pipeline for machine learning X-ray image analysis tasks, which we refer to as SyntheX. We demonstrate the clinical applications of SyntheX on two downstream tasks, namely hip imaging and surgical robotic tool detection.
- Through conducting precisely controlled experiments on a unique pelvic benchmark dataset, we isolate and quantify the role of domain shift in the deterioration of machine learning model performance from training in simulation to deployment on real data.

To our knowledge, no study to date has isolated the effect of domain generalization using precisely matched datasets across domains because generating or collecting matched data is a very difficult problem. It becomes possible in the case of X-ray image analysis due to advances in 2D/3D image registration technology. This work also, for the first time, demonstrates a feasible and cost-effective way to train AI X-ray image analysis models for clinical intervention on synthetic data in a way that provides comparable performance to training on real clinical data in multiple applications. Our efforts suggest that SyntheX catalyzes and drastically accelerates the design, conception, and evaluation of intelligent surgical systems for X-ray-guided procedures.

## CHAPTER 3. SYNTHEX

The work of concurrent localization and segmentation of the continuum manipulator is presented as a long abstract at the 10th International Conference on Information Processing in Computer-Assisted Interventions (IPCAI 2019) [53]. The work of SyntheX and the controlled experiments is published in an Arxiv preprint [28], and is under revision of Nature Machine Intelligence upon submission of this dissertation.

### 3.3 Background

Advances in robotics and artificial intelligence (AI) are bringing autonomous surgical systems closer to reality. However, developing the AI backbones of such systems currently depends on collecting training data during routine surgeries. This remains one of the largest barriers to widespread use of AI systems in interventional clinical settings, versus triage or diagnostic settings, as the acquisition and annotation of interventional data are time-intensive and costly. Furthermore, while this approach can contribute to the automation or streamlining of *existing* surgical workflows, robotic and autonomous systems promise even more substantial advances: novel and super-human techniques that improve outcomes, save time, or mitigate human error. This is perhaps the most exciting frontier of computer-assisted intervention research.

Conventional approaches for curating data for AI development (i.e., sourc-

## CHAPTER 3. SYNTHEX

ing it retroactively from clinical practice) are insufficient for training AI models that benefit interventions that use novel instrumentation, different access points, or more flexible imaging. This is because they are, by definition, incompatible with contemporary clinical practices and such data does not emerge from routine care. Furthermore, these novel systems are not readily approved, and thus not easily or quickly introduced into clinical practice. *Ex vivo* experimentation does not suffer the same ethical constraints; however, it is costly and requires mature prototypes, and therefore does not scale well.

A promising alternative to these strategies is simulation, i.e., the *in silico* generation of synthetic interventional training data and imagery from human models. Simulation offers a rich environment for training both human and machine surgeons alike, and sidesteps ethical considerations that arise when exploring procedures outside the standard of care. Perhaps most importantly, *in silico* surgical sandboxes enable rapid prototyping during the research phase. Simulation paradigms are inexpensive, scalable, and rich with information. While intra-operative data is generated in highly unstructured and uncontrolled environments and requires manual annotation, simulation can provide detailed ground truth data for every element of the procedure, including tool and anatomy pose, which are invaluable for AI development.

However, simulations can fall short of real surgery in one key aspect: realism. The difference in characteristics between real and simulated data is

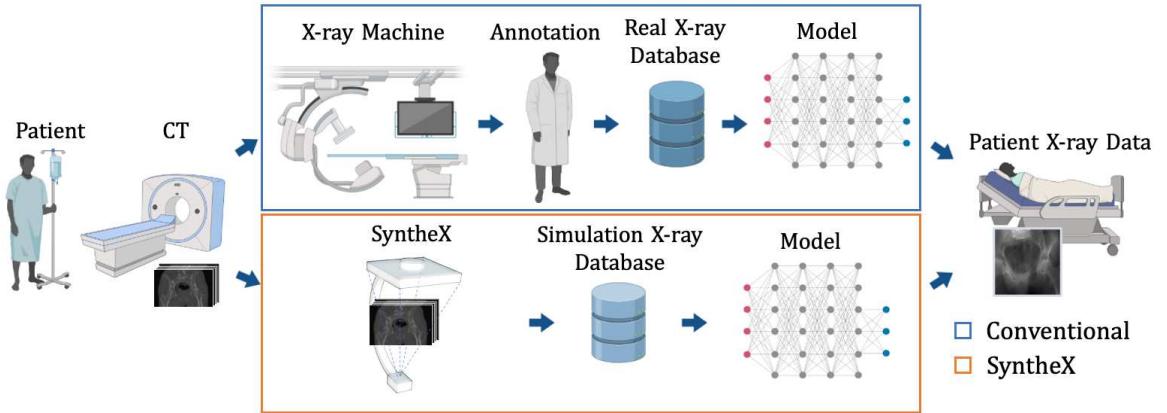
## CHAPTER 3. SYNTHEX

commonly referred to as the “domain gap.” The ability of an AI model to perform on data from a different domain, i.e. with a domain gap from the data it was trained on, is called “domain generalization.” Domain gaps are problematic because of the well-documented brittleness of AI systems [54], which exhibit vastly deteriorated performance across domain gaps. This may happen even with simple differences, such as noise statistics, contrast level, and other minutiae [55–58]. This unfortunate circumstance, which applies to all machine learning tasks, has motivated research in the AI field on *Sim2Real* transfer, and the development of domain transfer methods.

In this work, we describe advances in X-ray simulation and domain transfer methods, which, when combined, contribute feasible solutions for training AI algorithms on synthetic data while preserving their performance under domain shift for evaluation and deployment in the real world. The overall concept of training deep neural networks on realistically simulated data from annotated CT scans using domain randomization, which we refer to as SyntheX, is illustrated in Fig. 3.1 and we demonstrate its utility and validity in two clinical applications: hip imaging and surgical robotic tool detection. At the core of our report is an experiment on precisely controlled data that isolates and quantifies the effect of domain shift for AI-based X-ray image analysis. Using CT images from human cadavers and corresponding C-arm X-ray images acquired during surgical exploration, we generated a hip image dataset consisting of ge-

## CHAPTER 3. SYNTHEX

ometrically identical images across various synthetic and real domains to train AI models for hip image analysis.

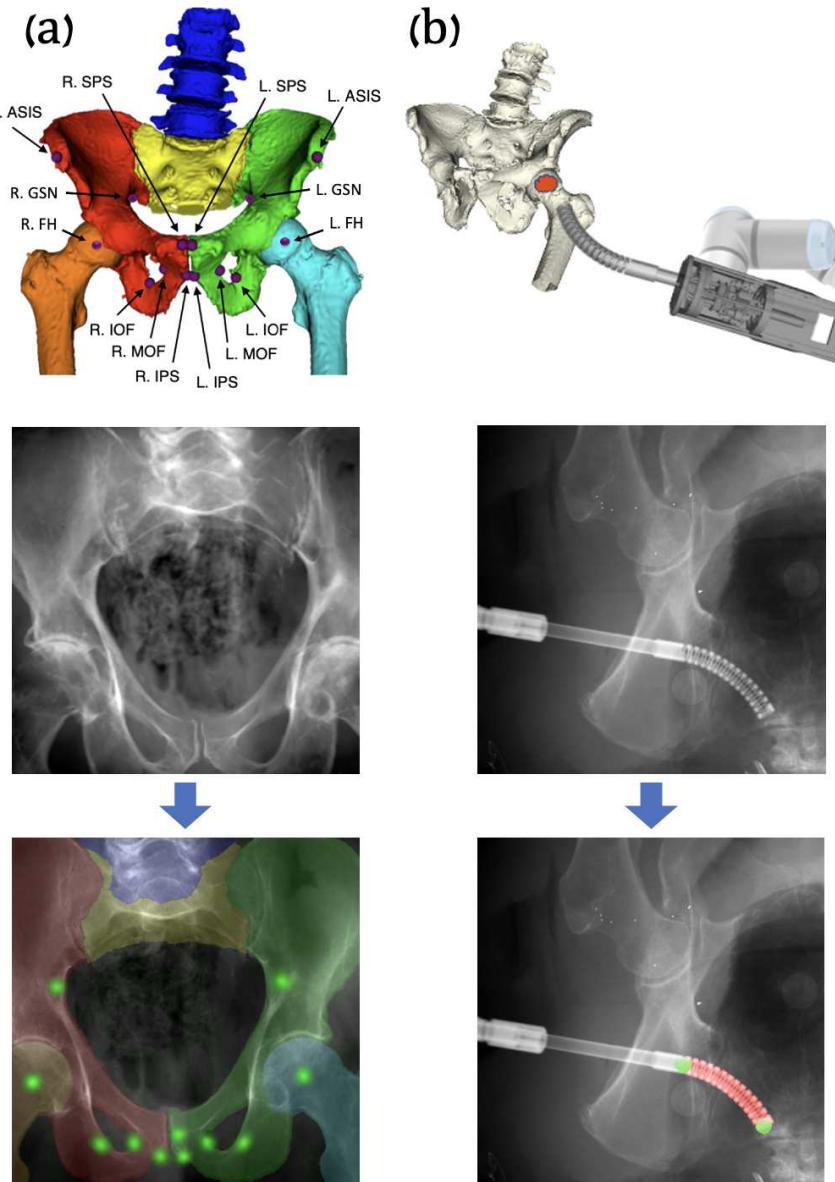


**Figure 3.1:** Overall concept of SyntheX. Blue: Conventional approach for learning-based tasks on medical imaging. Curating a relevant database of real X-ray samples requires real data acquisition and costly annotation from domain experts. Orange: SyntheX enables simplified and scaled-up data curation because data generation is synthetic and synthesized data can be annotated automatically through propagation from the 3D model. SyntheX results in deep learning image analysis models that perform comparably to or better than real-data trained models.

## 3.4 Down-Stream Clinical Tasks

We demonstrate the benefits of SyntheX on two X-ray image analysis downstream tasks: hip imaging and surgical robotic tool detection (Fig. 3.2). Both tasks use deep neural networks to make clinically meaningful predictions on X-ray images. We introduce the clinical motivations for each task in the following sections. Details of the deep network and training/evaluation paradigm are described in Section 3.7.

## CHAPTER 3. SYNTHEX



**Figure 3.2:** (a) Hip imaging. The hip anatomical structures include left and right hemipelvis, lumbar vertebrae, upper sacrum, and left and right femurs. The anatomical landmarks consist of left and right anterior superior iliac spine (ASIS), center of femoral head (FH), superior pubic symphysis (SPS), inferior pubic symphysis (IPS), medial obturator foramen (MOF), inferior obturator foramen (IOF), and the greater sciatic notch (GSN). These landmarks are useful in identifying the anterior pelvic plane (APP) and initializing the 2D/3D registration of both pelvis and femur [3, 4]. (b) Surgical robotic tool detection. An illustration of the robotic surgical system including a continuum manipulator is shown on the top. An example real X-ray image and the corresponding segmentation and landmarks of the continuum manipulator is shown on the bottom.

### 3.4.1 Hip Imaging

One of the main challenges in fluoroscopic image-based navigation is estimating spatial tool-to-tissue relationships from 2D transmission X-ray images. One effective approach to achieving spatial alignment is the identification of known structures and landmarks in the 2D X-ray image, which then are used to infer poses [24, 59]. In the context of hip imaging, we define six anatomical structures and fourteen landmarks as the most relevant known structures. They are shown in Fig. 3.2 (a). We trained deep networks using SyntheX to make these detections on X-ray images. Synthetic images were generated using CT scans selected from the New Mexico Decedent Image Database [60]. The 3D anatomical landmarks were manually annotated and the anatomical structures were segmented using the automatic method described in [61], which were then projected to 2D as labels following the simulation X-ray geometries. We evaluate the performance of our model on 366 real X-ray images collected from six cadaveric specimens. On real images, ground truth target structures were annotated semi-automatically. This real dataset also serves as the basis for our precisely controlled experiments that isolate the effect of the domain gap. More details on the creation, annotation and synthetic duplication of this dataset are provided in Section 3.5.1.

### 3.4.2 Surgical Robotic Tool Detection

Automatic detection of the surgical tool from intra-operative images is an important step for robot-assisted surgery since it enables vision-based control [62]. Because training a detection model requires sufficient image data with ground truth labels, developing such models is only possible after the surgical robot is mature and deployed clinically. We demonstrate AI model development for custom and pre-clinical surgical robotic tools.

We consider a specific continuum manipulator (CM) as the target object. This CM is custom-designed for minimally-invasive robot-assisted orthopedic procedures which have significant dexterity and stiffness [63, 64]. We introduce more details of the CM design, modeling, and use cases in Chapter 6.4. Using SyntheX, we address CM detection, which consists of segmenting the CM body and predicting distinct landmarks in the X-ray images. The semantic segmentation mask covers the 27 alternating notches which discern the CM from the other surgical tools; the landmarks are defined as the start and end points of the CM centerline [65]. These detection results are used to initialize CM 2D/3D registration described in Chapter 6.4.2.1. Synthetic images were generated using CT scans selected from the New Mexico Decedent Image Database [60] and a computer-aided design model of the CM. 3D CM segmentations and landmark locations were determined through forward kinematics and then projected to 2D as training labels using the X-ray geometry. The

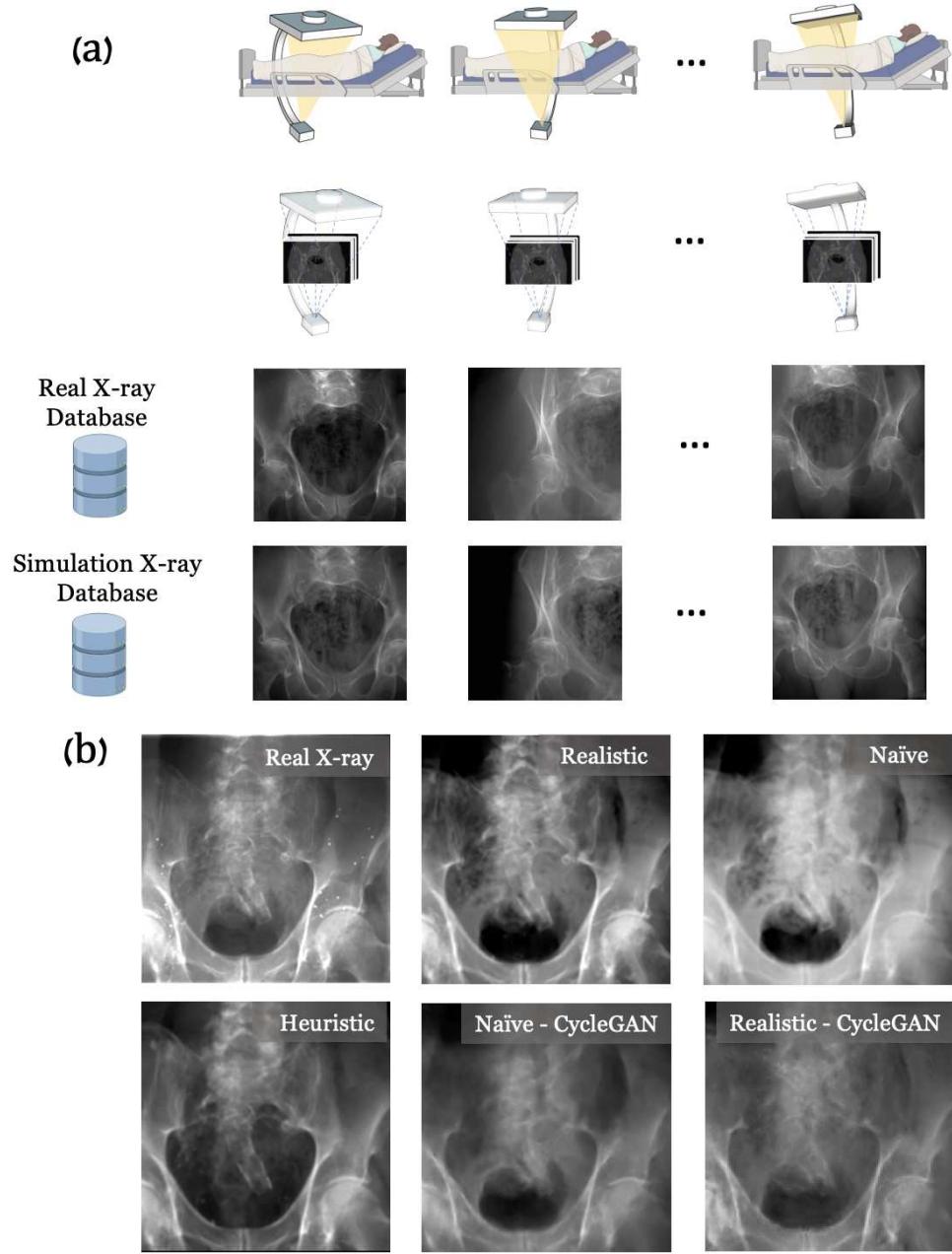
## CHAPTER 3. SYNTHEX

performance was evaluated on 264 real X-ray images of the CM during pre-clinical cadaveric testing. On real images, ground truth segmentation masks and landmark locations were annotated manually.

### **3.5 Precisely Controlled Investigations on SyntheX’ Sim2Real Performance Using the Hip Imaging Usecase**

Beyond presenting SyntheX for various clinical tasks, we present experiments on a unique dataset for hip imaging that enables the isolation of the effect that the domain gap has on simulation to reality (*Sim2Real*) AI model transfer. On the task of anatomical landmark detection and anatomy segmentation in hip X-rays, we study the most commonly used domain generalization techniques, namely domain randomization, and domain adaptation, and further consider different X-ray simulators, image resolution, and training dataset size. We introduce details on these experiments next.

## CHAPTER 3. SYNTHEX



**Figure 3.3:** (a) Generation of precisely matched synthetic and real X-ray database: Real X-rays and CT scans are acquired from cadaveric specimens and registered to obtain the relative camera poses. Using these poses, synthetic X-rays can be generated from the CTs that precisely match the real X-ray data in all aspects but appearance. (b) Changes in (synthetic) X-ray appearance based on simulation paradigm.

### 3.5.1 Precisely Matched Hip Dataset

We created an accurately annotated dataset of 366 real hip fluoroscopic images and corresponding high-resolution CT scans of six lower torso cadaveric specimens with manual label annotations [59], which constitutes the basis of our unique dataset that enables precisely controlled benchmarking of domain shift. For each of the real X-ray images, the X-ray camera pose was accurately estimated using a comprehensive 2D/3D image registration pipeline [59]. We then generated synthetic X-ray images (DRRs) that precisely recreate the spatial configurations and anatomy of the real X-ray images and only differ in the realism of the simulation (Fig. 3.3 (a)). Because synthetic images precisely match the real dataset, all labels in 2D and 3D apply equally.

We studied three different X-ray image simulation techniques, including naïve DRR generation, xreg DRR [18] and DeepDRR [66, 67], which we will refer to as *Naïve*, *Heuristic* and *Realistic* simulations. They differ in the considerations of modeling realistic X-ray imaging physical effects. Fig. 3.3 (b) presents a comparison of image appearance between the different simulators and a corresponding real X-ray image. Details of the dataset creation are introduced in the following Section. 3.6.

### 3.5.2 Domain Randomization and Adaptation

Domain randomization is a domain generalization technique that inflicts drastic changes on the input image appearances. This produces training samples with dramatically altered appearance, which forces the network to discover more robust associations between input image features and desired target. These more robust associations have been demonstrated to improve the generalization of machine learning models when transferred from one domain to another (here: from simulated to real X-ray images, respectively). We implemented two levels of domain randomization effects, namely regular domain randomization and strong domain randomization. Details are described in Section A.0.1.

Other than domain randomization which does not assume knowledge or sampling of the target domain at training time, domain adaptation techniques attempt to mitigate the domain gap’s detrimental effect by aligning features across the source (training domain, here: simulated data) and the target domain (deployment domain, here: real X-ray images). As such, domain adaptation techniques require samples from the target domain at training time. Recent domain adaptation techniques have increased the suitability of the approach for *Sim2Real* transfer because they now allow for the use of unlabeled data in the target domain. We conducted experiments using two common domain adaptation methods: CycleGAN [68] and adversarial discriminative do-

## CHAPTER 3. SYNTHEX

main adaptation (ADDA) [69]. The two methods are similar in that they attempt to align properties of real and synthetic domains and differ based on what properties they seek to align. While CycleGAN operates directly on the images, ADDA seeks to align higher-level feature representations, i. e., image features after multiple convolutional neural network layers. Example CycleGAN generated images are shown in Fig. 3.3 (b). More details of CycleGAN and ADDA training are provided in Section A.0.2.

## 3.6 Dataset and Experimental Details

The SyntheX simulation environment was set up to approximate a Siemens CIOS Fusion C-Arm, which has image dimensions of  $1536 \times 1536$ , isotropic pixel spacing of 0.194 mm/pixel, a source-to-detector distance of 1020 mm, and a principal point at the center of the image.

### 3.6.1 Hip Imaging

Synthetic hip X-rays were created using 20 CT scans from the New Mexico Decedent Image Database [60]. During simulation, we uniformly sampled the CT volume rotation in  $[-45^\circ, 45^\circ]$ , and translation Left/Right in [-50 mm, 50 mm], Interior/Superior in [-20 mm, 20 mm], Anterior/Posterior in [-100 mm, 100 mm]. We generated 18,000 images for training and 2,000 images for vali-

## CHAPTER 3. SYNTHEX

dation. Ground truth segmentation and landmark labels were projected from 3D using the projection geometry.

We consistently trained the model for 30 epochs and selected the best validation model for evaluation. Strong domain randomization was applied at training time (Section. A.0.1). During evaluation, a threshold of 0.5 was used for segmentation and the landmark prediction was selected using the highest heatmap response location.

### 3.6.2 Robotic Surgical Tool Detection

We created 100 voxelized models of the continuum manipulator (CM) in various configurations by sampling its curvature control point angles form a Gaussian distribution  $N(\mu = 0, \sigma = 2.5^\circ)$ . The CM base pose was uniformly sampled for in plane bending rotation in  $[-30^\circ, 30^\circ]$ , out of plane rotation in  $[-10^\circ, 10^\circ]$ , source-to-isocenter distance in [600 mm, 900 mm], and translation in x, y axes following a Gaussian distribution  $N(\mu = 0 \text{ mm}, \sigma = 10 \text{ mm})$ . We created DeepDRR synthetic images by projecting randomly selected hip CT scan from the 20 New Mexico Decedent Image Database CT scans used for hip imaging together with the CM model, which include 28,000 for training and 2,800 for validation. Ground truth segmentation and landmark labels were projected following each simulation geometry.

The network training details are in Section. A.0.3, and strong domain ran-

## CHAPTER 3. SYNTHEX

domization was applied (Section. A.0.1). The network was trained for 10 epochs and the best validation model was selected for evaluation. The performance was evaluated on 264 real CM X-ray images with manual ground truth label annotations. During evaluation, a threshold of 0.5 was used for segmentation and the landmark prediction was selected using the highest heatmap response location. The network training and evaluation routines are the same for the *Real2Real* 5-fold training and testing.

### 3.6.3 Benchmark Hip Imaging Investigation

For every X-ray image, ground truth X-ray camera pose relative to the CT scan were estimated using an automatic intensity-based 2D/3D registration of the pelvis and both femurs [59]. Every CT was annotated with segmentation of anatomical structures and anatomical landmark locations defined in Fig. 3.2 (A). 2D labels for every X-ray image were then generated automatically by forward projecting the reference 3D annotations using the corresponding ground truth C-arm pose.

We generated synthetic data using three DRR simulators: Naïve DRR, xreg DRR and DeepDRR. Naïve DRR generation amounts to simple ray-casting and does not consider any imaging physics. This amounts to the assumption of a mono-energetic source, single material objects, and no image corruption, e. g.,

## CHAPTER 3. SYNTHEX

due to noise or scattering. Heuristic simulation (xreg DRR<sup>1</sup>) performs a linear thresholding of the CT Hounsfield Units (HU) to differentiate materials between air and anatomy prior to ray-casting. While this results in a more realistic appearance of the resulting DRRs, in that the tissue contrast is increased, the effect does not model imaging physical effects. Realistic simulation (Deep-DRR) simulates imaging physics by considering the full spectrum of the X-ray source, and relies on machine learning for material decomposition and scatter estimation. It also considers both signal dependent noise as well as readout noise together with detector saturation.

## 3.7 Model and Evaluation Paradigm

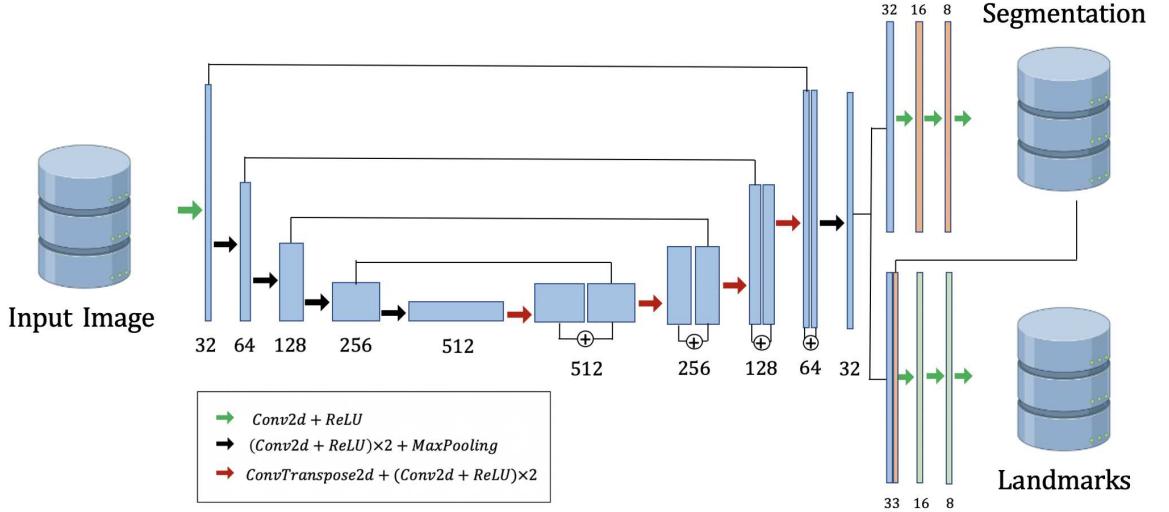
Because the goal of our experiments is to demonstrate convincing *Sim2Real* performance, we rely on a well-established backbone network architecture, namely the U-Net [70], for all tasks. Segmentation networks for all clinical applications are trained to minimize the Dice loss ( $L_{seg}$ ) [71], which evaluates the overlap between predicted and ground truth segmentation labels. For hip image analysis and surgical tool detection, we adjust the U-Net architecture as shown in Fig. 3.4 to concurrently estimate landmark locations. Reference landmark locations are represented as symmetric Gaussian distributions centered on the true landmark locations (zero when landmark is invisible). This

---

<sup>1</sup><https://github.com/rg2/xreg>

## CHAPTER 3. SYNTHEX

additional prediction target is penalized using ( $L_{ld}$ ), the mean squared error between network prediction and reference landmark heatmap.



**Figure 3.4:** U-net based concurrent segmentation and landmark detection network architecture for multi-task learning.

For evaluation purposes, we report the landmark accuracy as the  $l_2$  distance between predicted and ground truth landmark positions. Further, we use the Dice score to quantitatively assess segmentation quality for hip imaging and surgical tool detection.

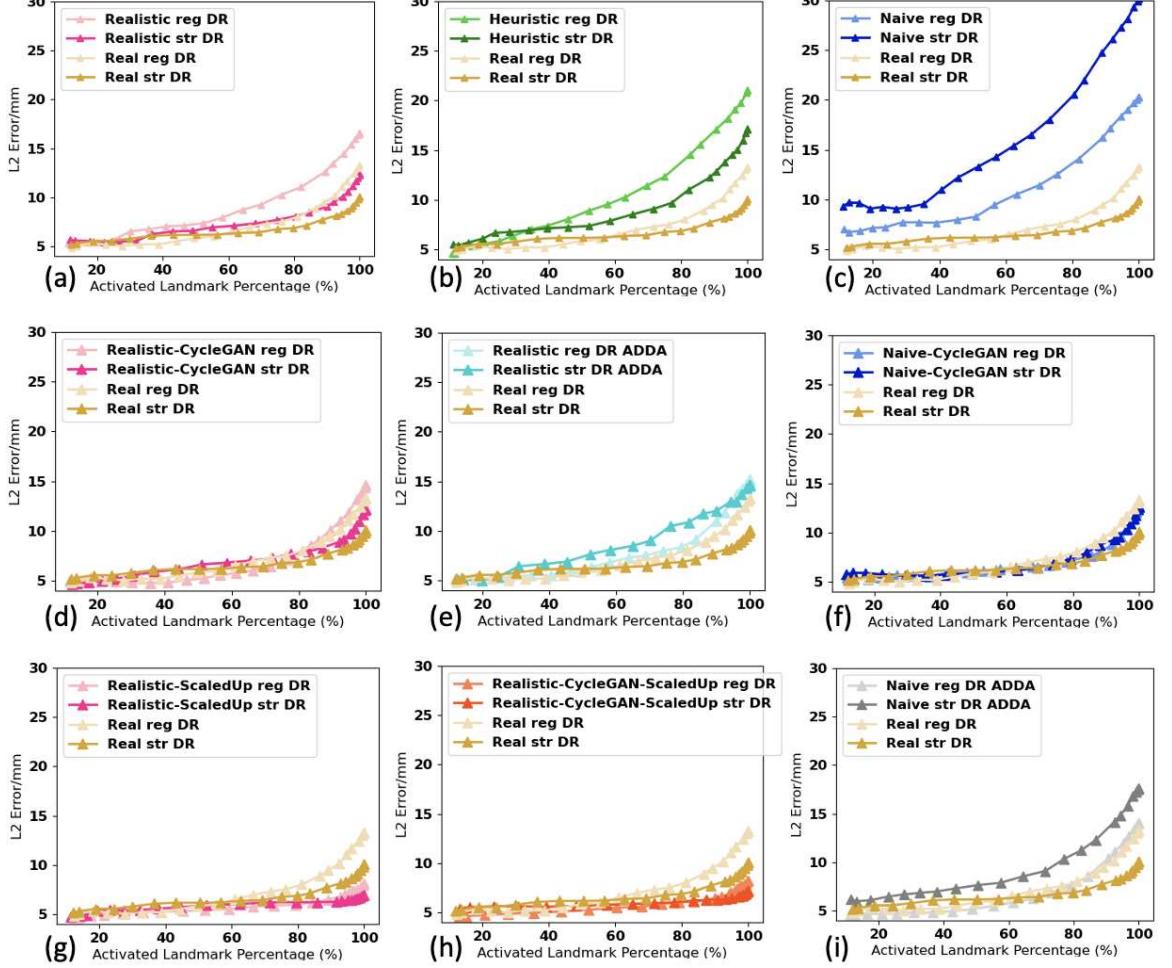
For all three tasks, we report both *Sim2Real* and *Real2Real* performances. The *Sim2Real* performance was computed on all testing real X-ray data. The *Real2Real* experiments were conducted using k-fold cross-validation, and we report the performance as an average of all testing folds. For the hip imaging benchmark studies, we further carefully designed the evaluation paradigm using leave-one-specimen-out paradigm. For each experiment, the training and

## CHAPTER 3. SYNTHEX

validation data consisted of all labeled images from all but one specimen while all labeled images from the remaining specimen were used as test data. The same data split was strictly preserved also for training of domain adaptation methods to avoid leakage and optimistic bias. On the scaled-up dataset, we used all synthetic images for training and evaluated all real data in the benchmark dataset.

A specially designed assessment curvature plot is used for reporting pelvic landmark detection performance. This way of measuring landmark detection performance provides detailed information on the two desirable attributes of such an algorithm: 1) completeness and 2) precision of detected landmarks. Landmarks are considered valid (activated) if their heatmap prediction is higher than a confidence threshold ( $\phi$ ). The mean landmark detection error ( $e^{ld}$ ) is reported as the average error over all activated landmarks. The ratio ( $p$ ) of the activated landmarks over all landmarks is a function of  $\phi$ . Thus, we created plots to demonstrate the relationship between  $e^{ld}$  and  $p$ , which shows the change of the error as we lower the threshold to activate more landmarks. Ideally, we would like a model to have a 0.0 mm error with a 100 % activation percentage, corresponding to a measurement in the bottom right corner of the plots in Fig. 3.5. We selected the activation percentage of 90% to report the numeric results for all ablation study methods in Table. 3.4.

## CHAPTER 3. SYNTHEX



**Figure 3.5:** Plots of average landmark detection error curves with respect to activated landmark percentage. *Real2Real* performance on the controlled dataset is shown in gold. An ideal curve should approach the bottom right corner: all landmarks detected with perfect localization. Each subplot compares the baseline *Real2Real* performance curve to varied *Sim2Real* methods that are evaluated on the same real data test set. The *Sim2Real* technique of the specific method is identified in the upper left corner of each subplot. Domain names followed by “CycleGAN” mean the training data are generated using CycleGAN trained between the specific image domain and the real image domain. “reg DR” and “str DR” refer to regular domain randomization and strong domain randomization, respectively. “ADDA” refers to adversarial discriminative domain adaptation. (a)-(c) present performance comparison of methods trained on precisely matched datasets. (d)-(f)(i) further evaluates the added effect of using domain adaptation techniques again using precisely matched datasets. (g),(h) demonstrate improvements in *Sim2Real* performance on the same real data test set when a larger, scaled-up synthetic training set is used. All the results correspond to an input image size of  $360 \times 360$  px.

## 3.8 Experiments and Results

We find that across both clinical tasks, namely hip imaging and surgical robotic tool detection, models trained using the SyntheX *Sim2Real* model transfer paradigm when evaluated on real data perform comparably to or even better than models trained directly on real data. This finding suggests that SyntheX, i. e., the realistic simulation of X-ray images from CT combined with domain randomization, is a feasible cost- and time-effective, and valuable approach to the development of learning-based X-ray image analysis algorithms that preserve performance during deployment on real data.

### 3.8.1 Hip Imaging

We present the multi-task detection results of hip imaging on images with  $360 \times 360$  px in Table. 3.1. Both landmark detection and anatomical structure segmentation performance achieved using SyntheX *Sim2Real* model transfer are superior to those of *Real2Real* when considering averaged metrics. The *Sim2Real* predictions are more stable with respect to their standard deviation: landmark error as 6.29 mm, Dice score as 0.221, compared to *Real2Real* with 15.30 mm and 0.248, respectively. We attribute this improvement to the flexibility of the SyntheX approach, providing the possibility of simulating a richer spectrum of image appearances from more hip CT samples and varied X-ray

### CHAPTER 3. SYNTHEX

geometries as compared to the limited data sourced from complex real-world experiments.

**Table 3.1:** Average performance metrics for hip imaging as a mean of 5-fold individual testing on 366 real hip X-ray images.

	Landmark Error (mm)	Dice Score
<i>Sim2Real</i>	$6.29 \pm 6.29$	$0.818 \pm 0.221$
<i>Real2Real</i>	$8.15 \pm 15.30$	$0.759 \pm 0.248$
P value	0.01590	0.00036

The p value is reported from T-Test score. The result is significant at  $p < 0.05$ .

### 3.8.2 Surgical Robotic Tool Detection

Results of the surgical tool detection task are summarized in Table. 3.2. The landmark detection errors of *Sim2Real* and *Real2Real* are comparable with a mean localization accuracy of 2.13 mm and 1.90 mm, respectively. However, the standard deviation of *Sim2Real* error is substantially smaller: 2.27 mm versus 5.49 mm. Further, with respect to segmentation Dice score, *Sim2Real* outperforms *Real2Real* by a large margin achieving a Dice score of 0.860 compared to 0.406, respectively. Overall, the results suggest that SyntheX is a viable approach to developing deep neural networks for this task, especially when the

## CHAPTER 3. SYNTHEX

robotic hardware is in the prototypic stages.

**Table 3.2:** Average performance metrics for surgical tool detection as a mean of 5-fold individual testing on 264 real X-ray images of the continuum manipulator.

	Landmark Error (mm)	Dice Score
<i>Sim2Real</i>	$2.13 \pm 2.27$	$0.860 \pm 0.115$
<i>Real2Real</i>	$1.90 \pm 5.49$	$0.406 \pm 0.194$

### 3.8.3 Sim2Real Benchmark Findings

Based on our precisely controlled hip imaging ablation studies including comparisons of 1) simulation environment, 2) domain randomization and domain adaptation effects, and 3) image resolution, we observed that training using realistic simulation with strong domain randomization performs on a par with models trained on real data or models trained on synthetic data but with domain adaptation, yet, *does not require any real data at training time*. Training using realistic simulation consistently outperformed naïve or heuristic simulations. Training using scaled-up realistic simulation data with domain randomization achieved the best performance on this task, even outperforming real data-trained models due to the effectiveness of larger training data. Thus, realistic simulation of X-ray images from CT combined with do-

## CHAPTER 3. SYNTHEX

main randomization, which we refer to as the SyntheX model transfer concept, is a most promising approach to catalyze learning-based X-ray image analysis. The specially-designed landmark detection error metric plot, which summarizes the results across all ablations on images with  $360 \times 360$  px is shown in Fig. 3.5. We plotted the *Real2Real* performance using golden curves as a baseline comparison to all the other ablation methods.

### 3.8.3.1 The Effect of Domain Randomization

Across all experiments, we observed that networks trained with strong domain randomization consistently achieved better performance than those with regular domain randomization. This is expected because strong domain randomization introduces more drastic augmentations, which samples a much wider spectrum of possible image appearance and promotes the discovery of more robust features that are less prone to overfitting. The only exception is the training on naïvely simulated images, where training with strong domain randomization results in much worse performance compared to regular domain randomization. We attribute this to the fact that the contrast of bony structures, which are most informative for the task considered here, are already much less pronounced in naïve simulations. Strong domain randomization then further increases problem complexity, to the point where performance deteriorates.

## CHAPTER 3. SYNTHEX

From Fig. 3.5(a)-(c), we see that realistic simulation (DeepDRR) outperforms all other X-ray simulation paradigms in both regular domain randomization and strong domain randomization settings. Realistic simulation trained using strong domain randomization even outperforms *Real2Real* with regular domain randomization. Since our experiments were precisely controlled and the only difference between the two scenarios is the image appearance due to varied stimulation paradigms in the training set, this result supports the hypothesis that realistic simulation of X-rays using DeepDRR performs best for model transfer to real data.

### 3.8.3.2 The Effect of Domain Adaptation

From Fig. 3.5 (d)(f), we observe that both Realistic-CycleGAN and Naïve-CycleGAN achieve comparable performance to *Real2Real*. This means that images generated from synthetic images via CycleGAN have a similar appearance, despite the synthetic training domains being different. The improvements over training purely on the respective synthetic domains (Fig. 3.5 (a)(c)) confirm that CycleGAN is useful for domain generalization. ADDA training also improves the performance over a non-adapted transfer but does not perform at the level of CycleGAN models. Interestingly, ADDA with strong DR shows deteriorated performance than regular DR (Fig. 3.5 (e)(i)). This is because the drastic and random appearance changes due to DR complicate do-

## CHAPTER 3. SYNTHEX

main discrimination, which in turn has adverse effects on overall model performance.

### 3.8.3.3 Scaling Up the Training Data

We selected the best-performing methods from the above domain randomization and domain adaptation ablations on the controlled dataset. These methods were realistic simulation with domain randomization and CycleGAN training based on realistic simulation, respectively, and trained on the scaled-up dataset, which contains a much larger variety of anatomical shapes and imaging geometry, i. e., synthetic C-arm poses.

With more training data and geometric variety, we found that all scaled-up experiments outperform the *Real2Real* baseline on the benchmark dataset, which are shown in Fig. 3.5 (g)(h). With 90 % of the landmarks activated, the model trained with strong domain randomization on realistically synthesized but large data (SyntheX, as reported above) achieved a mean landmark distance error of  $6.29 \pm 6.29$  mm. This is significantly better than the *Real2Real* baseline ( $p=0.032$ ), despite no real data being used for model training. This result suggests that training with strong domain randomization and/or adaptation on large-scale, realistically synthesized data is a feasible alternative to training on real data. Training on large-scale data processing by CycleGAN achieved comparable performance ( $6.59 \pm 7.25$  mm) as pure realistic simulation

## CHAPTER 3. SYNTHEX

with domain randomization but comes at the disadvantage that real data with sufficient variability must be available at training time to enable CycleGAN training.

**Table 3.3:** Segmentation Dice Score. The Dice score ranges from 0 to 1, with larger values corresponding to better segmentation performance.

	180x180		360x360		480x480	
	regular DR	strong DR	regular DR	strong DR	regular DR	strong DR
RealXray	0.775 ± 0.235	0.784 ± 0.214	0.783 ± 0.232	0.759 ± 0.248	0.739 ± 0.266	0.751 ± 0.265
Realistic	0.730 ± 0.240	0.787 ± 0.211	0.751 ± 0.241	0.760 ± 0.250	0.720 ± 0.256	0.700 ± 0.279
Heuristic	0.669 ± 0.273	0.737 ± 0.249	0.683 ± 0.265	0.682 ± 0.286	0.655 ± 0.277	0.668 ± 0.298
Naïve	0.689 ± 0.256	0.680 ± 0.278	0.687 ± 0.266	0.572 ± 0.309	0.653 ± 0.278	0.578 ± 0.305
Realistic-Cyc	0.778 ± 0.217	0.778 ± 0.220	0.760 ± 0.248	0.733 ± 0.267	0.741 ± 0.255	0.688 ± 0.291
Naïve-Cyc	0.784 ± 0.198	0.750 ± 0.230	0.739 ± 0.254	0.736 ± 0.258	0.726 ± 0.254	0.673 ± 0.292
Realistic-ADDA	0.767 ± 0.224	0.754 ± 0.231	0.726 ± 0.292	0.731 ± 0.256	0.704 ± 0.279	0.727 ± 0.256
Naïve-ADDA	0.491 ± 0.405	0.678 ± 0.266	0.693 ± 0.297	0.662 ± 0.265	0.693 ± 0.273	0.592 ± 0.306
Realistic-Scaled	<b>0.857 ± 0.184</b>	<b>0.853 ± 0.179</b>	<b>0.838 ± 0.221</b>	<b>0.818 ± 0.221</b>	0.783 ± 0.262	<b>0.823 ± 0.221</b>
Realistic-Cyc-Scaled	0.822 ± 0.213	0.794 ± 0.232	0.824 ± 0.225	0.789 ± 0.240	<b>0.789 ± 0.241</b>	0.812 ± 0.227
P value	< 0.00001	< 0.00001	0.00054	0.00036	0.00394	0.00004

Note: The best performance results are bolded. DR is short for domain randomization.  
Cyc is short for CycleGAN. “Scaled” means training on the scaled-up dataset.

ADDA refers to adversarial discriminative domain adaptation.

The p value is reported from T-Test score comparing the best performance and the Real2Real baseline results.  
The result is significant at  $p < 0.05$ .

### CHAPTER 3. SYNTHEX

**Table 3.4:** Landmark Detection Errors (mm) at 90% Activation Percentage. Lower values are better.

	180x180		360x360		480x480	
	regular DR	strong DR	regular DR	strong DR	regular DR	strong DR
RealXray	9.98 ± 22.58	7.78 ± 11.94	8.93 ± 19.76	8.15 ± 15.30	8.98 ± 21.38	7.59 ± 16.12
Realistic	14.33 ± 32.61	8.41 ± 14.47	12.62 ± 27.68	9.05 ± 19.37	14.96 ± 34.45	13.06 ± 26.52
Heuristic	20.84 ± 44.16	10.69 ± 22.92	14.54 ± 32.88	12.25 ± 26.82	17.59 ± 39.55	12.85 ± 29.49
Naïve	13.12 ± 30.87	13.03 ± 21.20	16.22 ± 39.71	20.55 ± 35.92	18.53 ± 40.88	19.46 ± 37.48
Realistic-Cyc	8.65 ± 19.84	8.19 ± 13.55	8.57 ± 18.69	8.90 ± 17.47	10.78 ± 26.9	8.75 ± 17.4
Naïve-Cyc	8.56 ± 18.46	9.05 ± 10.75	7.63 ± 16.10	9.29 ± 18.73	9.14 ± 21.82	11.73 ± 25.06
Realistic-ADDA	11.19 ± 24.80	11.43 ± 25.83	11.01 ± 24.69	12.92 ± 23.76	16.42 ± 37.61	9.24 ± 17.53
Naïve-ADDA	7.90 ± 17.56	11.84 ± 25.63	10.42 ± 25.05	14.17 ± 30.89	16.62 ± 40.12	22.88 ± 41.53
Realistic-Scaled	<b>5.91 ± 8.43</b>	<b>6.06 ± 7.10</b>	<b>6.80 ± 9.53</b>	<b>6.29 ± 6.29</b>	8.12 ± 19.35	5.99 ± 12.19
Realistic-Cyc-Scaled	6.79 ± 9.70	6.87 ± 9.58	6.87 ± 13.19	6.59 ± 7.25	<b>6.43 ± 13.67</b>	<b>5.52 ± 4.85</b>
P value	0.00065	0.00907	0.03186	0.01590	0.02750	0.00947

Note: The best performance result is bolded. The abbreviations follow Table 3.3.  
 The p value is reported from T-Test score comparing the best performance and the Real2Real baseline results.  
 The result is significant at  $p < 0.05$ .

## 3.9 Discussion

We present the general use cases of SyntheX on purely bony anatomy (the hip) and a metallic artificial surgical tool. Our experiments on two varied clinical tasks demonstrate that the performance of models trained using SyntheX - on real data - meets or exceeds the performance of real data-trained models. We show that generating realistic synthetic data is a viable resource for developing machine learning models compared to collecting largely annotated real clinical data.

The hip imaging ablation experiments reliably quantify the effect of the domain gap on real data performance for varied *Sim2Real* model transfer approaches. This is because all aleatoric factors that usually confound such experiments are precisely controlled for, with alterations to image appearances due to the varied image simulation paradigms being the only source of mismatch. The aleatoric factors we controlled for include anatomy, imaging geometries, ground truth labels, network architectures, and hyperparameters. In particular, the viewpoints and 3D scenes recreated in simulation were identical to the real images, which to our knowledge has not yet been achieved. From these results, we draw the following conclusions:

- Physics-based, realistic simulation of training data using the DeepDRR framework results in models that generalize better to the real data do-

## CHAPTER 3. SYNTHEX

main compared to models trained on less realistic, i. e., naïve or heuristic simulation paradigms. This suggests, not surprisingly, that matching the real image domain as closely as possible directly benefits generalization performance.

- Realistic simulation combined with strong domain randomization (SyntheX) performs on a par with both, the best domain adaptation method (CycleGAN with domain randomization) as well as real data training. However, because SyntheX does not require any real data at training time, this paradigm has clear advantages over domain adaptation. Specifically, it saves the effort to acquire real data early in development or design additional machine learning architectures that perform adaptation. This makes SyntheX particularly appealing for the development of novel instruments or robotic components, real images which can simply not be acquired early during conceptualization.

Realistic simulation using DeepDRR is as computationally efficient as naïve simulation, both of which are orders of magnitude faster than Monte Carlo simulation [66]. Further, realistic simulation using DeepDRR brings substantial benefits in regards to *Sim2Real* performance and self-contained data generation and training. These findings are very encouraging and strongly support the hypothesis that training on synthetic radiographs simulated from 3D CT is a viable alternative to real data training, or at a minimum, a strong candidate

## CHAPTER 3. SYNTHEX

for pre-training.

Compared to acquiring real patient data, generating large-scale simulation data is more flexible, timely efficient, low-cost, and avoids privacy concerns. For the hip image analysis use case, we performed experiments based on 10,000 synthetic images from 20 hip CT scans. Training with realistic simulation and strong domain randomization outperformed *Real2Real* training at the 90 % activation level but generally improved performance as seen by a flattened activation vs. error curve (Fig.3.5 (g)). The performance of training with CycleGAN with larger datasets was similar. These findings suggest that scaling-up data for training is an effective strategy to improve performance both in- and outside of the training domain. Scaling up training data is costly or impossible in real settings, but in comparison is easily possible using data synthesis. Having access to more varied data samples during training helps the network parameter optimization find a more stable solution, that also transfers better.

Interestingly, although domain adaptation techniques (CycleGAN and ADDA) have access to data in the real domain, these methods outperformed domain generalization techniques (here, domain randomization) only by a small margin in the controlled study. The performance of ADDA training heavily depends on the choices of additional hyperparameters, such as the design of the discriminator, the number of training cycles between task and discriminator network updates, and learning rates, among others. Thus, it is non-trivial to find the

## CHAPTER 3. SYNTHEX

best training settings, and these settings are unlikely to apply to other tasks. Because CycleGAN performs image-to-image translation, a complicated task, it requires sufficient and sufficiently diverse data in the real domain to avoid overfitting. Further, using CycleGAN requires an additional training step of a large model, which is memory intensive and generally requires long training time. In certain cases, CycleGAN models could also introduce undesired effects. Jaihyun et al. find that performance of CycleGAN is highly dependent on the dataset, potentially resulting in unrealistic images with less information content than the original images [72]. Finally, because real domain data is being used in both domain adaptation paradigms, adjustments to the real target domain, e.g., use of a different C-arm X-ray imaging device or design changes to surgical hardware, may require de novo acquisition of real data and re-training of the models. In contrast, SyntheX resembles a plug-and-play module, to be integrated into any learning-based medical imaging tasks, which is easy to set up and use. Similar to multi-scale modeling [73] and in silico virtual clinical trials [74, 75], SyntheX has the potential to envision, implement, and virtually deploy solutions for image-guided procedures and evaluate their potential utility and adequacy. This makes SyntheX a promising tool that may replicate traditional development workflows solely using computational tools.

Despite the promising outlook, our study has several limitations. First, while the real X-ray and CT data of cadaveric specimens used for the hip imag-

## CHAPTER 3. SYNTHEX

ing and robotic tool segmentation task is of respectable size for this type of application, it is small compared to some dataset sizes in general computer vision applications. However, the effort, facilities, time, and therefore, costs required to acquire and annotate a dataset of even this size are substantial due to the nature of the data. To counter this shortcoming, we performed leave-one-subject-out cross validations for all experiments and found that the clinical task of anatomical segmentation and landmark detection could be solved at an acceptable level that is comparable to previous studies on larger data [58, 76].

Second, the performance we report is limited by the quality of the CT and annotations. The spatial resolution of CT scans (between 0.5 to 1.0 mm in hip imaging and surgical robot tool segmentation, imposed a limitation on the resolution that can achieve in 2D simulation. Pixel sizes of conventional detectors are as small as 0.2 mm, smaller than the highest resolution scenario considered here. However, contemporary computer vision algorithms for image analysis tasks have only considered downsampled images in the ranges described here. The realism of simulation can be improved with higher quality CT scans, super-resolution techniques, and advanced modeling techniques to more realistically represent anatomy at higher resolutions.

## 3.10 Conclusion

In this Chapter, we demonstrated that realistic simulation of image formation from human models combined with domain generalization or adaptation techniques is a viable alternative to large-scale real data collection. We demonstrate its utility in two different clinical tasks, namely hip imaging and surgical robotic tool detection. Based on controlled experiments on a pelvic X-ray dataset, which is precisely reproduced in varied synthetic domains, we quantified the effect of simulation realism and domain adaptation/generalization techniques on *Sim2Real* transfer performance. We found promising *Sim2Real* performance of all models trained on realistically simulated data. The specific combination of training on the realistic synthesis and strong domain randomization, which we refer to as SyntheX, is particularly promising. SyntheX-trained models perform on a par with real-trained models, making the realistic simulation of X-ray-based clinical workflows and procedures a viable alternative or complement to real data acquisition. Because SyntheX does not require real data at training time, it is particularly promising to develop machine learning models for novel clinical workflows or devices, including surgical robotics, before these solutions physically exist.

## 3.11 Acknowledgement

I appreciate Mr. Yicheng Hu and Mr. Benjamin Killeen's efforts for their help processing the data and performing the controlled Sim2Real experiments. Sincere thanks to Dr. Robert Grupp for his efforts in collecting and registering the cadaveric pelvis X-ray data. I also thank Drs. Mathias Unberath, Russell Taylor, and Mehran Armand for their valuable suggestions and encouragement on this study. Special thanks to Mrs. Courtney McQueen from the Writing Center at Johns Hopkins University for improving the writing of this work.

# **Chapter 4**

## **Pose Estimation of Bone**

### **Anatomy**

#### **4.1 Introduction**

This Chapter introduces fiducial-free methods to estimate pose transformation of bone anatomy using multiple view C-arm fluoroscopic images. The proposed methods use purely image intensity information for registration without the need to attach a bone pin fiducial to close the calibration loop. The pose transformation is reported from the C-arm source coordinate frame with respect to the pre-operative computed tomography (CT) scan of the target bone anatomy. We use this pose estimation to resolve the relative transformation between the bone and robotic surgical tool in Chapter. 6, which is then integrated

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

into the pre-operative plan to navigate the robotic surgical tool.

We present image intensity-based 2D/3D registration methods of the pelvis, proximal femur, and lumbar spinal vertebrae, which are target anatomies in the investigated robot-assisted orthopedic applications in this dissertation (Chapter. 7). Accurate pose estimation of the proximal femur is challenging due to fewer distinct features present in the 2D projection domain than the other anatomies, such as the pelvis. We propose using multi-view pelvis 2D/3D registration to initialize and constrain the femur registration, which achieves clinically acceptable pose estimation performance for femoroplasty. Pose estimation of the spine vertebrae has additional challenges compared to the hip because the vertebrae are smaller and composed of multiple components, which deforms the spine's shape. For the spine, we propose a *vertebra-by-vertebra* registration method, which follows a coarse-to-fine manner to compensate for the intra-operative spine vertebrae shape difference from the pre-operative CT scan. This method yields satisfactory results when used to navigate the robotic needle injection device in our clinical study for transforaminal lumbar epidural injections.

We conducted intensive simulation and *ex vivo* cadaveric studies to validate our pose estimation methods. In simulation, synthetic X-ray images with known richly sampled anatomy geometries were generated by creating DRR images from the CT scans. We performed simulated registrations under

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

variant conditions, including randomized initialization poses, C-arm geometry views, and using single-view images versus multi-view images. Our approaches were also tested in clinical experiments using cadaveric specimens for each application. In this Chapter, we mainly present simulation and cadaver experiments' results to verify the registration accuracy. Our efforts on system-level clinical experiments and results are shown in Chapter 7.

## 4.2 Contributions

The contributions of this Chapter are:

- We develop an initialization and regularization strategy for intensity-based 2D/3D registration of the femur, which uses the pelvis registration as a reference. This strategy eliminates the need of feature extraction or manual annotations on the femur 3D model.
- We propose a coarse-to-fine, vertebra-by-vertebra 2D/3D registration method to address the intra-operative spine shape deformation.
- We evaluate the proposed bone anatomy registration methods in simulation experiments under various conditions and cadaveric experiments.

We extended the registration framework developed by Grupp et al. [77] for pelvis fragment pose estimation to different bone anatomies. The proposed

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

methods are fiducial-free and solve the unique challenges of each anatomy’s registration problem. Our results suggest that the pose estimation accuracy meets clinical requirements. From the system point of view, this work estimates the intra-operative pose of bone anatomy using C-arm fluoroscopic images, which is essential to estimating the tool-to-tissue relationship for navigation.

The 2D/3D registration method of the femur was presented at the 2020 SPIE Medical Imaging Conference [29], and is published in the July 2020 issue of IEEE Transactions on Medical Robotics and Bionics [3]. The 2D/3D registration method of the spine vertebrae is published in the IEEE Transactions on Medical Robotics and Bionics [30], and April 2022 issue of Global Spine Journal [78].

### 4.3 Related Work

Intra-operative shape and pose estimations of bony structures using X-ray images have been intensively studied in the literature. A great amount of work investigated on building the anatomy’s 3D shape model from calibrated X-ray images using statistical atlases to eliminate the requirement of taking pre-operative CT/MRI scans [79–81]. In this dissertation, we assume that the 3D anatomy model can be extracted from patient-specific pre-operative CT scans.

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

Therefore, we do not look into using statistical shape models for registration. Given pre-operative 3D models, there have been rich efforts on registering the femur and spine anatomies. Femur registration is challenging due to its lack of distinct features in 2D domain, and the challenges of spine registration arise from its overall shape deformation. Depending on the used image information, the registration approaches can be generally classified into feature-based and intensity-based methods.

Feature-based 2D/3D registration methods seek to extract prominent corresponding geometric features from the 3D anatomy models and 2D X-ray images for pose recovery. Shape matching methods were initially tested on an individual vertebrae [82]. The improved surface contour features were used to register the proximal femur [81, 83]. These methods were promising but the registration was sensitive to the contour extraction accuracy [84] and initial corresponding point set estimation. A number of different optimization frameworks have been studied for femur registration [85–87], but these algorithms require either control points or contours for initialization setup with additional annotations. Moreover, the femur X-ray images used for the validation of the mentioned frameworks did not include surrounding anatomies and soft tissues, and generally assumed that the C-arm views were calibrated, which is not common for a mobile C-arm. In recent years, anatomical landmark features have become popular for 2D/3D registration [58, 76, 88, 89]. However, it is challeng-

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

ing to detect precise landmark locations in X-ray images, which limits the pose estimation accuracy [58]. A common limitation for feature-based methods is that the features must be specifically selected and extracted for each anatomy of interest and clearly visible in the target images.

Intensity-based registration methods use purely image intensity information for pose estimation without defining features. Several works applied intensity-based 2D/3D registration to the spine [27, 90, 91]. These work aimed at localizing the vertebrae in a single X-ray image and concerned the target point reprojection errors, which was sufficient to differentiate the vertebrae levels. However, the accuracy of 3D vertebrae poses using intensity-based registration have not been investigated. Grupp et al. have studied image intensity-based 2D/3D registration method for pose estimation of periacetabular osteotomy fragments using multiple view X-ray images [18, 77, 92]. Because the pelvis is abundant in features and can be accurately registered, they treated the pelvis itself as a “fiducial” object to estimate the multi-view C-arm geometry for fragment registration. They have reported a clinically acceptable accuracy for the fragment pose and shape estimation.

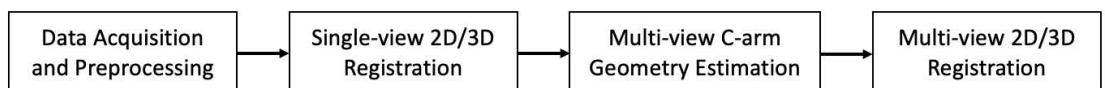
We extend the work of Grupp et al. [77] on intensity-based 2D/3D registration to the femur and spine anatomies. The method details are presented in the next section.

## 4.4 Methods

Fig. 4.1 presents the general registration workflow. After data acquisition and preprocessing, coarse pose estimation is achieved by running single-view 2D/3D registration. Multi-view C-arm projection geometry is estimated using either anatomy or a robotic device as “fiducial” object. Multi-view 2D/3D registration is then performed to achieve precise bone pose estimation. We introduce algorithmic details for femur and spine vertebrae registration in the following sections.

### 4.4.1 Data Acquisition and Preprocessing

Pre-operative CT scans of the target bone anatomy are acquired. The CT scans are resampled to 1.0 mm isotropic voxel spacing using linear interpolation. Segmentations of the bone are performed using CT scans. Bone anatomies within the hip, including pelvis and femur, are segmented using an automatic method described in [61] with manual touch-up refinement. The spine vertebrae are segmented using a coarse to fine vertebrae localization and segmentation method [93].



**Figure 4.1:** General workflow for pose estimation of bone anatomy.

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

We manually annotate the reference frame origin of each bone object, which serves as a rotation center during 2D/3D registration pose optimization. The frame origins for pelvis and spine vertebrae are selected as volume centers, and the origin for the proximal femur is chosen to be the femoral head center. We also annotate distinct anatomical landmarks to initialize 2D/3D registration. The meaningful hip anatomical landmarks consist of left and right anterior superior iliac spine (ASIS), center of femoral head (FH), superior pubic symphysis (SPS), inferior pubic symphysis (IPS), medial obturator foramen (MOF), inferior obturator foramen (IOF), and the greater sciatic notch (GSN). We select points on the extreme corners of the spinous and transverse processes to be vertebrae anatomical landmarks.

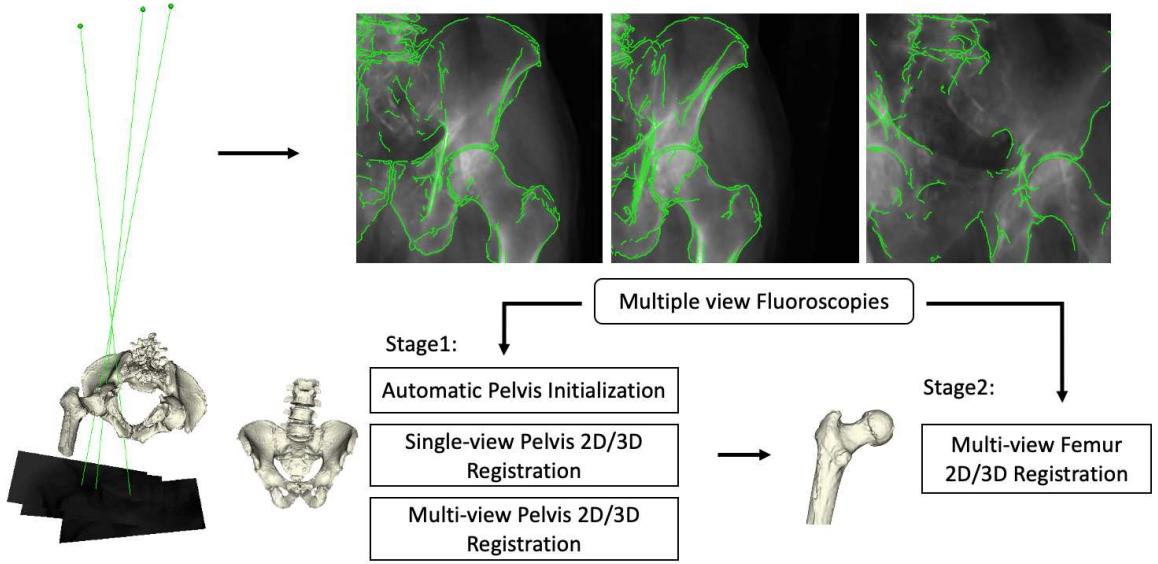
### 4.4.2 Femur Registration

Most of the reported work in the literature treats the proximal femur as an independent object to register, excluding the surrounding anatomies. In fact, the motion of the femur is articulated by the pelvis, and the articulated pelvis and femur compose a multi-component hip anatomy. Chintalapani et al. proposed a hierarchical 2D/3D registration framework using multi-component statistical hip atlases, which explicitly modeled the relationship between the pelvis and femur [79]. The experiment results suggested improvements of femur registration accuracy using multi-component atlas opposed to a single

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

component atlas. However, the focus of this study was the performance of statistical shape modeling. The multiple X-ray views were predefined and the registration was manually initialized.

In our work, We extend the idea of multi-component hip registration by treating the pelvis as a fiducial object to estimate the multi-view C-arm geometry, and initialize and constraint the rigid femur 2D/3D registration. We intend to take X-ray images which capture both the pelvis and the proximal femur. We first register the pelvis using multi-view intensity-based 2D/3D registration [18]. Because movement of the proximal femur head is covered by the pelvis acetabulum, we choose the femoral head center landmark as the femur's rotation center. The femur registration is initialized by the pelvis pose estimation and the optimization search space is constrained by the rotation center position. An overview of the femur registration pipeline is shown in Fig. 4.2. In stage1, the pelvis is automatically initialized by detecting anatomical landmarks in 2D X-ray image, and the pelvis is registered with a series of single-view and multi-view registrations. In stage2, the femur is initialized using the pelvis registration pose estimation and registered using multi-view fluoroscopic images. We present method details in the following sections.



**Figure 4.2:** Femur registration pipeline. Multiple view fluoroscopic projection geometries and simulation images with DRR-derived edges in green are presented.

#### 4.4.2.1 Automatic Pelvis Initialization

Intraoperatively, multiple view fluoroscopic images are acquired for registration (Fig. 4.2). In Chapter 3, we present a U-Net inspired network that performs concurrent pelvic anatomical structure segmentation and landmark detection. The landmarks are the same as introduced here in Section 4.4.1. The detection results are used here to automatically initialize pelvis registration. An initial pose estimate of the first view C-arm frame with respect to the pelvis volume ( $V^P$ ),  $(T_{Carm_0}^{V^P})_{ini}$ , is obtained by solving the PnP problem [94] using the detected 2D landmarks and their corresponding 3D anatomical landmarks. This pose estimation serves as an initialization for the following intensity-based 2D/3D registration.

#### 4.4.2.2 Intensity-based Pelvis Registration

Once the initial pose is estimated, a single-view intensity-based 2D/3D registration of the pelvis is performed. Single-view intensity-based 2D/3D registration is achieved by creating DRRs and calculating a similarity score between each DRR and the intra-operative image  $I$ . Using a pre-operative CT scan of the pelvis ( $V^P$ ), a DRR operator ( $\mathcal{P}$ ), similarity metric ( $\mathcal{S}$ ), and regularizer over plausible poses ( $\mathcal{R}$ ), the registration recovers the pelvis pose ( $T_{Carm_0}^{V^P}$ ) by solving the following optimization problem:

$$\min_{T_{Carm_0}^{V^P} \in SE(3)} \mathcal{S}\left(I, \mathcal{P}(V^P; T_{Carm_0}^{V^P})\right) + \mathcal{R}(T_{Carm_0}^{V^P}). \quad (4.1)$$

We use patch-based normalized gradient cross correlation score as the similarity metric [92]. DRR ray casting and similarity metric computation are accelerated using GPU. We downsample the 2D image four times in each dimension. The optimization is conducted using the CMAES optimization strategy [95]. The registration produces a single-view pose estimate of the pelvis,  $T_{Carm_0}^{V^P}$ .

Pose geometries of the other two C-arm views are estimated. We perform an exhaustive search starting from the first view pelvis pose estimation. The search space differs from the first view by a rotation along the C-arm orbit of  $\pm 45^\circ$  in  $0.5^\circ$  increments. The pose corresponding to the best similarity score is used to initialize single-view intensity-based registrations for the other two

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

C-arm views. To avoid local minima, both pelvis and femur are used for the exhaustive search. We keep the parameter setup the same for the other two single-view pelvis registrations.

The above single-view pelvis registration is used to produce pose estimates in three C-arm extrinsic views:  $T_{Carm_i}^{V^P}$  ( $i = 0, 1, 2$ ). Thus, the relative poses of these three C-arm views can be recovered using

$$T_{Carm_0}^{Carm_1} = T_{Carm_0}^{V^P} \cdot (T_{Carm_1}^{V^P})^{-1},$$

$$T_{Carm_0}^{Carm_2} = T_{Carm_0}^{V^P} \cdot (T_{Carm_2}^{V^P})^{-1}.$$

We then use the three-view geometry recovered from the pelvis coordinate frame to run multiple-view pelvis registration. Registration with multiple 2D views is accomplished by creating DRRs at each view and summing the similarity scores for each view [18, 96]. The multi-view registration optimizes the pelvis pose ( $T_{Carm_0}^{V^P}$ ) with 3 intraoperative views ( $I_0, I_1, I_2$ ):

$$\min_{T_{Carm_0}^{V^P} \in SE(3)} \sum_{m=0}^2 \mathcal{S}(I_m, \mathcal{P}(V^P; T_{Carm_0}^{V^P})) + \mathcal{R}(T_{Carm_0}^{V^P}). \quad (4.2)$$

The similarity metric, the downsampling rate, and the optimization algorithm remain the same as used for single view pelvis registration as described above. The multi-view registration will produce refinement of the pelvis pose estimate  $T_{Carm_0}^{V^P}$ , which is more accurate compared to single-view estimation with

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

the knowledge from multiple views.

### 4.4.2.3 Multi-view Femur Registration

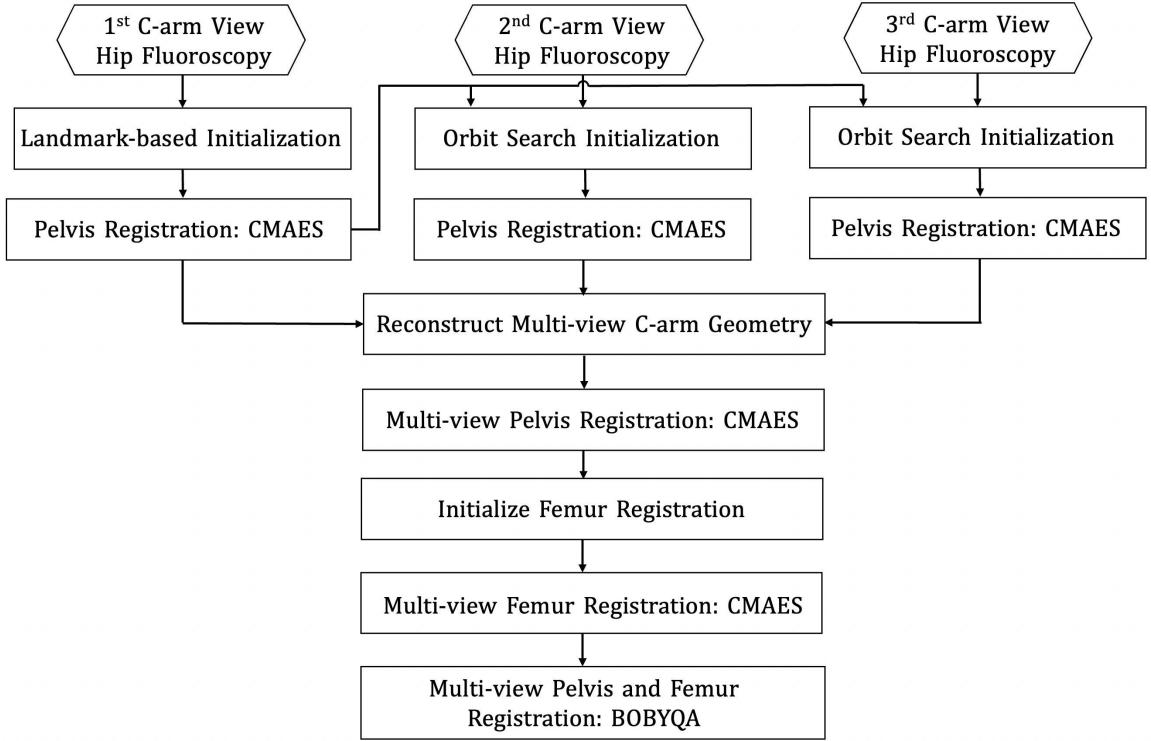
Once the pelvis is registered, we use the pelvis pose to initialize femur registration, and further constrain femur registration to be rotation only with respect to the femoral head center position. The registration shares the same formulation as in Section. 4.4.2.2, but the optimization target is changed to  $R_{Carm_0}^{V^F} \in SO(3)$ , where  $V^F$  is the femur CT scan volume. We keep the down-sampling and optimization algorithm the same. Multi-view femur registration will estimate the femur pose with respect to the first C-arm view,  $T_{Carm_0}^{V^F}$ .

After the multi-view pelvis and femur registration, we conduct a multi-view multiple object registration by jointly registering the pelvis and femur in order to perform a local search. The formulation is inspired by [18] as

$$\min_{T_{Carm_0}^{V^P}, T_{Carm_0}^{V^F} \in SE(3)} \sum_{m=0}^2 \mathcal{S}\left(I_m, \mathcal{P}(V^P, V^F; T_{Carm_0}^{V^P}, T_{Carm_0}^{V^F})\right) + \mathcal{R}(T_{Carm_0}^{V^P}, T_{Carm_0}^{V^F}) \quad (4.3)$$

The optimization strategy is the Bounded optimization by Quadratic Approximation (BOBYQA) [22]. BOBYQA is less robust to local minima as compared to CMAES, however, it runs significantly faster [18]. After the refinement, the

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY



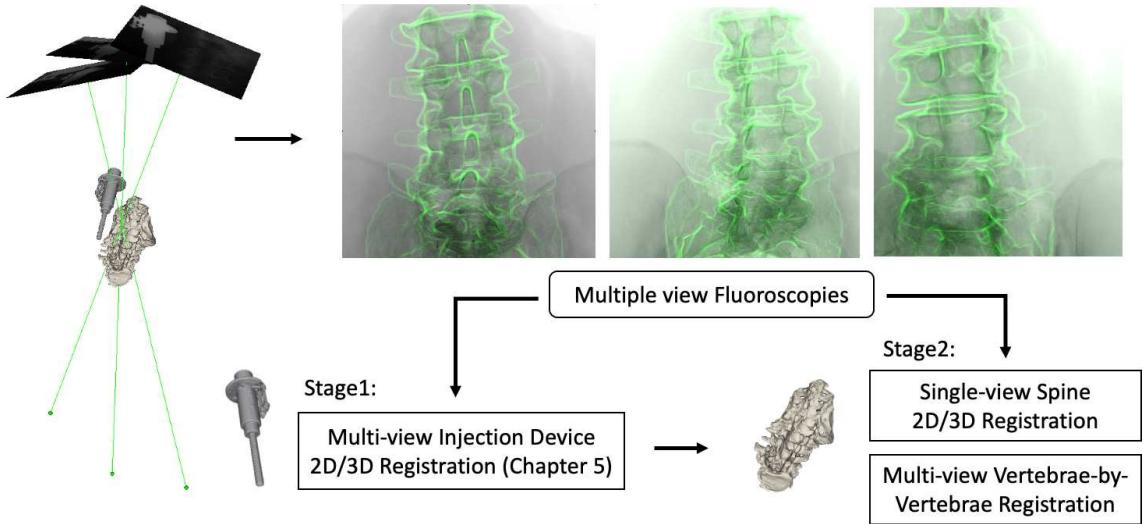
**Figure 4.3:** Workflow of pelvis and femur registration.

pose of the femur with respect to the first C-arm view source frame is estimated as  $T_{Carm_0}^{VF}$ . We provide a step-by-step registration workflow chart in Fig. 4.3.

### 4.4.3 Spine Vertebrae Registration

The spine vertebrae registration method is designed for the application of transforaminal lumbar epidural injections (TLEI). The fiducial-free multi-view image-based registration concept is similar to femur registration as described above. Multiple view fluoroscopic images are taken to capture both the spine vertebrae and a robotic injection device (Fig. 4.4). Because the spine vertebrae

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY



**Figure 4.4:** Spine registration pipeline: Multi-view fluoroscopy projection geometry of a robotic injection device and spine vertebrae is presented.

are smaller in size and composed of multiple components, the single-view spine registration is less robust than the pelvis registration, the multi-view fluoroscopic geometry is estimated using image-based registration of the injection device as a “fiducial” object. We present details of this multi-view geometry estimation method in Chapter 6. In this section, we focus on the pose estimation of the spine vertebrae, assuming that the multi-view geometry is accurately estimated.

Pose estimation of the spine has additional challenges compared to the hip because it consists of multiple small vertebrae components, which deforms its intra-operative shape. We propose a *vertebra-by-vertebra* registration method to compensate for the intra-operative spine vertebrae shape difference from the pre-operative CT scans. Intra-operative pose estimation of each individual

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

spine vertebrae is achieved in a coarse-to-fine manner. We introduce the spine registration strategies in the following sections.

### 4.4.3.1 Single-view Spine Registration

First, a single-view rigid 2D/3D registration is performed using the first C-arm view X-ray image and the rigid vertebrae segmentation from the pre-operative CT scans. Given the spine vertebrae ( $V_m^S, m \in \{1..M\}$ , where  $M$  is the total number of vertebrae for registration), X-ray image  $I_0^S$  (the first C-arm view X-ray image for spine registration), a DRR operator ( $\mathcal{P}$ ), a similarity metric ( $\mathcal{S}$ ), the single-view 2D/3D registration estimates the pose of rigid spine vertebrae ( $T_{Carm_0}^v$ ) by solving the following optimization problem:

$$\min_{T_{Carm_0}^v \in SE(3)} \mathcal{S}\left(I_0^S, \mathcal{P}\left(\sum_{m=1}^M V_m^S; T_{Carm_0}^v\right)\right). \quad (4.4)$$

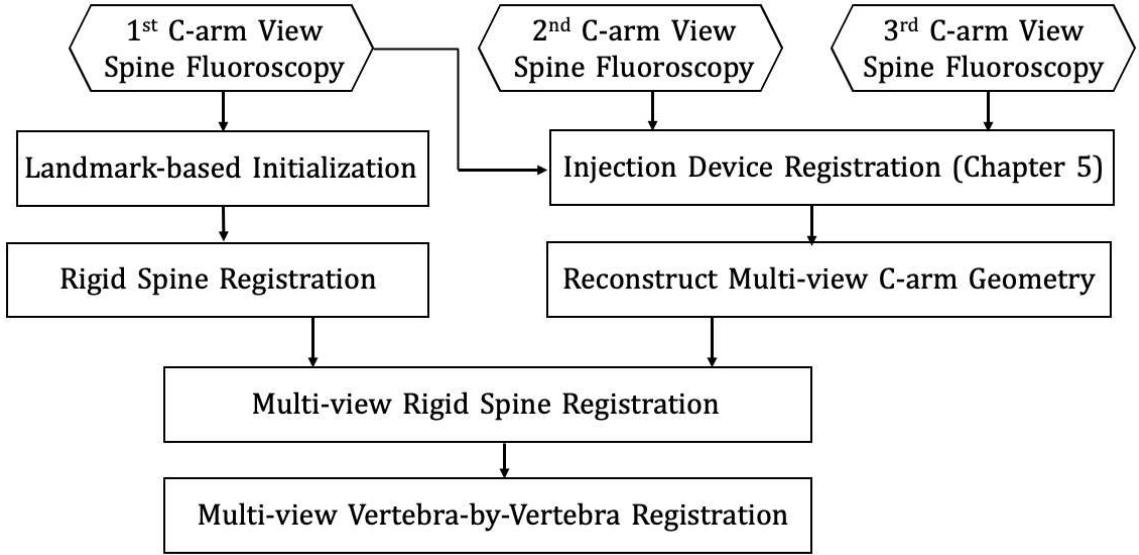
The similarity metric ( $\mathcal{S}$ ) is chosen to be patch-based normalized gradient cross correlation (Grad-NCC) [92]. The 2D X-ray image is downsampled four times in each dimension. We use the CMAES optimization strategy. Because of the shape difference and the ambiguity of single-view 2D/3D registration,  $T_{Carm_0}^v$  solved from equation 4.4 is prone to be less accurate. Precise intra-operative vertebrae pose estimation is achieved by performing multi-view *vertebra-by-vertebra* 2D/3D registration.

### 4.4.3.2 Vertebra-by-Vertebra Registration

The pose of each individual vertebra is optimized independently under multi-view C-arm geometries. The multiple C-arm geometries ( $T_{Carm_0}^{Carm_k}$ ) are estimated from the joint injection device registration. The registration is initialized by  $T_{Carm}^v$  and estimates deformable spine vertebrae poses ( $T_{Carm_0}^{v_m}, m \in \{1..M\}$ ) by solving the optimization:

$$\min_{T_{Carm_0}^{v_m} \in SE(3)} \sum_{k=1}^K \mathcal{S}\left(I_k^S, \mathcal{P}\left(\sum_{m=1}^M V_m; T_{Carm_0}^{v_m}, T_{Carm_0}^{Carm_k}\right)\right). \quad (4.5)$$

The registration setup and optimization strategies are the same as single-view registration. Multi-view spine vertebrae registration functioned as an accurate local search of each vertebra component of the deformable spine object. The vertebrae pose estimation ( $T_{Carm_0}^{v_m}$ ) is reported in the reference C-arm frame, which is used to compute the relative pose transformation to navigate the robot to the injection position. We provide a step-by-step registration workflow chart in Fig. 4.5.

**Figure 4.5:** Workflow of spine registration.

## 4.5 Experiments and Results

### 4.5.1 Femur Registration

#### 4.5.1.1 Simulation Setup and Evaluation Metrics

We verified the accuracy of the proposed femur registration pipeline with a series of simulation studies with randomized projection geometries and anatomical poses. The multi-view C-arm geometry was simulated including a perturbed AP view and two views at random rotations about the C-arm orbit with a mean and STD of  $+10 \pm 3^\circ$  and  $-15 \pm 3^\circ$ . Random movements of the pelvis were sampled uniformly to simulate patient pose variations, including translation from 0 to 10 mm and rotation from -10 to 10 degrees. We sampled ran-

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

dom rotations of the proximal femur with respect to the center of the femoral head (FH). The axis of rotation was sampled uniformly between -15 and 15 degrees. The full pipeline was initialized using the annotations as determined by the pre-trained deep network model. We used the registration workflow described in Section 4.4.2 to produce the pose estimation of the femur with respect to the C-arm frame.

We report the performance of both pelvis and femur registrations based on our simulated ground truth poses of the objects with respect to the C-arm source frame. The rotation errors across anatomical axes are computed by decomposing the rotation matrix of the delta frame into Euler angles using the XYZ convention. The total rotation error is the axial angle magnitude of the rotation matrix. Pelvis registration is reported with respect to the rotation center frame (PC) at the center of pelvis volume to the first view C-arm frame ( $C_{arm_0}$ ),  $T_{C_{arm_0}}^{PC} = T_{C_{arm_0}}^{VP} \cdot T_{VP}^{PC}$ , where PC is annotated from the segmented pelvis volume. The transformation error is reported using

$$\delta T_{pel} = (T_{C_{arm_0}}^{PC})_{gt} \cdot (T_{C_{arm_0}}^{PC})_{regi}^{-1} \quad (4.6)$$

where  $(T_{C_{arm_0}}^{PC})_{gt}$  is the ground truth transformation when the C-arm is set to first view. Femur registration accuracy is reported with respect to the rotation center frame (FH) at the center of femoral head when C-arm is set to the first

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

view ( $Carm_0$ ) using  $T_{Carm_0}^{FH} = T_{Carm_0}^{FV} \cdot T_{FV}^{FH}$ . The transformation error is:

$$\delta T_{fem} = (T_{Carm_0}^{FH})_{gt} \cdot (T_{Carm_0}^{FH})_{regi}^{-1} \quad (4.7)$$

where  $(T_{Carm_0}^{FH})_{gt}$  is the ground truth transformation when the C-arm is set to the first view.

We also report the multi-view C-arm pose estimation accuracy by calculating the relative transformation of the second and third C-arm frames ( $Carm_1, Carm_2$ ) with respect to the first view C-arm frame. The transformation error is reported using

$$\delta T_{Carm_0}^{Carm_1} = (T_{Carm_0}^{Carm_1})_{gt} \cdot (T_{Carm_0}^{Carm_1})_{regi}^{-1}, \quad (4.8)$$

$$\delta T_{Carm_0}^{Carm_2} = (T_{Carm_0}^{Carm_2})_{gt} \cdot (T_{Carm_0}^{Carm_2})_{regi}^{-1}. \quad (4.9)$$

### 4.5.1.2 Simulation Results

We performed a total of 1,000 simulations with varying projection geometries and pelvis initializations. The mean femur registration error ( $\delta T_{fem}$ ) was  $0.81 \pm 0.76$  mm and  $0.73 \pm 0.23^\circ$ , a median of 0.74 mm and  $0.71^\circ$ ; mean injection device registration error ( $\delta T_{inj}$ ) of  $1.00 \pm 0.77$  mm and  $0.23 \pm 0.21^\circ$ , a median of 0.80 mm and  $0.20^\circ$  reported in translation and rotation, respectively. The details of registration error about each axis is tabulated in Table 4.1. The mean error of multi-view C-arm pose estimation ( $\delta T_{C_1}^{C_2}, \delta T_{C_1}^{C_3}$ ) was  $1.37 \pm 0.82$  mm,

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

$0.23 \pm 0.16^\circ$  and  $0.69 \pm 0.45$  mm,  $0.16 \pm 0.14^\circ$ , respectively. Fig. 4.6 presents the statistical histogram distribution of pelvis, femur and the multi-view C-arm pose errors.

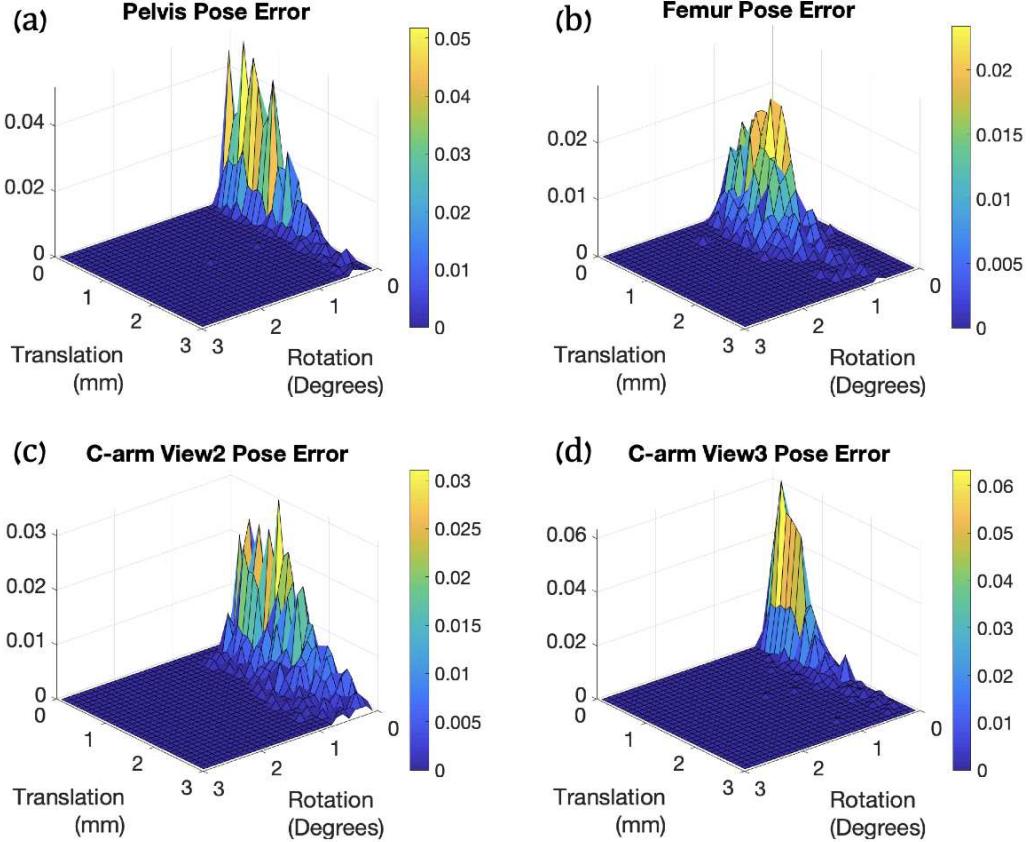
**Table 4.1:** Simulation Results of Pelvis and Femur Registration Errors

		Translation error (mm)				Rotation error (degrees)			
		x (IS)	y (LR)	z (AP)	total	x (IS)	y (LR)	z (AP)	total
Pelvis	mean	$0.23 \pm 0.23$	$0.11 \pm 0.23$	$0.85 \pm 0.76$	$0.92 \pm 0.80$	$0.11 \pm 0.28$	$0.08 \pm 0.23$	$0.06 \pm 0.09$	$0.17 \pm 0.36$
	median	0.20	0.07	0.72	0.77	0.08	0.05	0.04	0.13
Femur	mean	$0.14 \pm 0.15$	$0.12 \pm 0.17$	$0.77 \pm 0.74$	$0.81 \pm 0.76$	$0.65 \pm 0.23$	$0.25 \pm 0.15$	$0.11 \pm 0.10$	$0.73 \pm 0.23$
	median	0.12	0.09	0.72	0.74	0.64	0.24	0.09	0.71
C-arm	mean	$0.36 \pm 0.34$	$0.19 \pm 0.18$	$1.23 \pm 0.84$	$1.37 \pm 0.82$	$0.09 \pm 0.08$	$0.15 \pm 0.14$	$0.10 \pm 0.10$	$0.23 \pm 0.16$
	median	0.27	0.14	1.11	1.25	0.07	0.11	0.06	0.19
view2	mean	$0.22 \pm 0.22$	$0.23 \pm 0.18$	$0.55 \pm 0.43$	$0.69 \pm 0.45$	$0.10 \pm 0.13$	$0.08 \pm 0.08$	$0.05 \pm 0.05$	$0.16 \pm 0.14$
	median	0.15	0.20	0.47	0.60	0.08	0.06	0.04	0.14
view3	mean	$0.22 \pm 0.22$	$0.23 \pm 0.18$	$0.55 \pm 0.43$	$0.69 \pm 0.45$	$0.10 \pm 0.13$	$0.08 \pm 0.08$	$0.05 \pm 0.05$	$0.16 \pm 0.14$
	median	0.15	0.20	0.47	0.60	0.08	0.06	0.04	0.14

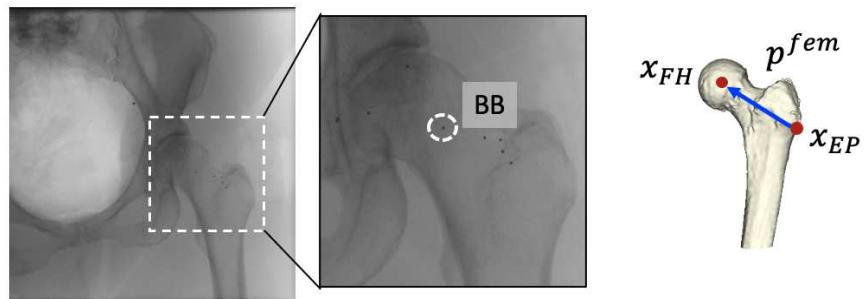
### 4.5.1.3 Cadaver Experiments and Results

We further tested our algorithm in cadaver experiments. A female specimen, including the lower torso, pelvis, and femurs, was used for the study. A Siemens CIOS Fusion C-arm with a 30 cm flat panel detector was used to collect intraoperative fluoroscopy. To obtain the ground truth poses for the femur, metallic BBs were implanted into the femoral head as shown in Fig. 4.7. The BBs were implanted closer to the trochanter and the femoral head center region in order to accurately estimate the femoral head pose. An injection

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY



**Figure 4.6:** Normalized 2D histograms of (a) pelvis pose ( $\delta T_{pel}$ ), (b) femur pose ( $\delta T_{fem}$ ), (c) C-arm view2 ( $\delta T_{C_1}^{C_2}$ ), (d) C-arm view3 ( $\delta T_{C_1}^{C_3}$ ) error for the femur registration simulation studies.



**Figure 4.7:** Left: An example intraoperative fluoroscopic image is presented. The injected BBs are zoomed in. One example BB location is marked with a dotted circle. Right: Illustration of femur head injection path. The injection entry point ( $x_{EP}$ ) and femur head center point ( $x_{FH}$ ) are marked with red dots. The path is shown using a blue vector ( $p^{fem}$ ).

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

path ( $p^{fem}$ ) for femoroplasty was planned on the pre-operative CT scan of the specimen femur based on the biomechanical analysis (Fig. 4.7) [1]. We calculate the femur entry point displacement error with respect to the C-arm source frame using

$$\delta x_{Carm_0}^{EP} = \|(x_{Carm_0}^{EP})_{gt} - (x_{Carm_0}^{EP})_{regi}\|_2 \quad (4.10)$$

where  $x_{Carm_0}^{EP}$  is calculated using  $x_{Carm_0}^{EP} = T_{Carm_0}^{V^F} \cdot x_{V^F}^{EP}$  ( $V^F$  is the femur CT volume). The injection path direction error ( $\delta\theta_{Carm_0}^{fem}$ ) is calculated using

$$\delta\theta_{Carm_0}^{fem} = \arccos\left(\frac{(\mathbf{p}_{Carm_0}^{fem})_{gt} \cdot (\mathbf{p}_{Carm_0}^{fem})_{regi}}{\|(\mathbf{p}_{Carm_0}^{fem})_{gt}\| \cdot \|(\mathbf{p}_{Carm_0}^{fem})_{regi}\|}\right)^1 \quad (4.11)$$

The proposed registration algorithm was tested on our collected intra-operative fluoroscopic images. Ground truth femur poses were estimated by solving a PnP problem using the corresponding BB locations manually annotated in the 3D CT scans and acquired 2D fluoroscopic images, respectively. We reported the errors on the femur entry point and the path direction. Table 4.2 presents the numeric results of six independent trials. The mean entry point error ( $\delta x_{EP}$ ) was  $2.71 \pm 1.12$  mm, and mean injection path direction error ( $\delta\theta^{fem}$ ) was  $1.36 \pm 0.78$  degrees. The clinical acceptable accuracy for femur head drilling and injection is 2 - 3 mm in translation and 2 degrees in rotation.

**Table 4.2:** Cadaver Study Results of Femur Pose Error Metrics

---

<sup>1</sup>Subscripts *gt* and *regi* refer to ground truth and registration, respectively.

Trial ID	I	II	III	IV	V	VI
Entry Point (mm)	1.34	2.44	2.41	1.99	3.67	4.38
Femur Path (degree)	1.35	0.73	1.82	0.83	0.73	2.69

## 4.5.2 Spine Vertebrae Registration

### 4.5.2.1 Simulation Study Setup

We conducted simulation studies to verify the feasibility of the proposed spine vertebrae registration method. Lower torso CT scan images of a male caudaveric specimen were acquired for fluoroscopic simulation and spine vertebrae registration. Vertebrae S1, L2, L3, L4, and L5 were segmented. The C-arm simulation environment keeps the same as introduced in Section 4.5.1.1.

To simulate the intra-operative spine shape difference from the pre-operative CT scans, we applied random rotation change to the consecutive vertebrae CT segmentations. Fig. 4.8 presents an example of this simulated spine deformation. This deformed spine model was used to perform single-view rigid spine registration and initialize the vertebrae pose in multi-view deformable spine registration. Reference frames of spine vertebrae were manually annotated at the center of each individual vertebra in the pre-operative CT scan.

We tested the registration performance under various settings, including single-view and multi-view C-arm geometries:

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

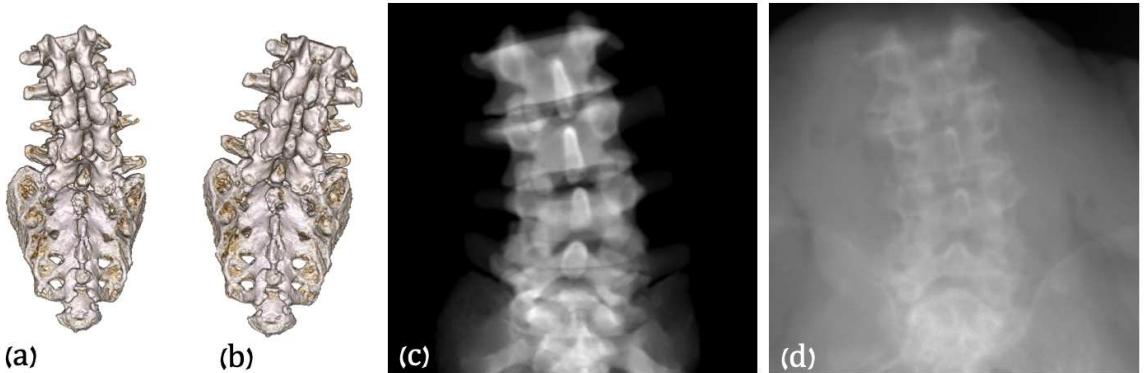
- *Single-view Registration* We performed single-view 2D/3D registration workflows for the rigid spine and the deformable spine. For every registration running, uniformly sampled rotations from -5 to 5 degrees in all three axes were applied to the vertebrae segmentations. Random initializations of the spine were uniformly sampled including translation from 0 to 10 mm and rotation from -10 to 10 degrees.
- *Multi-view Registration* Three multiple C-arm pose geometries were estimated with a uniformly sampled random separation angle between 20 and 25 degrees for the two side views. The three registration workflows tested in single-view were performed with the same settings under this multi-view setup.

We report the registration accuracy based on our simulated ground truth poses of the objects using  $\delta T_{Carm}^V = (T_{Carm}^V)_{gt} \cdot (T_{Carm}^V)_{regi}^{-1}$ , where *gt* and *regi* refer to ground truth and registration estimation, respectively.

### 4.5.2.2 Results

1,000 simulation studies were performed with randomized poses of the injection device and the spine for each registration workflow. Numeric results and statistical plots are presented in Table 4.3 and Fig. 4.9.

Table 4.3 summarizes the magnitudes of translation and rotation errors. The vertebrae error is computed as a mean error of vertebra S1, L2, L3, L4,



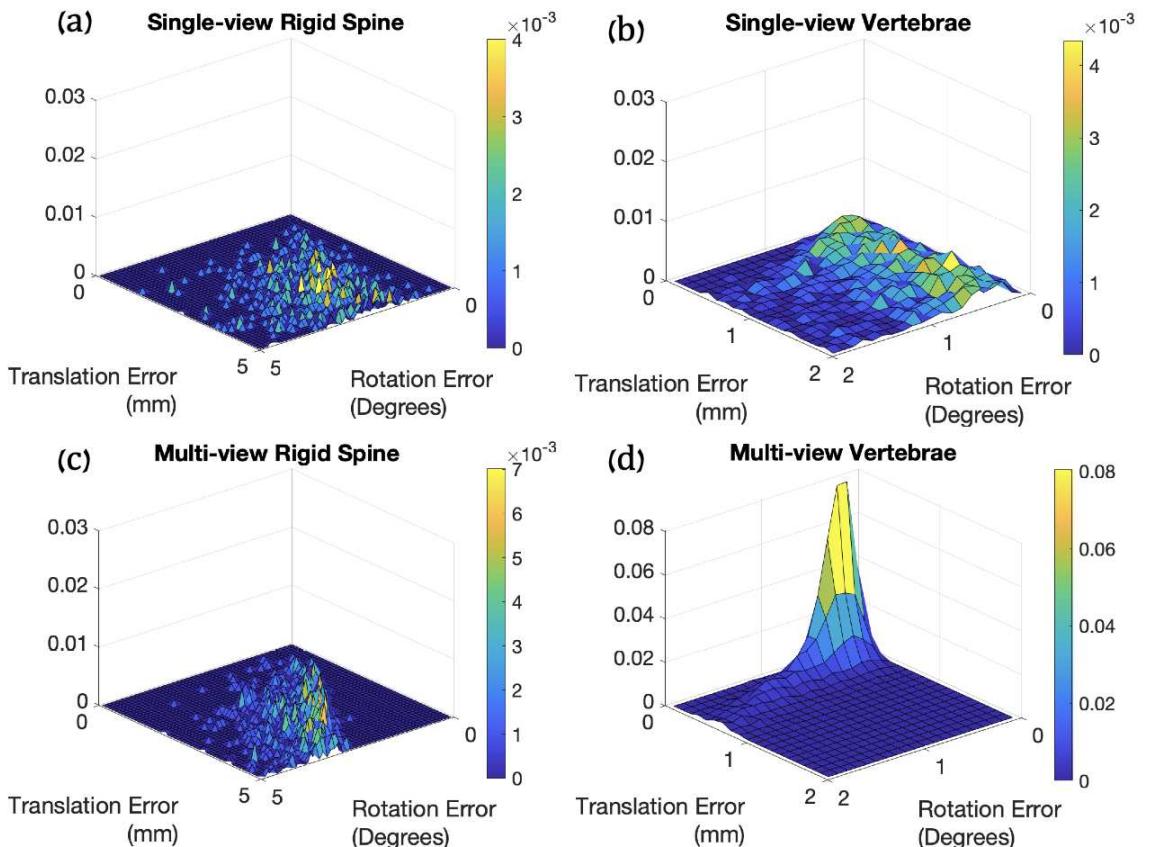
**Figure 4.8:** (a) Rendered vertebrae segmentation from pre-operative CT scans. (b) An example of a randomly simulated spine shape. (c) An example DRR image of the spine vertebrae. (d) An example simulation an X-ray image.

and L5. For single-view registration, we achieved a mean translation error of  $3.50 \pm 2.91$  mm and a mean rotation error of  $1.05 \pm 1.88$  degrees using vertebra-by-vertebra registration. In multi-view registration, the mean vertebra registration error was  $0.76 \pm 0.28$  mm and  $0.88 \pm 0.68$  degrees in translation and rotation respectively. A joint histogram of the translation and rotation errors is presented in Fig. 4.9. From the plots, we clearly observed the multi-view vertebrae registration has the best error distribution with the cluster close to zero errors. The clinical requirement for TLEI is delivering the needle tip within the triangle safety zone, which we will introduce in Chapter 7.

## 4.6 Discussion

The results of both femur and spine vertebrae registration suggest the feasibility of applying our proposed fiducial-free 2D/3D registration method for

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY



**Figure 4.9:** Normalized 2D histograms of spine registration pose error ( $\delta T_{Carm}^{ID}$ ,  $\delta T_{Carm}^v$ ) reported in joint magnitudes of translation and rotation.

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

**Table 4.3:** Mean Registration Error in Spine Vertebrae Registration Simulation Study

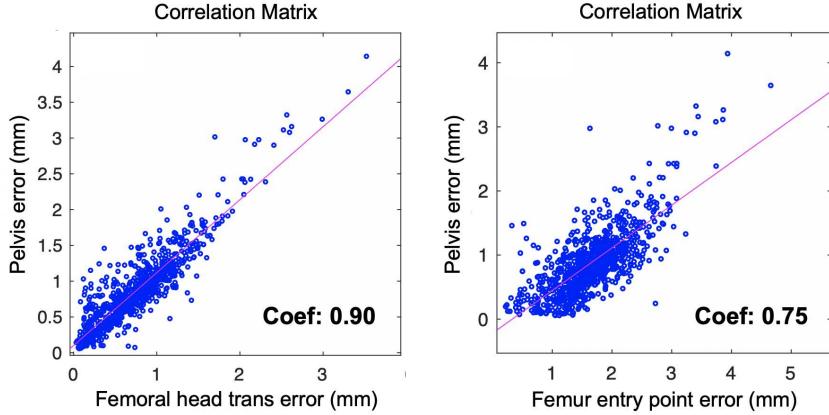
		Translation Error (mm)	Rotation Error (degrees)
Single-View	Rigid Spine	$4.79 \pm 2.36$	$2.79 \pm 1.70$
	Vertebrae	$3.50 \pm 2.91$	$1.05 \pm 1.88$
Multi-View	Rigid Spine	$3.69 \pm 1.60$	$2.89 \pm 1.23$
	Vertebrae	$0.76 \pm 0.28$	$0.88 \pm 0.68$

pose estimation of bone anatomies.

The use of the pelvis as a fiducial gives an accurate estimation of the multi-view C-arm poses. The pelvic bone has detailed features that are very helpful for multi-view registration. Also, an accurate initialization of the femoral head center (FH) helps constrain the femur registration search space and avoids bad local minima. Fig. 4.10 shows the correlation matrix scatter plots between pelvis error and femur error. We report the pelvis translation error as measured from the femoral head center (FH) frame to better illustrate the relationship with the femur translation error. The femur registration accuracy and the entry point accuracy are heavily related to the pelvis registration accuracy with the coefficient factors of 0.90 and 0.75, respectively. The result supports the conclusion that a more accurate pelvis registration leads to a better femur registration result.

The registration accuracy is reduced in cadaver studies. The potential reasons are 1) the spectrum and exposure of the real fluoroscopic images are different from the DRR image; 2) the simulated C-arm projection geometries are dif-

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY



**Figure 4.10:** Left: Scatter plot of correlation matrix between femoral head center translation error and pelvis translation error reported in the femoral head center. Right: Scatter plot of correlation matrix between femur entry point error and pelvis translation error reported in the femoral head center. Correlation coefficients are marked on the right bottom of each plot.

ferent from the cadaveric studies; 3) the BB injection, annotation, and segmentation are likely to introduce errors. The registration accuracy is also related to the relationship between magnification factor and the integral appearance of the pelvis. Capturing a larger portion of the pelvis includes more features, however, the magnification factor is reduced.

Our spine registration simulation studies have shown that multi-view registration is significantly more accurate and stable than single-view registration in all the ablation registration workflows (Table. 4.3). This is because multi-view projection geometries fundamentally improve the natural ambiguity of single-view registration. Our specially designed vertebra-by-vertebra registration solves the problem of spine shape deformation between pre-operative CT scan and intra-operative patient pose. In simulation, the mean multi-

## CHAPTER 4. POSE ESTIMATION OF BONE ANATOMY

view registration error decreased from  $3.69 \pm 1.60$  mm,  $2.89 \pm 1.23$  degrees to  $0.76 \pm 0.28$  mm,  $0.88 \pm 0.68$  degrees in translation and rotation, using pre-operative rigid spine segmentation compared to multiple vertebrae.

The registration was tested running on an NVIDIA GeForce RTX 2080 graphics processor. The single-view registration takes about 20 seconds and the multi-view registration takes around 30 seconds. The full registration process for the pelvis and femur takes about 3 minutes, The full registration process for the spine vertebrae takes about 2 minutes.

One common concern of the fluoroscopic navigation system is excessive radiation exposure to the patient. Our approach requires ten to twelve X-ray images to register the patient to our injection device. Since X-ray images are commonly used in the clinician's manual injections to checking the needle position, this amount of radiation is acceptable for this procedure. Our pipeline is designed to be fully automated, however, our current implementation required a few manual annotations from the clinician to initialize the registration. Future work would consider automating the intra-operative landmark detection to further simplify the workflow, similar to the automatic pelvis initialization.

## 4.7 Conclusion

We have developed a fiducial-free 2D/3D registration pipeline, which uses multiple view fluoroscopic images to register the bone anatomies including femur and spine vertebrae. The method was evaluated through simulations with varying geometries and initializations, and a cadaveric specimen study for femur registration verification. The proposed method showed the feasibility of an image-based and fiducial-free registration approach for the pose estimation of bone anatomies.

## 4.8 Acknowledgement

Sincere thanks to Dr. Robert Grupp's efforts in building the 2D/3D registration framework and his kind guidance in familiarizing the software. I thank Dr. Amirhossein Farvardin and Mrs. Mahsan Bakhtiarinejad for their efforts in performing femoroplasty path planning using biomechanical analysis. I appreciate the efforts of Mr. Henry Phalen, Mr. Adam Margalit, Dr. Liuhong Ma, and Ms. Mareike Thies for their help in the cadaveric experiments. I also thank Drs. Mehran Armand, Russell Taylor, Amit Jain, and Mathias Unberath for their insightful remarks on the method development.

# **Chapter 5**

## **Projective Spatial Transformers for 2D/3D Registration**

### **5.1 Introduction**

As discussed in Chapter 1, one of the challenges of conventional intensity-based 2D/3D registration is that the capture range is very limited due to the presence of local minima in hand-crafted loss functions. Thus, the initial object pose needs to be close enough to the ground truth pose to achieve global optimum during registration.

In this Chapter, we introduce an end-to-end differentiable deep network architecture which learns to approximate a convex-shape similarity function, aiming at extending the registration capture range. The network uses a novel

## CHAPTER 5. PROST

**Projective Spatial Transformer (ProST)** module that generalizes spatial transformers to projective geometry. ProST computes differentiable gradients of 2D image with respect to 3D pose parameters and volumes. Our pipeline is trained using a specially designed double backward gradient-driven loss function. To the best of our knowledge, we are the first to describe the spatial transformers in the context of projective transmission imaging, including rendering and pose estimation.

We present the superiority of ProST network similarity on simulation and real X-ray data with intensive controlled studies, using CMAES intensity-based registration methods as benchmark. The studies were performed on pelvis and spine anatomies. The work in this Chapter presents a promising direction of using learning-based method to overcome the challenge of conventional 2D/3D registration methods in a data-driven manner. This presented approach is a proof-of-concept, which was not included in the actual clinical study validation in Chapter. 7.

## 5.2 Contributions

The contributions of this work include:

- We propose a Projective Spatial Transformer module (ProST), which enables differentiable volumetric rendering of X-ray images using CT vol-

## CHAPTER 5. PROST

umes and analytical gradient calculation with respect to pose parameters.

- We demonstrate its application to 2D/3D image registration by end-to-end learning a convex-shape similarity function.
- Comprehensive ablation experiments were performed to compare the conventional CMAES image intensity-based registration methods and the state-of-the-art learning-based registration methods.

The novel design of ProST, for the first time, enables analytical gradient propagation to the pose parameters for projective X-ray transmission imaging. Our ProST module can be used as a plug-and-play to any modern deep network architecture and makes end-to-end learning feasible under projection geometry. Our efforts to solve the non-convexity problem of conventional intensity-based 2D/3D registration present one use case of ProST. However, we believe it has broader benefits to 3D medical imaging research.

The design of ProST and preliminary results were presented at the 23th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2020) [97].

### 5.3 Related Work

Due to the limited capture range of conventional 2D/3D registration, researchers have worked on extracting dominant features of the object for pose

## CHAPTER 5. PROST

recovery. Such examples include SIFT features for femur [86], Generalized Hough Transform (GHT) for spine vertebrae [98], and learning-based shape encodings for metal implants [99], etc. In recent years, pose estimation using corresponding 2D and 3D anatomical landmarks has become popular. Researches have proposed deep learning models which automatically detect 2D landmarks to solve the PnP problem [94] for pose estimation [58, 59, 89]. However, these features and anatomical landmarks must be object-specifically defined with domain knowledge and clearly visible in the target image for detection.

Direct pose regression from image observations has gained interest in literature. Miao et al. were the first that employ CNN regressors to directly estimate the pose parameters [100]. Li et al. proposed DeepIM which performs iterative pose updates using learned gradient predictions [101]. Gu et al. designed an iterative scheme for X-ray registration by predicting Riemannian pose gradients [57]. Jaganathan et al. also learns pose updates by using Point-to-Plane Correspondences [102]. Several researchers proposed reinforcement learning paradigms to iteratively estimate a transformation [103] [104] [105]. These methods aim at fitting a mapping function from 2D image appearance to 3D pose or pose updates by using deep networks as powerful encoders. However, such mapping is a highly complex and ill-posed problem in transmission X-ray imaging. Thus, these approaches are prone to have strong overfitting on training domain. Because the learned mapping is unconstrained, the pose

## CHAPTER 5. PROST

predictions can be easily altered or explode by a tiny shift of the input image appearance.

Compared to learning a mapping function, a more desirable solution is learning an effective convex “similarity function” for iterative optimization, because 1) the problem complexity of fitting the neural network to a target shape is lower than mapping to precise pose values, and 2) the iterative optimization is more likely to converge to global minima compared to unconstrained prediction mapping. However, such approach requires computing the *differentiable gradient* of the similarity function with respect to the 3D pose parameters for end-to-end network training and iterative registration updates. Unfortunately, this differentiable gradient computation is not yet feasible in literature for X-ray imaging tasks.

Mathematically, the mapping from volumetric 3D object  $V$  to projective transmission DRR  $I_m$  can be modeled as  $I_m = A(\theta) \cdot V$ , where  $A(\theta)$  is the system matrix that depends on pose parameter  $\theta \in \text{SE}(3)$ . In conventional intensity-based 2D/3D registration, we seek to retrieve the pose parameter  $\theta$  such that the moving image  $I_m$  simulated from  $V$  is as similar as possible to the target image  $I_t$ :

$$\min_{\theta} S(I_t, I_m) = \min_{\theta} S(I_t, A(\theta)V), \quad (5.1)$$

where  $S$  is the similarity objective function, such as P-Grad-NCC or MI. Gradient decent-based optimization methods require the gradient  $\frac{\partial S}{\partial \theta} = \frac{\partial S}{\partial A(\theta)} \cdot \frac{\partial A(\theta)}{\partial \theta}$

## CHAPTER 5. PROST

at every iteration. Although the mapping was constructed to be differentiable, analytic gradient computation is still impossible due to the excessively large memory footprint of  $A$  for all practical problem sizes<sup>1</sup>. Traditional stochastic optimization strategies, such as CMAES [21] or BOBYQA [22], use numeric-based methods *without* derivatives. Due to such challenges, the similarity functions, including both hand-crafted and learning-based designs, are limited to be computed only on 2D image domain by generating DRR projection at every iteration.

Spatial transformer networks (STNs) [107] have been applied to estimate deformation field [108] [109] for 3D registration problems. Yan et al. proposed perspective transformer nets which applied STNs for 3D volume reconstruction [110].

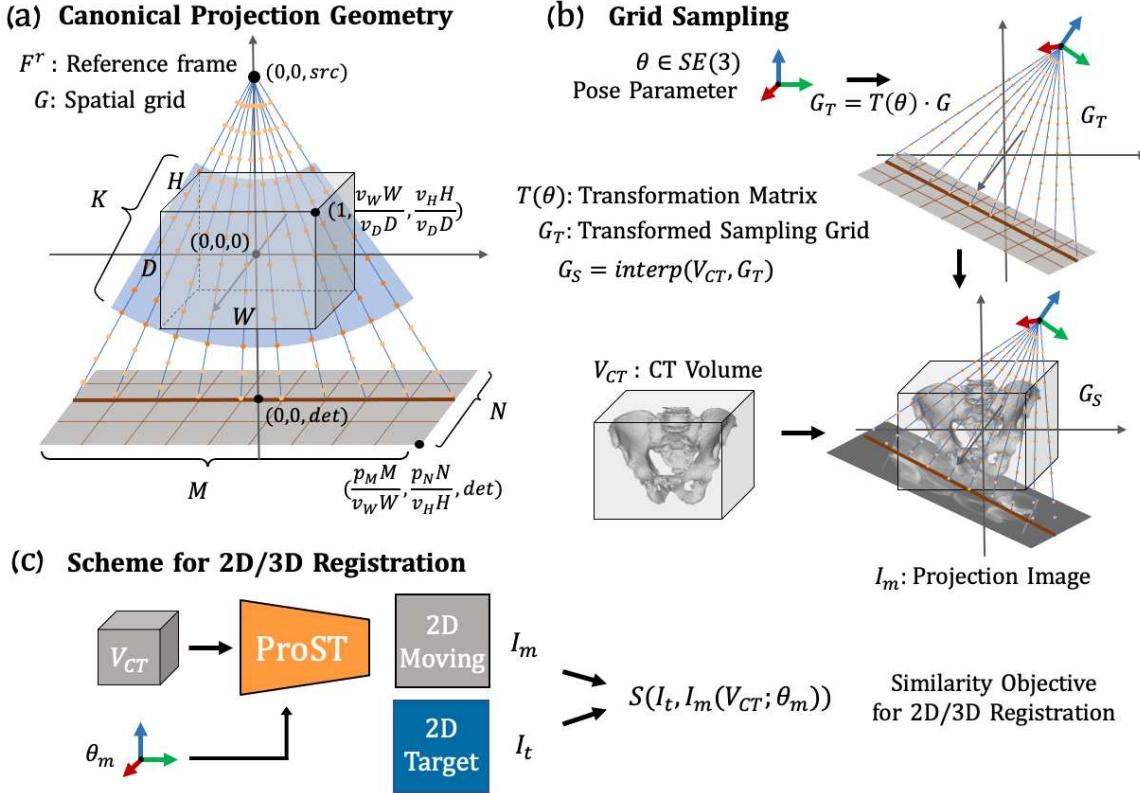
In this Chapter, we propose an analytically differentiable volume renderer that follows the terminology of spatial transformer networks and extends their capabilities to spatial transformations in projective transmission imaging. In the following section, we present our novel design of projective spatial transformer (ProST), which computes analytical gradients from 2D projection domain to 3D pose parameter and volume. We then present that a convex-shape similarity function can be learned end-to-end from 2D to 3D domains, which significantly extends the capture range of intensity-based 2D/3D registration.

---

<sup>1</sup>It is worth mentioning that this problem can be circumvented via ray casting-based implementations if one is interested in  $\partial S \partial V$  but not in  $\partial S \partial \theta$  [106].

## 5.4 Methods

### 5.4.1 Projective Spatial Transformer (ProST)



**Figure 5.1:** (a) Canonical projection geometry and a slice of cone-beam grid points are presented with key annotations. The blue fan covers the control points which are used for further reshaping. (b) Illustration of grid sampling transformer and projection. (c) Scheme of applying ProST to 2D/3D registration.

The proposed ProST module is presented in Fig. 5.1. In this section, we introduce its geometric design and how it can be used for 2D/3D registration.

### 5.4.1.1 Canonical Projection Geometry

Given a CT volume  $V_{CT} \in \mathbb{R}^{D \times W \times H}$  with voxel size  $(v_D, v_W, v_H)$ , we define a reference frame  $F^r$  with the origin at the center of  $V_{CT}$ . We use the volume depth ( $Dv_D$ ) to normalize coordinates in the canonical geometry. The volume corner point coordinate  $(Dv_D/2, Wv_W/2, Hv_H/2)$  is transformed as  $(1, \frac{Wv_W}{Dv_D}, \frac{Hv_H}{Dv_D})$  after normalization. Given an X-ray projection camera intrinsic matrix  $\mathcal{K} \in \mathbb{R}^{3 \times 3}$ , we denote the associated source point as  $(0, 0, src)$  in  $F^r$ . The spatial grid  $G$  of control points, shown in Fig. 5.1(a), lies on  $M \times N$  rays originating from this source. Because the control points in regions where no CT voxels exist will not contribute to the line integral, we cut the grid point cloud to a cone-shaped structure that covers the exact volume space for memory concern (Fig. 5.1(a) blue fan). Thus, each ray has  $K$  control points uniformly spaced within the volume  $V$ , so that the matrix  $G \in \mathbb{R}^{4 \times (M \cdot N \cdot K)}$  of control points is well-defined, where each column is a control point in homogeneous coordinates. The uniform distribution of control points along each ray follows the routine of setting a fixed step size in conventional DRR ray casting. The intensity value of the control point is linearly sampled from the volumetric CT voxels. These rays describe a cone-beam geometry which intersects with the detection plane, centered on  $(0, 0, det)$  and perpendicular to the  $z$  axis with pixel size  $(p_M, p_N)$ , as determined by  $\mathcal{K}$ . The X-ray source  $(0, 0, src)$  and center of detector  $(0, 0, det)$  coordinates are all applied with the normalization factor  $Dv_D$ . The upper-right

## CHAPTER 5. PROST

corner of the detection plane is at  $(\frac{p_M M}{v_W W}, \frac{p_N N}{v_H H}, det)$ .

### 5.4.1.2 Grid Sampling Transformer

ProST extends the canonical projection geometry by learning a transformation of the control points  $G$ . Given six-degrees-of-freedom (6 DoF) rigid pose parameter  $\theta \in \text{SE}(3)$ , we obtain a transformed set of control points via the affine transformation matrix  $T(\theta)$ :

$$G_T = T(\theta) \cdot G, \quad (5.2)$$

as well as source point  $T(\theta) \cdot (0, 0, src, 1)$  and center of detection plane  $T(\theta) \cdot (0, 0, det, 1)$ . Since these control points lie within the volume  $V$  but in between voxels, we perform differentiable linear interpolation of  $V$  at the control points  $G_T$ , producing sampled control point values  $G_S$ :

$$G_S = \text{interp}(V, G_T), \quad (5.3)$$

where  $G_S \in \mathbb{R}^{M \times N \times K}$ . Finally, we obtain a 2D image  $I_m \in \mathbb{R}^{M \times N}$  by integrating along each ray. This is accomplished by “collapsing” the  $k$  dimension of  $G_S$ :

$$I^{(m,n)} = \sum_{k=1}^K G_S^{(m,n,k)} \quad (5.4)$$

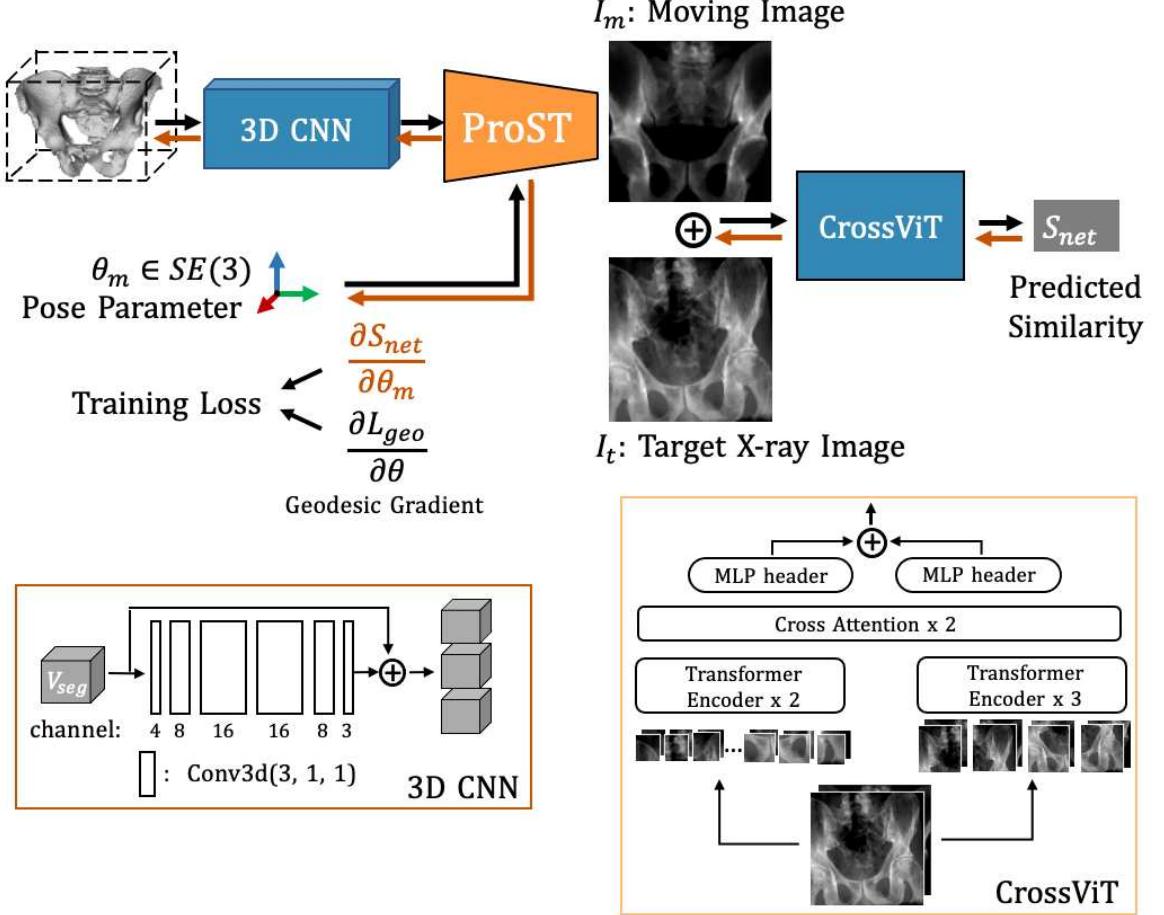
## CHAPTER 5. PROST

Of note, the integrating operation (Eqn. 5.4) can be achieved as a single-step time-efficient matrix summation operation. The process above takes advantage of the spatial transformer grid ( $G$ ), which reduces the projection operation to a series of linear transformations. The intermediate variables are reasonably sized for modern computational graphics processors, and thus can be loaded as a tensor variable. We implement the grid generation function using the C++ and CUDA extension of the PyTorch framework and embed the projection operation as a PyTorch layer with tensor variables. With the help of the PyTorch autograd function, this projection layer enables analytical gradient flow from the projection domain back to the spatial domain. Fig. 5.1 (c) shows how this scheme is applied to 2D/3D registration. Without any learning parameters, we can perform registration with PyTorch’s powerful built-in optimizers on large-scale volume representations. Furthermore, by integrating deep convolutional layers, we show that ProST makes it feasible to approximate a convex shape image similarity in a data-driven manner.

### 5.4.2 Approximating Convex Image Similarity

Geodesic loss,  $L_{geo}$ , which is the square of the geodesic distance in  $SE(3)$ , has been studied for registration problems due to its convexity with respect to pose transformations [111] [112]. Using ProST, we propose an end-to-end DeepNet architecture that learns to approximate the convex shape of  $L_{geo}$ , aiming at

## CHAPTER 5. PROST



**Figure 5.2:** ProST DeepNet Architecture for 2D/3D registration. A pelvis segmentation  $V_{seg}$  is illustrated as 3D input. The rigid pose parameter  $\theta_m$  is shown using RGB cross arrows. An example target X-ray image  $I_t$  and moving image  $I_m$  generated from ProST are presented. Forward pass follows the black arrows. Backward pass follows orange arrows, where the gradient  $\frac{\partial S_{net}}{\partial \theta_m}$  is computed. Detailed structures of CrossViT and 3D CNN are illustrated in blocks on the bottom.

extending the capture range of 2D/3D registration. Given a sampling pose  $\theta_m$  and a target pose  $\theta_t$ , we took the implementation of geomstats [113] to calculate the geodesic distance,  $L_{geo}(\theta_m, \theta_t)$ , and the geodesic gradient,  $\frac{\partial L_{geo}(\theta_m, \theta_t)}{\partial \theta_m}$ .

Fig. 5.2 shows the proposed DeepNet architecture. The input includes a 3D

## CHAPTER 5. PROST

segmentation volume:  $V_{seg}$ , a pose parameter  $\theta_m \in SE(3)$  and a target image:  $I_t$ , which implies a target pose  $\theta_t$ . 3D to 2D projections are performed using the ProST projection module in orange. The network learnable parameters are colored in blue with detailed structures presented on the right side. The 3D CNN is a skip connection from the input volume to multi-channel expansion just to learn the residual. The projected moving image  $I_m$  and the target image  $I_t$  are concatenated together and then pass through a cross vision transformer (CrossViT) encoder. CrossViT is an advanced version of the standard vision transformer, which learns multi-scale feature representations by combining image patches of different sizes [114]. The final output is the network predicted similarity,  $S_{net}$ . We will then explain the network training techniques and their application to 2D/3D registration in the following sections.

### 5.4.3 Double Backward Training Loss

Since the target is to make  $S_{net}$  similar to  $L_{geo}$ , a straightforward solution is defining it as a regression task, for example, using mean squared error,  $MSE(S_{net}, L_{geo})$ , as training loss. However, learning a 6 DoF convex loss landscape from image appearance is particularly challenging by regressing the absolute loss scales. In reality, the scale values do not contribute to the iterative optimization, but the shape of loss matters. Thus, we downgrade the task to focus on learning the convex shape, which essentially refers to learning the

## CHAPTER 5. PROST

second-order gradients.

Specifically, the goal is to make the gradient of our network similarity function with respect to pose parameters,  $\frac{\partial S_{net}}{\partial \theta_m}$ , close to the geodesic gradient  $\frac{\partial L_{geo}}{\partial \theta_m}$ . The black arrows in Fig. 5.2 show the forward pass in a single iteration. The network output can be mathematically formulated as  $S_{net}(\phi; V_{seg}, \theta_m, I_t)$ , where  $\phi$  are the network parameters. The gradients,  $\frac{\partial S_{net}}{\partial \theta_m}$  and  $\frac{\partial S_{net}}{\partial \phi}$ , are computed by applying back-propagation, illustrated with orange arrows in Fig. 5.2. Of note, we do not update network parameters  $\phi$  during this back-propagation. The network training loss is designed by calculating a distance measure,  $M_{dist}\left(\frac{\partial S_{net}}{\partial \theta}, \frac{\partial L_{geo}}{\partial \theta}\right)$ , of the network pose gradient  $\frac{\partial S_{net}}{\partial \theta_m}$  and geodesic gradient  $\frac{\partial L_{geo}}{\partial \theta_m}$ . Since the gradients are also 6 DoF, we again use the geodesic distance  $L_{geo}$  as this distance measurement metric  $M_{dist}$ , which is the true network loss function during training.

We then perform a second forward pass, or “double backward” pass, to get  $\frac{\partial M_{dist}}{\partial \phi}$  for updating network parameters  $\phi$ . To this end, we formulate the network training as the following optimization problem

$$\min_{\phi} M_{dist}\left(\frac{\partial S_{net}(\phi; V_{seg}, \theta_m, I_t)}{\partial \theta_m}, \frac{\partial L_{geo}(\theta_m, \theta_t)}{\partial \theta_m}\right). \quad (5.5)$$

Due to the unique goal of shaping the network function landscape to be convex, our double backward training design uses the gradient of geodesic distance to

drive the entire network training. This is feasible only because ProST makes the network end-to-end differentiable with respect to the pose parameter  $\theta_m$  and 3D volume. We will demonstrate the advantage of this training loss design in ablation studies (Section. 5.5.2).

#### 5.4.4 Domain Randomization

Target images during training were generated using a physically-realistic X-ray simulation framework – DeepDRR [66], which is shown to be effective on learning-based X-ray imaging tasks compared to naive DRR [67]. We applied domain randomization on DeepDRR target images to improve the generalization ability on unseen real data. Domain randomization is a domain generalization technique that introduces drastic changes in the training image appearances, which forces the network to learn more robust features between training and target domains [115]. During each training iteration, we applied the following domain randomization methods sequentially each with a probability of 50% on the target image  $I_t$ :

- *Inverting:*  $\max(I_t) - I_t$  the maximum intensity value was subtracted from all image pixels;
- *Gaussian noise injection:*  $I_t + N(0, \sigma)$ , where  $\sigma$  was uniformly chosen from the interval (0.005, 0.1) multiplied by the image intensity range.

## CHAPTER 5. PROST

- *Gamma transform:*  $\text{norm}(I_t)^\gamma$ , where  $I_t$  was normalized by its maximum and minimum values, and  $\gamma$  was uniformly selected from the interval  $(0.7, 1.3)$
- *Box corruption:* a random number of box regions of  $I_t$  were corrupted with large noise.

### 5.4.5 Application to 2D/3D Registration

When the trained network is applied to registration, the network parameters  $\phi$  are fixed, and the well-trained network including the ProST module is treated as a similarity objective function. Because the pose gradients  $\frac{\partial S_{\text{net}}}{\partial \theta_m}$  are differentiable, the iterative registration optimization can be performed analytically using gradient-based methods such as stochastic gradient descent (SGD) rather than numerically sampling. Following the math descriptions of ProST in Section. 5.4.1 and Fig. 5.1 and 5.2, the analytical gradient can be computed following:

$$\frac{\partial S_{\text{net}}}{\partial \theta_m} = \frac{\partial S_{\text{net}}}{\partial I_m} \cdot \frac{\partial I_m}{\partial G_S} \cdot \frac{\partial G_S}{\partial G_T} \cdot \frac{\partial G_T}{\partial T(\theta_m)} \cdot \frac{\partial T(\theta_m)}{\partial \theta_m}. \quad (5.6)$$

Since the entire architecture is implemented using PyTorch, the 2D/3D registration is performed using PyTorch built-in optimizers, such as SGD or Adam, and learning rate schedulers.

The network similarity is more effective when the initial pose is far away

## CHAPTER 5. PROST

from the target, while less sensitive to local textures compared to traditional hand-crafted image similarity functions. In the next section, we present our efforts in demonstrating its use case of extending the registration capture range with controlled ablation experiments.

## 5.5 Experiments

We performed intensive studies of single-view 2D/3D registration on simulation and real X-ray images to evaluate our approach. We selected the state-of-the-art image intensity-based 2D/3D registration method as benchmark, which uses Patch-based Normalized Gradient Cross Correlation (P-Grad-NCC) [92] score as the similarity metric and CMAES [95] as the optimization strategy. Due to its well-known robustness to local minima, the CMAES benchmark method for pose estimation of bone anatomy and surgical devices has been tested to meet clinical requirements in various orthopedic applications, such as osteotomy [18, 59], femoroplasty [3, 116], core decompression of the hip [65], and transforaminal lumbar epidural injections [78], etc. Our experiments aim at demonstrating the superiority of extending the CMAES registration capture range using ProST network. The initial pose geometry of the registration object was randomly sampled in a wide range. We compared the registration performance of running CMAES from initial poses against CMAES from learning-

## CHAPTER 5. PROST

based pose estimations in all precisely controlled ablation studies.

We introduce our experiment design as follows: In Section. 5.5.1, we describe the general environment setup, dataset, and processing details. In Section. 5.5.2, we present the ablation studies, which include 1) architecture ablation: variations of ProST architecture and two representative learning-based registration methods in the literature, namely DeepIM [101] and DMW [57]; 2) image with overlays; 3) anatomy ablations: pelvis and spine. In Section. 5.5.3, we describe the initial pose sampling strategies and network training hyperparameters, which were precisely controlled overall ablation experiments. In Section. 5.5.4, we present the evaluation metrics of the registration performance.

### 5.5.1 Simulation Environment and Dataset

Our X-ray simulation environment approximates a Siemens CIOS Fusion C-Arm, which has image dimensions of  $1536 \times 1536$ , isotropic pixel spacing of 0.194 mm/pixel, and a source-to-detector distance of 1020 mm. 50 pixels on each dimension were cropped to eliminate the X-ray detector's collimator boundary effects. The images were downsampled to have dimensions of  $128 \times 128$  with a pixel spacing of 2.176 mm/pixel. The source to iso-center distance is 800 mm.

We selected twenty high-quality CT scans from the New Mexico Decedent

## CHAPTER 5. PROST

Image Database (NMDID) [60] for training and simulation study. The CT scans were manually cropped to focus on the hip region and resampled to preserve an isotropic cubic shape of 128 voxels in each dimension. The pelvis anatomy was automatically segmented using the algorithm described in [61]. A separate set of twenty CT scans of the spine anatomy were selected from the same NMDID. The spine vertebrae were segmented using a coarse-to-fine vertebrae localization and segmentation method [117]. For both the pelvis and spine dataset, 19 CTs were used for network training and we left 1 CT for testing.

Our real hip X-ray data were selected from the cadaveric X-rays released by Grupp et al [118]. The six corresponding CT scans and pelvis segmentations were processed using the same methods as described above. Ground truth poses of the pelvis were estimated using the comprehensive image intensity-based 2D/3D registration pipeline described in [118]. The coordinate frames of DeepDRR and xReg were calibrated to the ProST geometry convention as introduced in Section 5.4.1.1.

### 5.5.2 Ablation Study Design

#### 5.5.2.1 Architecture

We conducted ablation studies on variations of the ProST registration architectures to demonstrate the effects of each key component design:

## CHAPTER 5. PROST

- *Baseline*: Baseline ProST network using all techniques introduced in Section 5.4.1.
- *No DR*: The architecture is the same as a baseline but domain randomization (DR) was removed during training time.
- *No 3D Net*: The 3D CNN part (Fig. 5.2) was removed. The goal is to compare the effect of 3D learnable parameters in approximating the desired convex similarity function.
- *MSE Loss*: To show the importance of gradient-driven double backward training loss design, we performed an ablation experiment using  $MSE(S_{net}, L_{geo})$  to train the network. The other training settings were kept the same as baseline.

We also included comparisons with the other representative learning-based 2D/3D registration architectures in the literature:

- *MICCAI*: This is the conference published version of the ProST registration network at MICCAI 2020 [97]. The architecture and training loss was kept the same, while domain randomization was applied and training/testing data was set in line with the controlled experiments in this work.
- *DeepIM*: DeepIM is a popular deep iterative matching network for 6 DoF object pose estimation, which predicts direct pose updates by learning

## CHAPTER 5. PROST

from the optical flow between moving and target images [101]. DeepIM was shown to achieve state-of-the-art results on benchmark computer vision datasets.

- *DMW*: DMW is also proposed to extend the capture range of 2D/3D registration by regressing pose updates from moving and target images [57]. It uses a sequence of DenseNet [119] blocks as the backbone and learns to predict direct geodesic gradients of the two relative poses. DMW was tested to be effective in recovering large pose initialization errors on pelvis X-rays.

DeepIM and DMW were selected as representatives of direct pose regression methods from 2D images in literature. Target and moving image pairs of these two ablations were generated using DeepDRR from the same CT data and pose distributions as ProST baseline training. Of note, except for *No DR*, domain randomization was consistently applied in all the other ablation experiments.

### 5.5.2.2 Image with Overlays

We simulated the challenging intra-operative imaging condition with the surgical tool in the C-arm capture range as overlays. The surgical tool was chosen an integrated drilling and injection device, which was custom designed for the application of femoroplasty [3]. CT scans of the injection device were taken. For experiments on images with tool overlay, we randomly sampled the

## CHAPTER 5. PROST

pose of the injection device and generated DRR from the CT. The tool DRR was overlayed on the original bone image as the target image. We present an example target X-ray image with tool overlay in Fig. 5.3.

### 5.5.2.3 Anatomy

We trained and tested the baseline model on the spine anatomy to demonstrate the generalization ability of ProST. Due to the scarcity of paired spine CT and X-ray images, we only tested the performance on the spine in simulation.

## 5.5.3 Pose Sampling Strategies and Training Parameters

We performed precisely controlled experimental training to benchmark the performance. All experiments were fed with training data from the same distribution and applied the same training strategies. The canonical geometry is the Anterior-Posterior (AP) view, which is the most common case in clinical use. Target training images ( $I_t$ ) were generated using DeepDRR, following a uniformly sampled pose geometry with the random translation of (-25, 25) mm and rotation of (-15, 15) degrees in all three axes. Moving training image ( $I_m$ ) poses were randomly sampled following Gaussian distributions with translation in mm of  $N(0, 25)$  for in-plane (X and Y) direction and  $N(0, 60)$  for depth (Z)

## CHAPTER 5. PROST

direction, with rotation in degrees of  $N(0, 25)$  for all three axes. During testing in simulation, we followed the same training target image pose distribution to generate testing target images. We sampled an exhaustive initialization pose space following uniformly sampled translation of (-200, 200) mm for in-plane direction and (-300, 300) mm for depth direction, uniformly sampled rotation of (-50, 50) degrees. The sample initialization pose sampling strategy was applied during real X-ray image testing.

During each training epoch, one CT was randomly selected, and target and moving image pairs were randomly generated following the distributions described above. 50 iterations were trained for every epoch. The ProST network architecture was trained using an SGD optimizer with a cyclic learning rate between 1e-6 and 1e-4 every 100 steps [120] and a momentum of 0.9. The batch size was chosen as 4 and we trained 300 epochs until full convergence. Following their original training strategies, DeepIM and DMW were trained using Adam optimizer with a learning rate schedule that starts from 1e-4 and decreases by 10% every 20 epochs as well as a momentum of 0.8. The batch size was chosen as 16 and the networks were trained for 150 epochs until full convergence.

## 5.5.4 Registration Testing and Evaluation Metrics

We performed learning-based and conventional CMAES iterative 2D/3D registration testing using randomly sampled target images and initial poses for all ablation experiments on simulation and real X-ray data, separately. For ProST architecture variants including the MICCAI ablation, we used an SGD optimizer to optimize over the 6 DOF pose parameter ( $\theta_m$ ) iteratively until convergence up to a fixed number of steps with a learning rate of 0.1, which decays by a factor of 0.5 for every 15 steps. Because DeepIM and DMW do not learn a similarity score, poses were updated using direct network predictions by generating a DRR image and retrieving the network inference results at each iteration. The registration was considered to be converged if the predicted gradient magnitude went below a pre-defined threshold.

In simulation testing, we used the left-out NMDID CT to simulate target images and its bone segmentation to perform registration. 1,000 images and poses were randomly sampled following the distribution described in Section 5.5.3. In real pelvic X-ray data testing, we manually selected 100 standard AP view images and 100 challenging view images. The challenging views include images with only partial pelvis visible or drastic orientations, which fall outside of the training target image distribution. Five registrations were per-

## CHAPTER 5. PROST

formed on each image with the initial pose sampled from the same distribution as in simulation, resulting in 500 registrations for standard and challenging views, respectively.

For each testing, the CMAES registrations were performed from both the initial pose and the network estimated pose to compare the capture range. The CMAES optimizer was set with a pose sampling population size of 300 in each iteration, and sigmas of 45 degrees and 200 mm for rotation and translation, respectively, which is sufficient to cover the testing pose sampling space.

We report the following error metrics for evaluation:

- *Registration Pose Error*: A residual pose transformation was computed using the estimated ( $\theta_{est}$ ) and ground truth ( $\theta_{gt}$ ) pose parameters:

$$\delta T(\theta_{gt}, \theta_{est}) = T(\theta_{gt}) \cdot T(\theta_{est})^{-1} \quad (5.7)$$

We report the magnitude of translation and rotation errors in mean and median after decoupling  $\delta T(\theta_{gt}, \theta_{est})$  into each degree of freedom.

- *Target Registration Error (TRE)*: TRE is computed by calculating the average point to point distance of the bone segmentation ( $P_{seg}$ ):

$$TRE = \text{mean}(l_2(P_{seg} \cdot T(\theta_{gt}), P_{seg} \cdot T(\theta_{est}))) \quad (5.8)$$

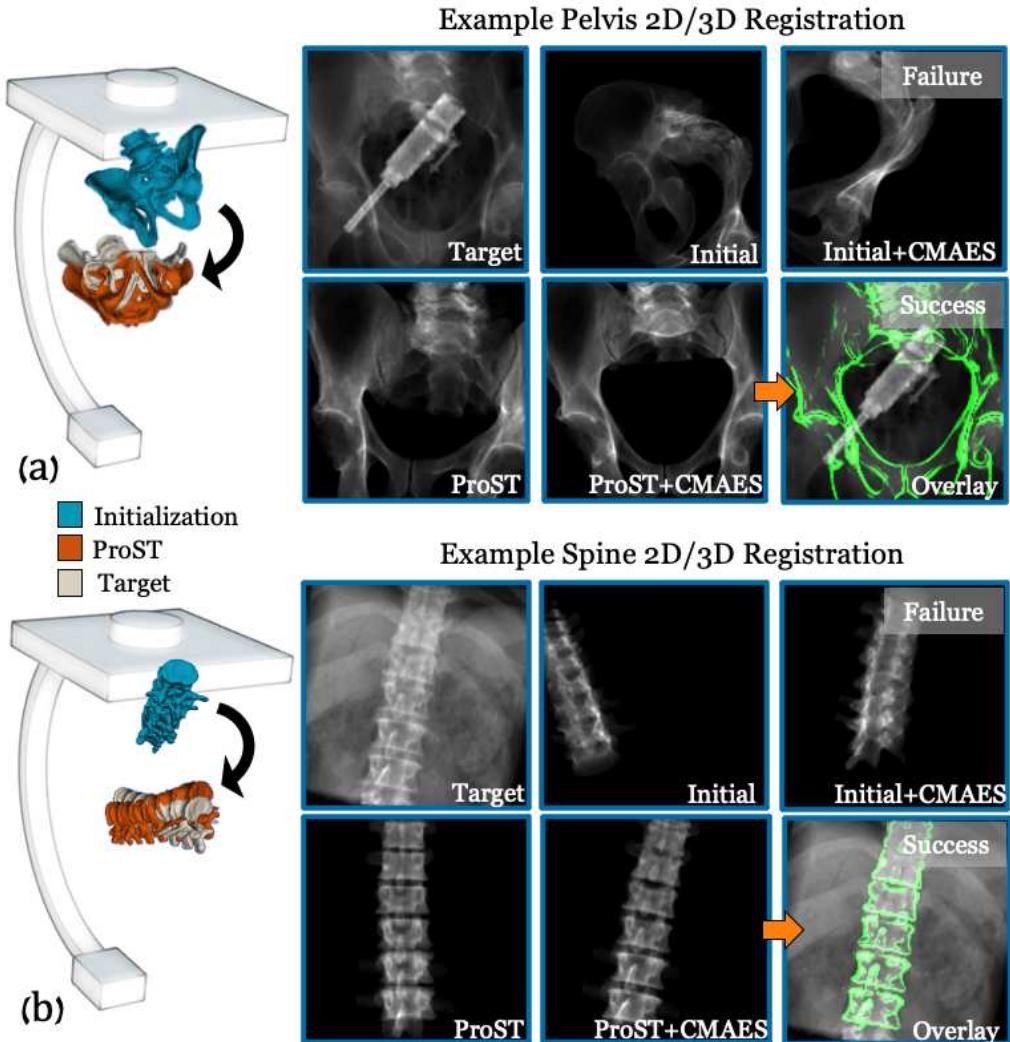
- *Success Rate (SR)*: We define a threshold for successful registrations with TRE less than 10 mm. The success rate is computed as a percentage of the number of successful testings over all testings. The success rate is computed only for the CMAES registrations.

## 5.6 Results

Qualitative results of the iterative 2D/3D registration process are shown in Fig. 5.3. We present two example registrations using a real pelvis X-ray image with surgical tool overlay and a simulated spine DRR image as targets. The CMAES registrations from initialization failed at local minima. The 3D pose rendering shows that the object poses were much closer to ground truth after ProST registration, which then successfully converged to global minima using CMAES.

Numeric results of registration pose errors are presented in Table. 5.1. Because our experiments were initialized with fairly large offsets, failed registrations led to large pose errors. The mean errors were heavily biased by the large-scale outliers, which do not represent the actual distribution. Thus, we only report the median errors at several percentage levels. The CMAES from initialization achieved a 75% median error of 56.95 degrees, 170.40 mm in simulation, and 57.90 degrees, 134.55 mm in real data, in rotation and translation,

## CHAPTER 5. PROST



**Figure 5.3:** Illustration of iterative 2D/3D registration using examples of (a) a real pelvis X-ray image with tool overlay and (b) a simulated spine X-ray image. Renderings of the initialization, ProST estimation, and ground truth pose of the anatomies are illustrated on the left. The black arrow shows the difference between initialization and ProST estimated poses. In each example, the target image, initial projection image, and the failed CMAES registration image from the initial are shown in the first row. The ProST registration image, successful CMAES registration image from ProST estimation, and an overlay image with DRR-derived edge in green are shown in the second row.

## CHAPTER 5. PROST

respectively. The ProST baseline model followed by CMAES performed the best across all the ablation studies, which achieved a 75% median error of 0.25 degrees, 3.91 mm in simulation, and 13.63 degrees, 19.96 mm in real data, in rotation and translation, respectively. In Table. 5.2, we report the TRE and success rate. Using our ProST baseline model, the CMAES registration success rate improved from 32.5% to 82.5% in simulation and 36.0% to 72.0% in real data. In Fig. 5.4, we plot the histograms of rotation and translation errors for the baseline model. We clearly observe that ProST shifted the error distribution from initialization closer to zero, which resulted in a much higher success rate for CMAES, compared to CMAES from initialization.

The results of our comparison experiments all performed worse than the baseline model. We present the results using ProST baseline on the spine anatomy in Table. 5.3. The success rate was 47.1%, compared to 16.2% using CMAES from initialization. We present discussions and analysis of the results in greater detail in the next Section. 5.7.

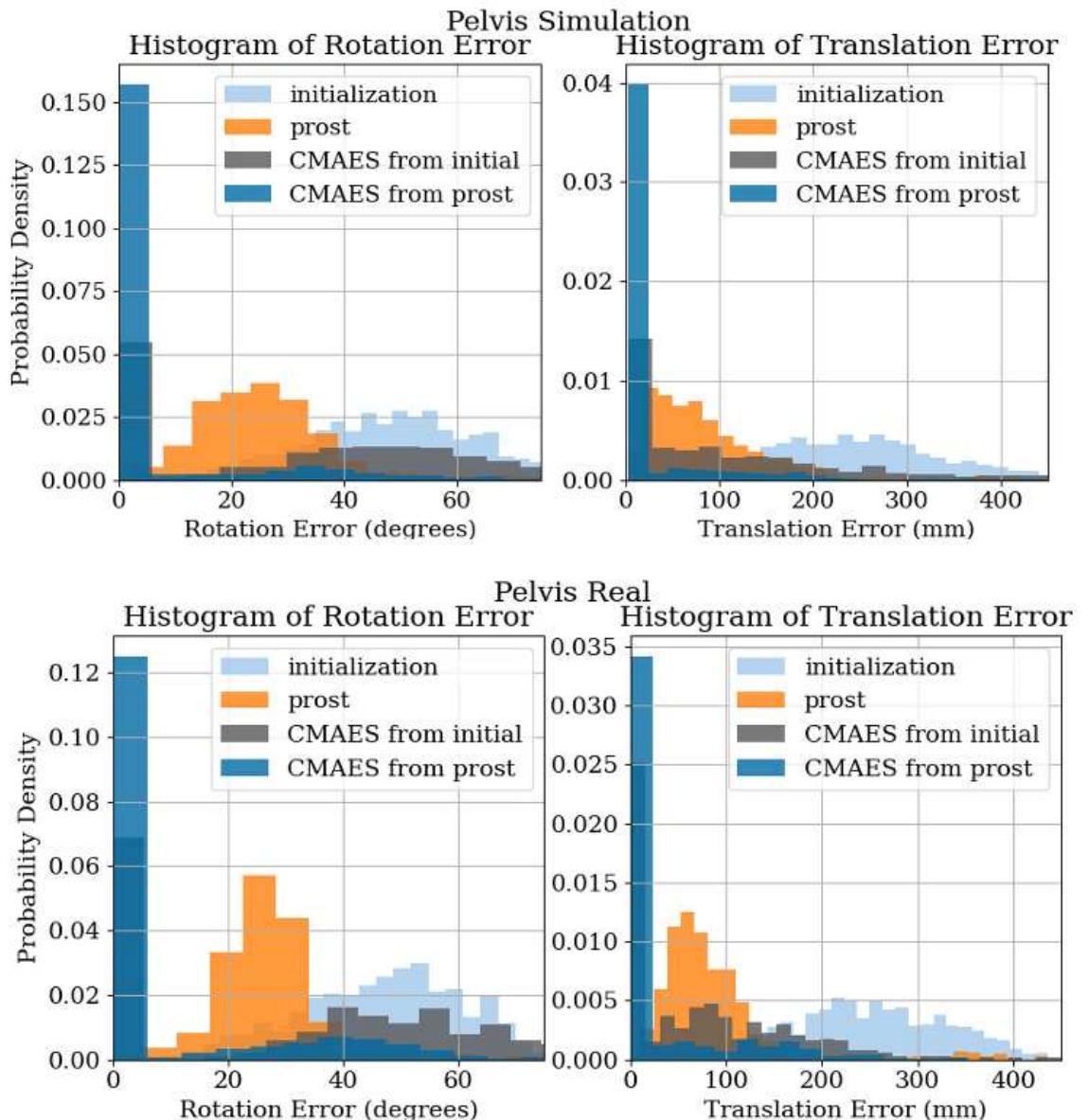
## CHAPTER 5. PROST

**Table 5.1:** Registration Pose Error of Pelvis

	Simulation Study						Standard AP View Real X-ray		
	Rotation Error (degrees)		Translation Error (mm)		Rotation Error (degrees)		50% Med	75% Med	50% Med
	50% Med	75% Med	50% Med	75% Med	50% Med	75% Med	50% Med	75% Med	75% Med
Initialization + CMAES	48.85	59.12	244.88	306.08	48.61	58.00	246.88	307.57	
	38.05	56.95	73.42	170.40	39.65	57.90	62.20		134.55
ProST Baseline + CMAES	25.11	31.40	72.75	116.90	26.79	31.14	71.50	99.70	
	0.19	0.25	3.67	3.91	0.47	13.63	1.67		19.96

In each block, we present the initialization or network estimated registration results, and the *followed* CMAES registration results. Med is short for Median. We present both 50 and 75 percentile median errors.

## CHAPTER 5. PROST



**Figure 5.4:** Histogram of registration pose errors in translation and rotation for pelvis simulation and standard AP view real X-ray study, respectively, from ProST baseline model registrations.

## CHAPTER 5. PROST

**Table 5.2:** Target Registration Error (TRE) and Success Rate of Pelvis

	Simulation Study			Standard AP View Real X-ray		
	TRE (mm)		SR (%)	TRE (mm)		SR (%)
	50% Med	75% Med		50% Med	75% Med	
Initialization	253.39	327.24	-	255.27	324.81	
CMAES	116.79	212.40	32.5	106.27	189.81	36.0
ProST Baseline	90.28	145.32	-	96.10	127.59	
CMAES	3.75	4.06	<b>82.5</b>	2.16	31.51	<b>72.0</b>
ProST Tool Overlay	93.77	150.65	-	96.54	129.73	
CMAES	3.80	15.18	74.8	24.10	155.59	47.4
ProST No DR	92.62	142.08	-	96.45	126.06	
CMAES	3.78	4.14	78.7	2.12	70.51	69.0
ProST No 3D Net	123.43	234.79	-	128.10	247.45	
CMAES	3.99	144.23	57.6	23.07	161.78	49.2
ProST MSE Loss	259.24	338.18	-	260.50	353.73	
CMAES	133.82	241.68	28.3	120.34	198.24	28.6
ProST MICCAI	209.14	293.43	-	236.58	314.84	
CMAES	45.45	175.98	48.2	2.89	156.64	51.2
DeepIM	75.57	295.18	-	226.44	351.73	
CMAES	3.83	137.47	68.4	99.72	196.82	41.2
DeepIM Tool Overlay	80.51	303.81	-	256.50	375.92	
CMAES	3.97	208.37	57.6	171.17	246.52	11.0
DMW	360.13	523.01	-	265.99	393.36	
CMAES	103.19	254.20	39.9	94.45	193.87	38.4
DMW Tool Overlay	360.71	506.34	-	284.39	394.64	
CMAES	98.07	237.68	40.2	170.76	240.53	8.4

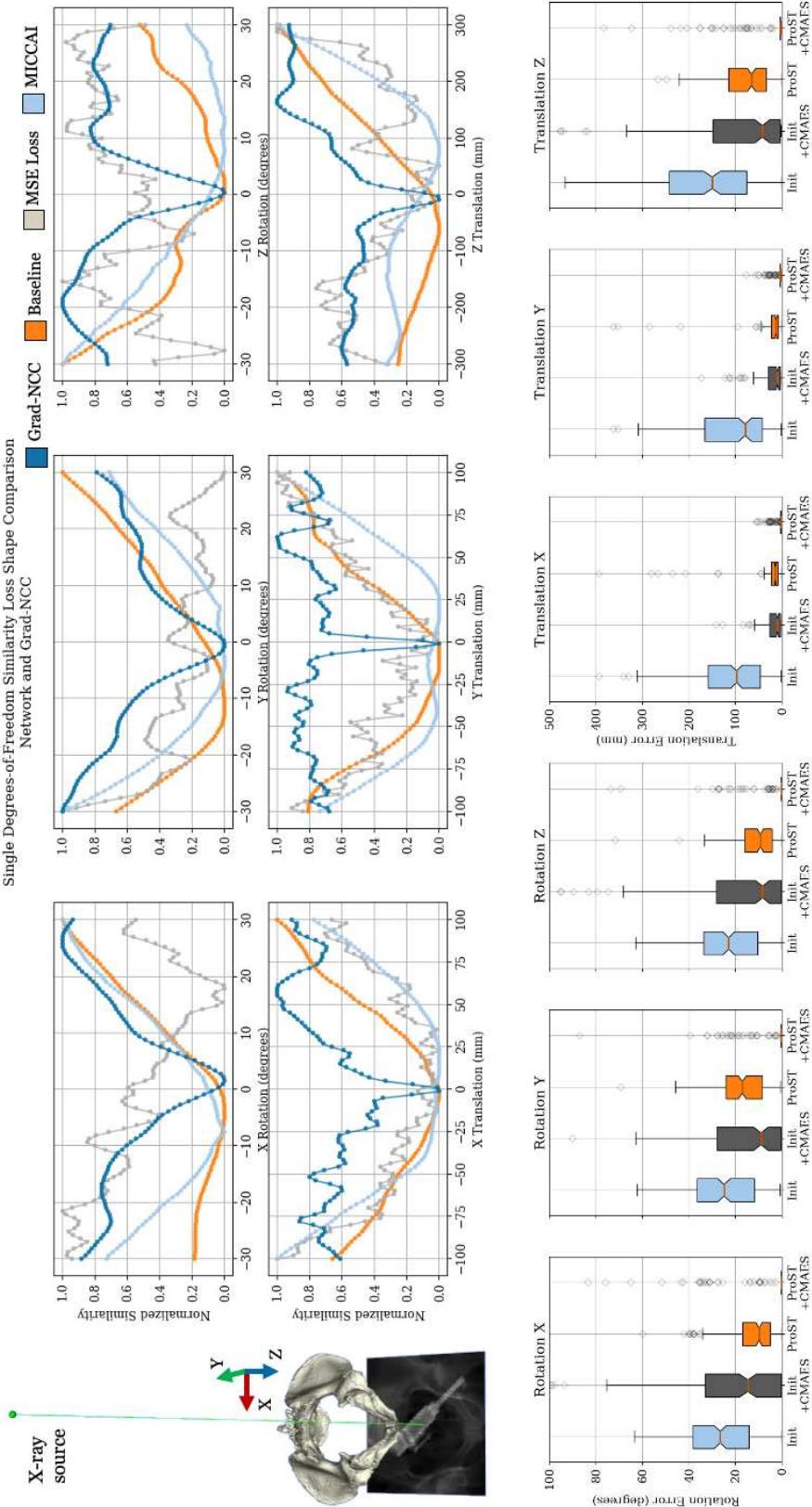
Note: SR refers to Success Rate. The highest success rate is bolded.

**Table 5.3:** Target Registration Error (TRE) and Success Rate of Spine

	TRE (mm)		SR (%)
	50% Med	75% Med	
Initialization	255.07	333.91	
CMAES	141.33	246.72	16.2
ProST Baseline	122.07	198.69	
CMAES	34.84	119.01	<b>47.1</b>

Note: The highest success rate is bolded.

## CHAPTER 5. PROST



**Figure 5.5:** Upper: Single Degrees-of-Freedom similarity loss shape comparison between network similarity and Grad-NCC. An example of projection geometry with axis directions is illustrated on the left. Bottom: Box plot of registration pose errors of baseline model in simulation decoupled to the individual degree of freedom.

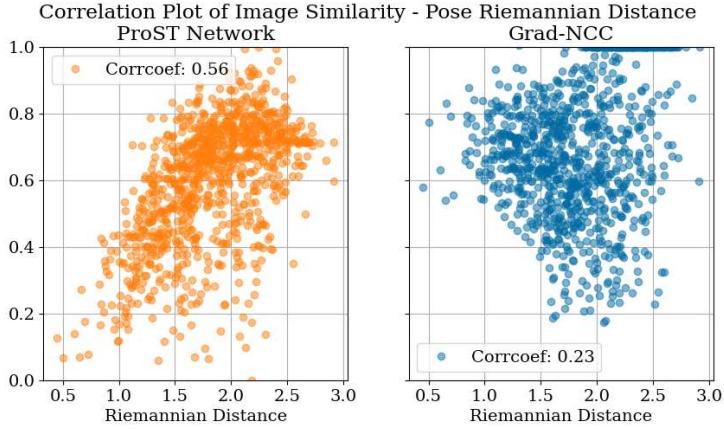
## 5.7 Discussion

Our results suggest that ProST largely increases the capture range of conventional CMAES intensity-based 2D/3D registration. This improvement is because the network similarity function has much fewer local minima compared to the hand-crafted image similarity functions, such as Grad-NCC. To our knowledge, this is the first time that the similarity function of X-ray image 2D/3D registration is learned to a target convex shape in an end-to-end fashion. Such breakthrough comes from the unique property of ProST: enabling differentiable gradient flow from the 2D domain to 3D, especially to the 3D pose parameters. We present a qualitative comparison of the network and Grad-NCC similarity function landscapes in Fig. 5.5. Since the similarity function is high dimensional with respect to the 6 DoF pose parameter, we plot each DoF individually by fixing the other DoF parameters as zeros. The similarity plots in Fig. 5.5 show that the network similarity is smoother and contains fewer local minima, especially in translations.

By learning a convex shape image similarity, the network similarity function has a higher correlation with respect to the registration error than the hand-crafted similarities. In Fig. 5.6, we present a correlation plot of the similarity values with respect to TRE. The correlation coefficient is 0.56 for the network and 0.23 for Grad-NCC. This suggests that our learned similarity can

## CHAPTER 5. PROST

be used as an indicator of registration uncertainty.



**Figure 5.6:** Correlation of similarity and pose Riemannian distance.

The current learning-based iterative 2D/3D registration methods in literature aim at predicting a pose update from 2D image observations by either framing it as a reinforcement learning task or direct gradient update regression. We included two typical methods in our ablation experiments, namely DeepIM [101] and DMW [57]. Such methods have the following limitations: 1) Since the registration does not optimize a cost function but the iterative pose prediction is extracted from 2D images, there is no guarantee for convergence to global or local minima. The prediction can explode if the image appearance alters by a small change. 2) Learning from only a 2D domain without 3D parameters limits the network ability to make predictions on 3D poses, and is likely to overfit to the training domain. Our ProST registration method follows the same iterative optimization paradigm as the intensity-based registration methods, where the only difference is that we take advantage of the great ex-

## CHAPTER 5. PROST

pressivity of the deep network to learn a set of more complicated filters than the conventional hand-crafted ones. This potentially makes generalization easier because the mapping that our method needs to learn is simple. In our ablation experiments, the success rates of ProST baseline model in simulation and real are 82.5% and 72.0%, respectively. DeepIM performs comparable to the baseline model in simulation with a success rate of 68.4%, but the performance drops substantially on real X-ray data with a success rate of 41.2%. When testing on images with tool overlays, the success rate of DeepIM was only 11.0%, which suggests the poor generalization ability of such methods.

To demonstrate the importance of 3D parameters in the architecture, we performed a controlled ablation study by removing the 3D CNN part and keeping the other training/testing settings the same. The results are worse than the baseline in Table. 5.1 and 5.2) for both the network and the following CMAES estimations. Although it is difficult to interpret the meaning of 3D features, the results suggest that the 3D part makes it easier to learn for this task.

Our architecture was trained with the novel double backward gradient loss. Our ablation experiment using MSE loss regression performs much worse than the baseline model with success rates of 28.3% in simulation and 28.6% in real data, which suggests that the network failed to learn meaningful mappings in this regression task. The double backward gradient loss design is a more effective way to learn the shape of target loss. We performed an ablation study to

## CHAPTER 5. PROST

show the effect of domain randomization (DR). The No DR results are slightly worse than the baseline model with success rates of 78.7% in simulation and 69.0% in real, which reveals that DR is useful to improve the model generalization ability. In this work, we used the advanced CrossViT as a 2D image encoder. The multi-scale patch-based feature extraction design learns more robust features compared to the ResNet block encoders used in the MICCAI published version [97]. Our ablation experiment of MICCAI architecture achieves success rates of 48.2% in simulation and 51.2% in real data, which are both worse than the CrossViT baseline model.

Single-view 2D/3D registration by nature has its ambiguity in depth translation and out-of-plane rotations. We plot the registration pose error distributions decoupled in individual DoF in Fig. 5.5. We have observed that ProST registration produces larger errors in translation Z (depth) compared to translation X and Y. This is because the appearance of the projection changes much less when the object translates along the depth Z direction compared to X or Y. It is particularly challenging for the network to detect such small differences from the target image using an encoder architecture compared to hand-crafted image similarities. This ambiguity can be visualized in the loss landscape plot in Fig. 5.5, where the network similarity presents flattened bottoms while Grad-NCC has a sharp curve around global minima. ProST network registration is particularly effective when the object is far away, but its accuracy

## CHAPTER 5. PROST

**Table 5.4:** Target Registration Error (TRE) and Success Rate of Pelvis

	Challenging View Real X-ray		
	TRE (mm)		SR (%)
	50% Med	75% Med	
Initialization	263.12	336.75	-
CMAES	149.04	228.09	20.6
ProST Baseline	119.08	155.96	-
CMAES	65.02	192.86	41.6
ProST Tool Overlay	117.77	160.51	-
CMAES	158.76	229.35	15.8
MICCAI	246.32	334.59	-
CMAES	124.15	208.14	33.6
DeepIM	315.16	496.17	-
CMAES	167.02	275.70	26.6
DMW	390.00	421.04	-
CMAES	186.27	209.82	10.0

Note: SR refers to Success Rate. The highest success rate is bolded.

is limited by this ambiguity when the pose is close to the target.

We tested the ablation study models on real pelvis X-rays of challenging views. Three examples of challenging view X-ray images are shown in Fig. 5.7. The pelvis in these types of images is only partially visible. Thus the present features were very different from the AP view images during training. Numeric registration results are present in Table. 5.4. They were all deteriorated from the standard view image results with the highest success rate of 41.6% corresponding to the ProST baseline model. This result suggests that the network similarity function is overfitted to training image appearance to a certain extent. Future work will include learning features that are more robust to drastic view changes.

## CHAPTER 5. PROST



**Figure 5.7:** Examples of challenging view pelvis X-rays.

We are aware that there are popular feature-based 2D/3D registration methods, such as solving a PnP problem using corresponding anatomical landmarks [58, 59, 89], which may achieve higher accuracy than our ProST registration. These methods need to manually define meaningful landmark features. Our method, however, learns the similarity function in a purely data-driven manner. We demonstrated our method using the spine as alternative anatomy. Pose estimation of the spine is more challenging than the pelvis because the spine is symmetric and the vertebrae are smaller in size. Using ProST registration, the success rate improves from 16.2% to 47.1%, which suggests its generalization ability on different anatomies.

In our experiments, the CT data were downsampled to a cubic volumetric size of 128 voxels in each dimension, and the projection images were of size  $128 \times 128$ , which consumes about 15 GB of GPU memory in pipeline training. Although the design grid sampler in ProST eases the excessive use of memory computation, it is still challenging to fill in an actual full-size CT data into a modern graphic card's memory. Such downsampling dropped information from

## CHAPTER 5. PROST

the original data and limited the registration accuracy. The iterative ProST registration takes about 30 seconds to converge, which is not feasible for real-time usage. Our future work includes improving the architecture design to fit data with higher resolution and speed up the registration inference time.

## 5.8 Conclusion

This work presents a learning-based 2D/3D registration pipeline, which learns a differentiable similarity function using a novel Projective Spatial Transformer module (ProST). We performed controlled studies using simulation and real X-ray images with intensive ablation experiments. Our results suggest that the proposed registration method largely extends the registration capture range and performs more robustly than the other learning-based pose regression methods in the literature. As a result, we believe that ProST has the potential to benefit relevant X-ray imaging applications.

## 5.9 Acknowledgement

I gratefully thank Dr. Xingtong Liu for his help designing ProST, neural network architecture, and double backward gradient loss. Sincere thanks to Ms. Anqi Feng for her help in data processing and ablation experiments. I

## CHAPTER 5. PROST

also thank Mr. Wenhao Gu and Mr. Benjamin Killeen for their helpful suggestions. Finally, my appreciation to Drs. Mathias Unberath, Russell Taylor, and Mehran Armand for their valuable feedback and support on this work.

# **Chapter 6**

## **Fluoroscopic Navigation of Rigid and Flexible Surgical Tools and Manipulators**

### **6.1 Introduction**

In this chapter, we present fluoroscopic image-based 2D/3D registration methods for navigation of rigid and flexible surgical tools and manipulators, explicitly referring to a robotic bone drilling and injection device and a snake-like continuum manipulator. These surgical tools are custom-designed for the orthopedic applications investigated in this dissertation. We use multiple view C-arm fluoroscopic images to estimate the intra-operative pose and shape of

## CHAPTER 6. SURGICAL TOOL NAVIGATION

the surgical tool with respect to the C-arm source frame. Combining the pose estimation of bone anatomy introduced in Chapter 4, the relative pose transformation between the surgical tool and bone anatomy can be computed. The surgical tool is then navigated to the planning position after registration. In this chapter, we introduce the design, use case, and registration methods for these two types of surgical devices in separate sections.

The injection device is a single rigid body. Therefore, its 2D/3D registration concept is similar to that for rigid bone anatomy, such as the pelvis (Chapter 4). However, the shape of the injection device is highly symmetric and made of metal, which produces fewer textures in 2D projection than the pelvis does. Thus, this ambiguity makes single-view 2D/3D registration of the injection device prone to large errors. To resolve this ambiguity problem, we propose using the robot's information. The robot arm can position the injection device to multiple configurations, and the relative poses can be computed using the robot forward kinematics. These configurations can be *jointly* registered to balance the single-view registration ambiguity. In this chapter, we present this idea of joint injection device registration for the injection device pose estimation and multi-view C-arm geometry estimation.

Pose estimation of the continuum manipulator is challenging due to its small size and high dexterity. In this chapter, we present a novel registration pipeline to estimate the intra-operative pose and shape of the continuum

## CHAPTER 6. SURGICAL TOOL NAVIGATION

manipulator using fluoroscopic images. This pipeline includes 1) methods that automatically detect the continuum manipulator in X-ray images, identifying landmark features that are used to initialize and regularize image registration; 2) a modified hand-eye calibration method that numerically optimizes the hand-eye matrix during image registration; 3) intensity-based deformable 2D/3D registration methods.

We performed intensive simulation, robotic system testing, and cadaver experiments to validate the proposed pose estimation approaches. In this chapter, we focus on the registration methods and present the simulation studies and results. We discuss our efforts on system-level calibration and clinical cadaveric studies in Chapter 7.

## 6.2 Contributions

The contributions of this chapter are as follows:

- We developed a multi-view intensity-based 2D/3D registration method for the injection device pose estimation using the robot information.
- We proposed an automatic 2D/3D registration pipeline for the continuum manipulator pose estimation. The pose initialization is computed from the extracted features in X-ray images of the continuum manipulator using a learning-based framework.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

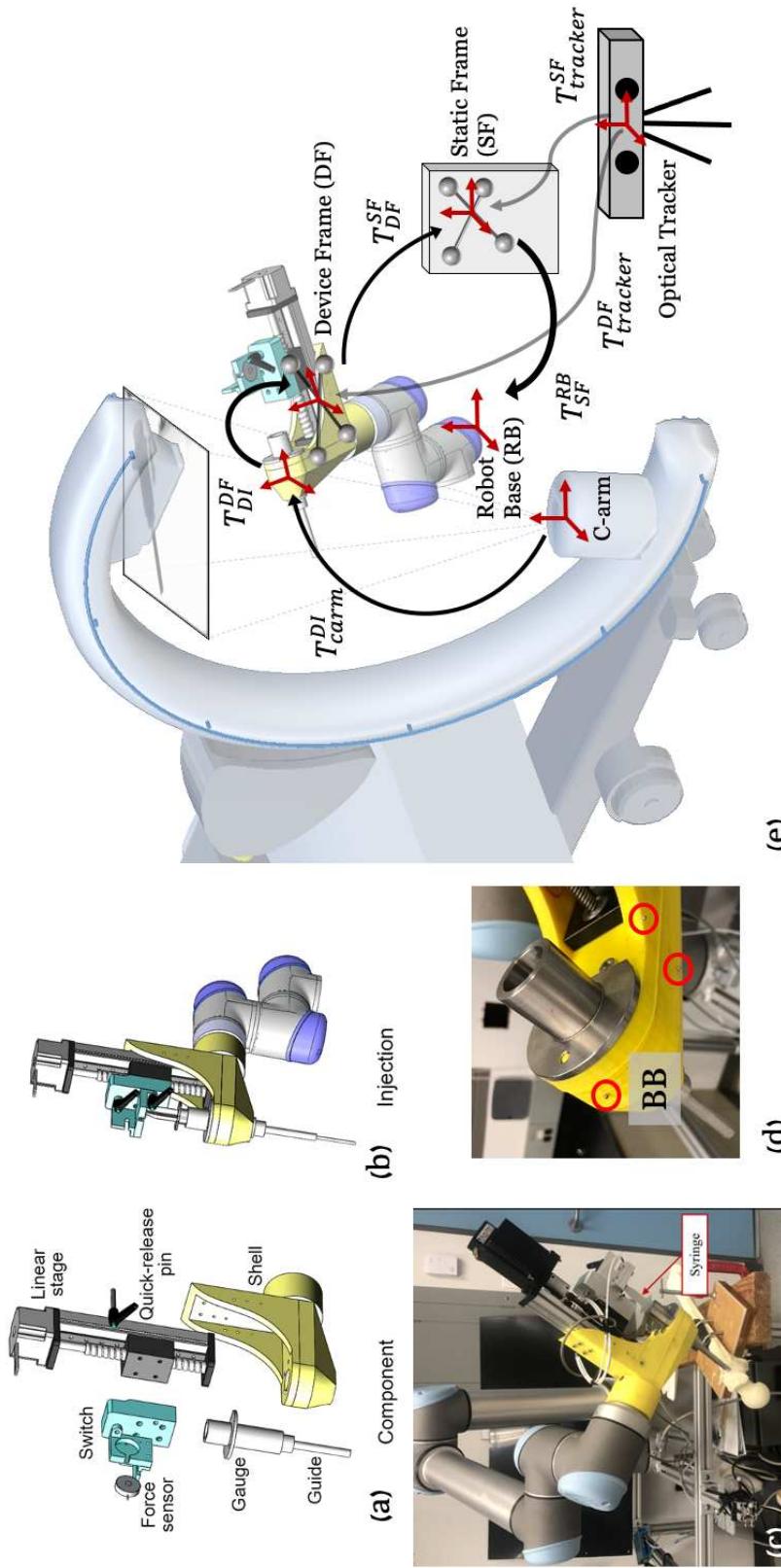
- We evaluated the proposed registration methods with comprehensive simulation experiments under numerous conditions.

Accurate pose and shape estimation of the surgical tool using external imaging is essential for robotic system navigation and control. Using fluoroscopic images, we present 2D/3D registration methods for pose estimation of a rigid surgical tool and a flexible continuum manipulator. The proposed methods are fiducial-free and achieve accurate pose and shape estimation by utilizing distinct surgical tool features from the X-ray image and the robot kinematics information. The surgical tool navigation methods have the potential to be generalized for more applications.

The navigation method of the rigid injection device for robot-assisted femoro-plasty is published in two articles of IEEE Transactions on Medical Robotics and Bionics [3,30]. The navigation method of the flexible continuum manipulator is presented at the 10th International Conference on Information Processing in Computer-Assisted Interventions (IPCAI 2019) [53], and is published as a featured article in the January 2022 issue of IEEE Transactions on Biomedical Engineering [32].

### 6.3 Rigid Drilling/Injection Device

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.1:** (a) Models of the integrated drilling and injection component. (b) Cement injection with an inserted syringe. (c) An example picture of the actual integrated robotic system for cement injection. (d) Picture of the metallic BBs glued on the surface of the injection device. Three example BBs are marked with red circles. (e) Illustrations of transformations of frames used for injection device registration. Key frames are marked with red arrows.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

The integrated Drilling and Injection component (DI) was designed and fabricated for the application of femoroplasty, which is capable of both guiding the drill trajectory and controlled the injection of the cement to the proximal femur [121]. For simplicity, we call it an injection device or DI in this dissertation. Fig. 6.1(a) shows the individual model of the injection device which consists of a supporter shell, a linear stage, a load cell, a gauge with a cylinder guide, and a switch block to interchange between drilling and injection<sup>1</sup>. A 3D printed shell attached to the end-effector of the robot is used as a supporter for other parts of the DI. A linear stage provides the push force on the plunger of the syringe and serves as a hard stop for drilling. A gauge made of stainless steel is the supporter of the syringe and the slender cylinder of the gauge is used as a guide for both the drill bit and cannula (both are 4 mm). The length of the guide is 75 mm and it is placed on the surface of the tissue of the femur. This length keeps a safe distance from the DI shell to human tissue accounting for changes in the soft tissue thickness. Fig. 6.1(b) and (c) present the model of the integrated DI for injection, and an actual picture of the integrated robotic injection system, respectively. We took CT scans of the DI and cropped the volume of the metallic gauge and guide for 2D/3D registration. To obtain ground truth poses of the DI, metallic BBs were glued onto the surface of the device as shown in Fig. 6.1(d).

---

<sup>1</sup>Pictures of (a)(b)(c) are from manuscript [121].

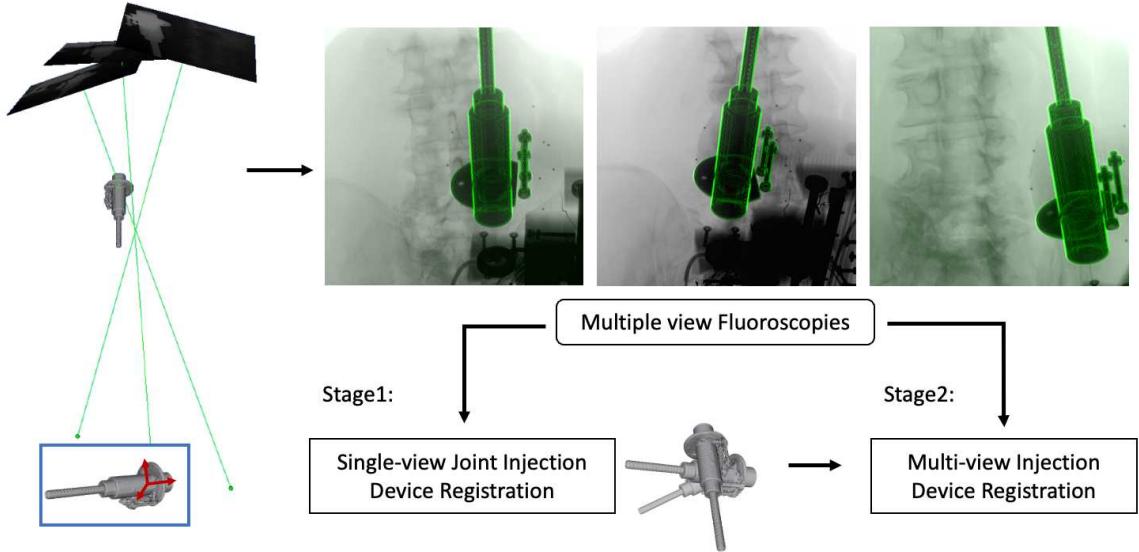
## CHAPTER 6. SURGICAL TOOL NAVIGATION

The injection device is attached to a 6-DOF UR-10 (Universal Robots, Odense, Denmark). Unfortunately, the forward kinematic accuracy of the UR-10 robot is insufficient to meet the clinical accuracy requirement. This insufficiency is further amplified by the weight of the injection unit. To ameliorate these inaccuracies, and NDI Polaris (Northern Digital Inc., Waterloo, Ontario, Canada) system was used to achieve closed-loop position control of the robotic system. Key transformations of the robotic system are illustrated in Fig. 6.1 (e). Hand-eye calibrations are required to achieve closed-loop navigation, which estimates transformation between the robot base frame and static marker frame ( $T_{SF}^{RB}$ ), and transformation between the injection device and the feed-forward control fiducial frame ( $T_{DI}^{DF}$ ). We present the hand-eye calibration methods in Chapter. 7.3.1. In this Chapter, we focus on the registration methods.

### 6.3.1 Injection Device Registration Methods

Intra-operative pose estimation of the injection device was achieved using multi-view X-ray image-based 2D/3D registration. The general workflow is shown in Fig. 6.2. The robot base was fixed relative to the patient bed. The C-arm was positioned at multiple geometric views with separate angles of 40-50 degrees. At each C-arm view, we positioned the robotic injection device at varied configurations within the C-arm capture range. Fluoroscopic images of the injection device were taken for each injection device pose. These robot

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.2:** Injection device registration pipeline: Multiple view fluoroscopy projection geometries and real injection device X-ray images with DRR-derived edges in green are presented. The rotation center frame is set at the center of the guide tube, which is illustrated using red cross arrows at the bottom left corner.

configurations were saved and kept the same while the C-arm was positioned at different views. These images were used for 1) stage 1: single-view joint injection device registration to estimate the C-arm view geometry; 2) stage 2: multi-view registration for injection device pose estimation. We introduce the algorithmic details of these two registration stages in the following subsections.

### 6.3.1.1 Joint Injection Device Registration

Intensity-based 2D/3D registration of the injection device was performed at each C-arm view. It optimizes a similarity metric between the target X-ray image and a DRR image simulated from the 3D injection device volume ( $V^{DI}$ ).

## CHAPTER 6. SURGICAL TOOL NAVIGATION

Given  $J$  tracker observations  $T_{DF_1}^{SF}, T_{DF_2}^{SF}, \dots, T_{DF_J}^{SF}$ , the hand-eye calibration matrix  $T_{DI}^{DF}$ , the injection device poses in the static frame are  $T_{DI_j}^{SF} = T_{DI}^{DF} \cdot T_{DF_j}^{SF}, j \in \{1, \dots, J\}$ . We used the first pose as reference and the rest of the poses can be computed relative to the reference pose using  $T_{DI_0}^{DI_j} = T_{DI}^{DF} \cdot T_{DF_j}^{SF} \cdot (T_{DF_0}^{SF})^{-1} \cdot (T_{DI}^{DF})^{-1}$ . Given an X-ray image  $I_k^{Dj}$  (the  $k^{\text{th}}$  C-arm view and the  $j^{\text{th}}$  injection device pose), a DRR operator ( $\mathcal{P}$ ), a similarity metric ( $\mathcal{S}$ ), the joint 2D/3D registration estimates the injection device pose ( $T_{Carm_k}^{DI_0}$ ) by solving the following optimization problem:

$$\min_{T_{Carm_k}^{DI_0} \in SE(3)} \sum_{j=1}^J \mathcal{S}(I_k^{Dj}, \mathcal{P}(V^{DI}; T_{Carm_k}^{DI_0}, T_{DI}^{DF}, T_{DF_j}^{SF})) \quad (6.1)$$

The similarity metric ( $\mathcal{S}$ ) was chosen to be patch-based normalized gradient cross correlation [92]. The 2D X-ray image was downsampled four times in each dimension. The optimization strategy was selected as CMAES [95]. The registration gives an accurate injection device pose estimation at each C-arm view ( $T_{Carm_k}^{DI_0}$ ).

### 6.3.1.2 Multi-view Injection Device Registration

Given the first C-arm view as reference, poses of the rest C-arm views can be calculated using  $T_{Carm_k}^{Carm_0} = (T_{Carm_0}^{DI_0})^{-1} \cdot T_{Carm_k}^{DI_0}, k \in \{1, \dots, K\}$ , where  $K$  is the total number of C-arm views. We then performed a multi-view injection device

## CHAPTER 6. SURGICAL TOOL NAVIGATION

registration to estimate the reference injection device pose in the reference C-arm view ( $T_{Carm_0}^{DI_0}$ ) by solving the optimization:

$$\min_{T_{Carm_0}^{DI_0} \in SE(3)} \sum_{k=1}^K \mathcal{S}\left(I_k^{D_0}, \mathcal{P}(V^{DI}; T_{Carm_0}^{DI_0}, T_{Carm_0}^{Carm_k})\right). \quad (6.2)$$

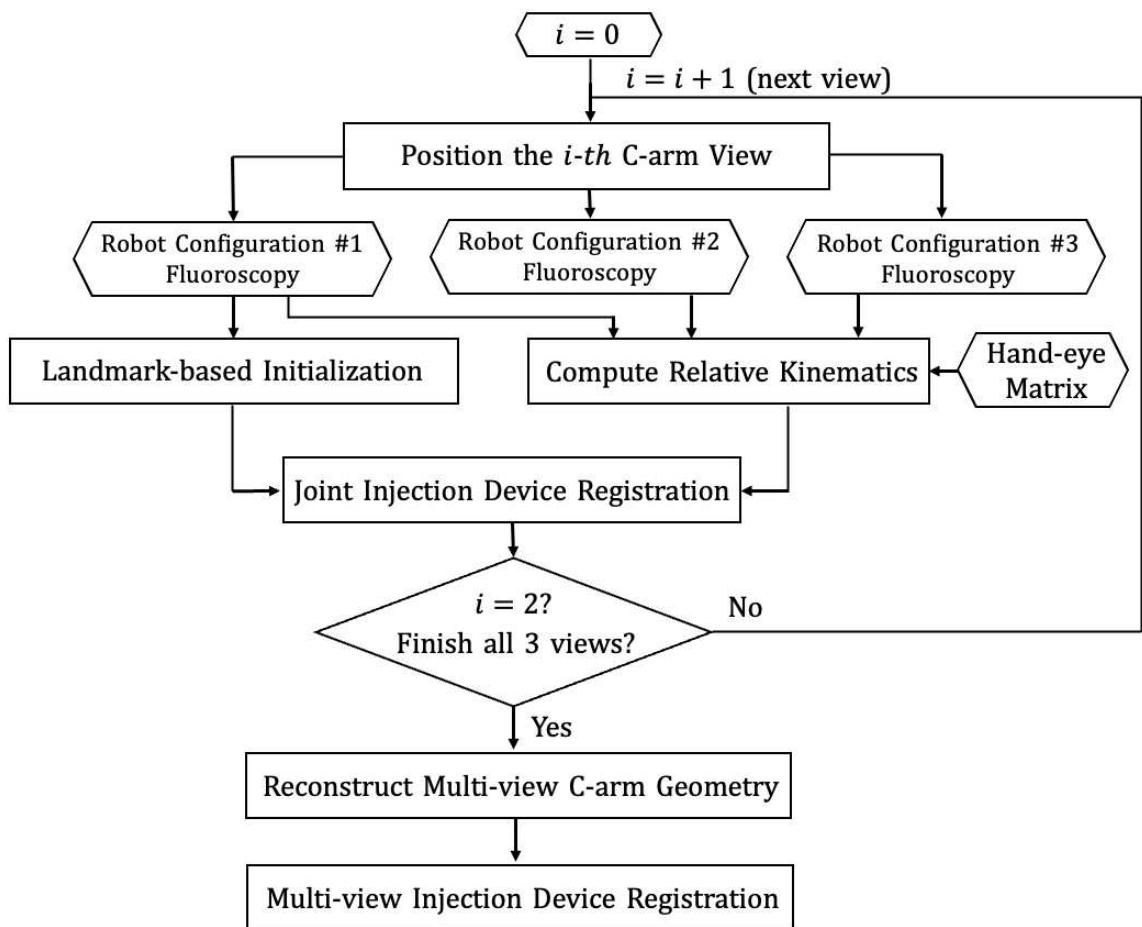
We used the same similarity metric, image processing, and optimization strategy as introduced in the joint injection device registration.  $T_{Carm_0}^{DI_0}$  derived from multi-view registration further refined the result of joint registration under single-view. We present a step-by-step injection device registration workflow in Fig. 6.3.

### 6.3.2 Experiments and Results

#### 6.3.2.1 Image-based Registration Simulation Study

We tested the injection device 2D/3D registration performance under various settings, including single-view and multi-view C-arm geometries. 1,000 simulation studies were performed with randomized poses of the injection device for each registration workflow. We defined the reference frame of the injection device model at the center of the injector tube. The registration pose,  $T_{Carm}^{DI}$  as defined in Section 6.3.1 refers to the rigid transformations from the simulated C-arm source to the reference frame. We report the registration accuracy based on our simulated ground truth poses of the objects using

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.3:** Workflow for injection device registration.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

$\delta T_{Carm}^{obj} = (T_{Carm}^{obj})_{gt} \cdot (T_{Carm}^{obj})_{regi}^{-1}$ ,  $obj \in \{DI, v\}$ , where  $gt$  and  $regi$  refer to ground truth and registration estimation, respectively. We described the detailed simulation setup in the following subsections. Numeric results and statistical plots are presented in Table 6.1 and Fig. 6.4.

### 6.3.2.1.1 SINGLE-VIEW REGISTRATION

We performed 2D/3D registration of the injection device by simulating single-view X-ray images. We also tested the proposed joint injection device registration by simulating three variant robot configurations under single-view C-arm geometry. These three device poses are jointly registered using the simulated relative robot configurations. For every registration running, random initializations of the injection device were uniformly sampled including translation from 0 to 10 mm and rotation from -10 to 10 degrees. Table 6.1 summarizes the magnitudes of translation and rotation errors. The single injection device registration errors were  $2.15 \pm 1.57$  mm and  $1.62 \pm 1.40$  degrees, respectively. We achieved lower errors using joint injection device registration, which was  $1.73 \pm 1.17$  mm and  $0.91 \pm 0.92$  degrees, respectively.

### 6.3.2.1.2 MULTI-VIEW REGISTRATION

Three multiple C-arm pose geometries were estimated with a uniformly sampled random separation angle between 20 and 25 degrees for the two side

## CHAPTER 6. SURGICAL TOOL NAVIGATION

**Table 6.1:** Mean Registration Error in Injection Device Registration Simulation Study.

		Translation Error (mm)	Rotation Error (degrees)
Single-View	Injection Device	$2.15 \pm 1.57$	$1.62 \pm 1.40$
	Joint Injection Device	$1.73 \pm 1.17$	$0.91 \pm 0.92$
	P value	$< 0.00001$	$< 0.00001$
Multi-View	Injection Device	$0.17 \pm 0.60$	$1.21 \pm 1.31$
	P value	$< 0.00001$	$< 0.00001$

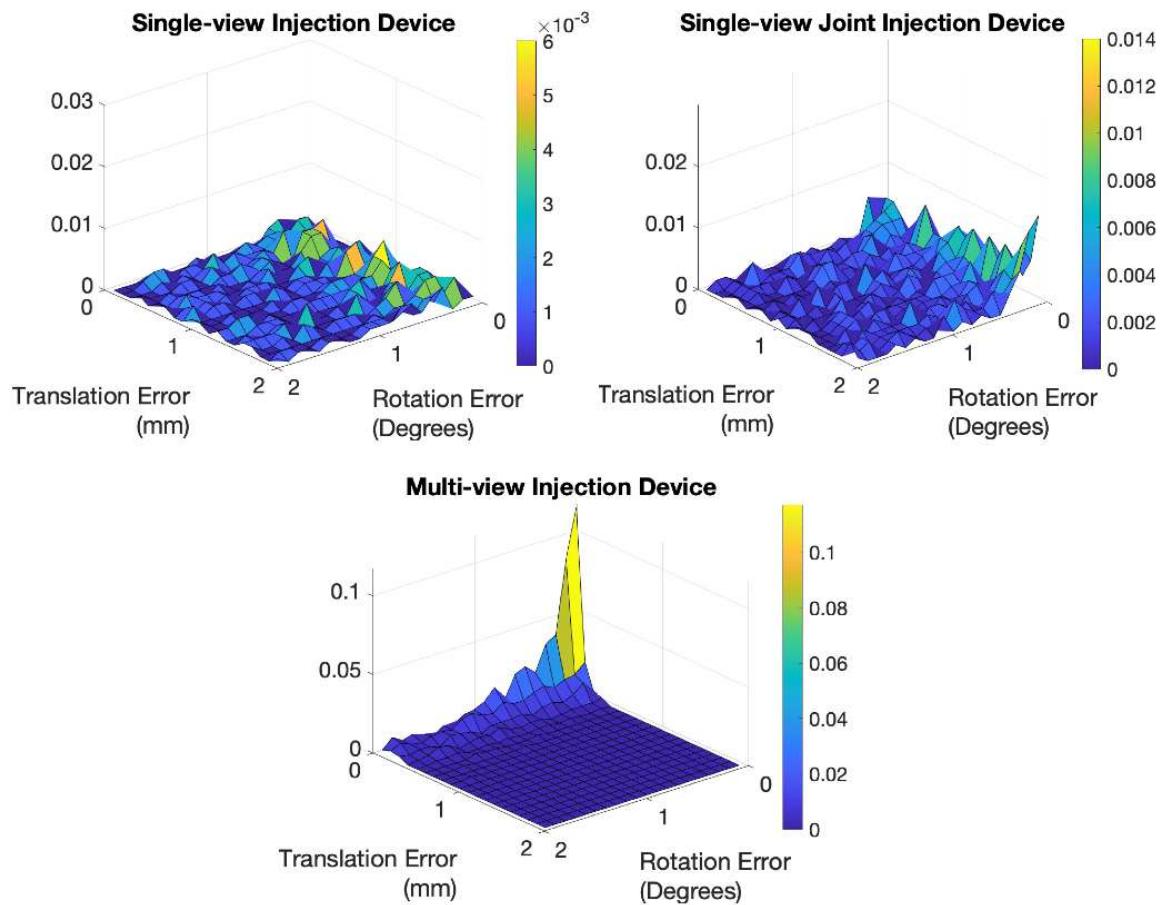
The p value is reported from T-Test score compared to single-view injection device registration results. The result is significant at  $p < 0.05$ .

views. The three registration workflows tested in single-view were performed with the same settings under this multi-view setup. The registration accuracy was improved. The registration error was  $0.17 \pm 0.60$  mm and  $1.21 \pm 1.31$  degrees in translation and rotation, respectively. A joint histogram of the translation and rotation errors is presented in Fig. 6.4. From the plots, we clearly observed the multi-view injection device registration has better error distribution with the cluster close to zero errors.

### 6.3.3 Discussion

Our proposed registration pipeline achieves accurate pose estimation of the injection device in simulation. Our simulation result has shown that multi-view registration is significantly more accurate and stable than single-view registration (Table. 6.1). This is because multi-view projection geometries fundamentally improve the natural ambiguity of single-view registration. Our

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.4:** Normalized 2D histograms of injection device registration pose error ( $\delta T_{Car_m}^{ID}$ ) reported in joint magnitudes of translation and rotation.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

proposed joint injection device registration also improves from single-view registration. In simulation, we have shown that this method is superior to single injection device registration in accuracy: the mean errors decreased from 2.15 mm, 1.62 degrees to 1.73 mm, 0.91 degrees in translation and rotation, respectively. This is because the joint registration of multiple injection devices poses balances the ambiguity of single-view single object registration. The injection device was positioned close to the C-arm rotation center so that the C-arm detector can be rotated to large side angles while maintaining both the injection device and the spine vertebrae within the capture range.

The joint injection device registration takes about 20 seconds and the multi-view injection device registration takes about 30 seconds. Considering using three C-arm view X-rays, the full registration pipeline takes around 2 minutes running on an NVIDIA GeForce RTX 2080 graphics processor. This time is acceptable for a one-time intra-operative robotic device registration. This injection device supplies a positioning device for both applications of femoroplasty and transforaminal lumbar epidural injections. In Chapter 7, we will present our efforts on system-level integration and clinical experiments using the registration methods developed in this Chapter.

## 6.4 Flexible Continuum Manipulator

Continuum manipulators (CMs) have the potential to advance minimally-invasive surgical procedures due to their high dexterity and enhanced accessibility [122]. Robotic systems equipped with CMs have been studied in the context of soft environment surgical applications such as percutaneous intracardiac surgery [123], fetoscopic interventions [124], laryngeal surgery [125], gastroscopy [126] and endoscopic orifice surgery [127–129].

The application of CMs for soft tissue surgery has attracted a great amount of research energy. However, similar applications in orthopedic surgery have been limited. The stiffness needed for cutting and debriding harder tissues such as bone contrasts with the shape-compliance inherent to CMs, presenting design challenges and further complicating the shape sensing and control of CMs. Despite these challenges, CMs may be very useful in the surgical treatment of bone defects, such as femoroacetabular impingement (FAI), metastatic bone disease, and severe osteoporosis in areas including the pelvis/acetabulum, femoral neck, peri- and sub-trochanteric regions, and traumatic fracture repair. For example, in core decompression of the hip, which consists of using a drill to remove the 8-10 mm cylindrical core from an osteonecrotic lesion [44]. Complete removal of a lesion in the femoral head requires access by drilling through the narrow femoral neck and then debriding a larger volume of necrotic bone

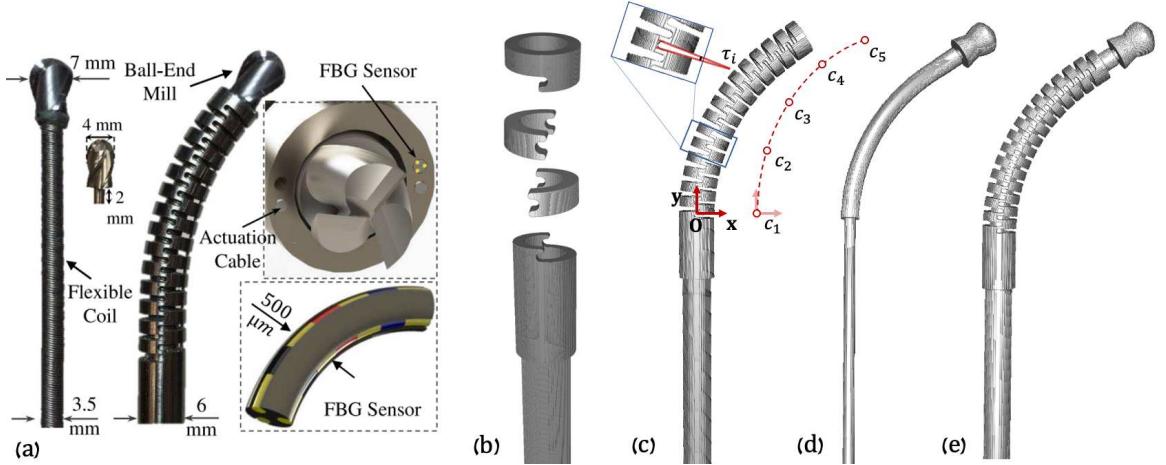
## CHAPTER 6. SURGICAL TOOL NAVIGATION

beyond this access point. High accuracy is needed to remove the minimal healthy bone so as to maintain structural integrity and stability, especially in the narrowest regions of the femoral neck. However, to debride a larger volume once having passed through the narrow femoral neck, significant dexterity of the tool is required. CMs with embedded shape sensing has been demonstrated to meet these requirements [63, 130].

Intra-operative navigation and shape sensing of the CMs are key challenges to achieving the clinical goal. Because of the small size and dexterity of the CM, accurate shape and pose estimation is very challenging. Otake et al. first applied fluoroscopic image-based 2D/3D registration to estimate the CM pose and shape [131]. However, there were several major limitations: 1) the algorithm was only tested with simulation and camera images instead of real fluoroscopic images; 2) the registration was manually initialized close to the global optima; 3) the method only used a single-view image method, which leads to large ambiguity.

In this Chapter, we present a custom-designed continuum manipulator for orthopedic applications used in this dissertation work. We present our novel X-ray image-based navigation solution for a surgical robotic system with a CM. We introduce the CM modeling, navigation methods, and our simulation experiments and results in the following sections.

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.5:** (a) Figure of the continuum manipulator with flexible cutting/debriding tool [5]. The actuation cables and the FBG sensor are demonstrated. (b)-(e) CM Model. (b) Basic model components including base, notch and top segments. (c) An example CM configuration. The notch joint angle is illustrated as  $\tau_i$ . The cubic spline control points are shown aside as  $c_1, c_2, \dots, c_5$ . The CM base reference frame is shown in RGB cross arrows. (d) An example flexible tool model. (e) Integrated CM with flexible tool inside.

### 6.4.1 Continuum Manipulator Model

The CM used in our surgical robotic system is constructed from a hollow superelastic nitinol tube with 27 alternating notches on two sides, using a 4 mm diameter lumen as a flexible instrument channel, with an outside diameter of 6 mm and a total length of 35 mm. It achieves flexibility and compliance in the direction of bending while preserving high stiffness in the perpendicular direction to the bending plane [132–134]. Two stainless steel cables are embedded through channels on two opposing sides of the CM wall to provide bidirectional planar motion [132, 133]. FBG fibers are integrated into the CM wall channels. The 3D shape of the CM can be inferred in real-time from FBG

## CHAPTER 6. SURGICAL TOOL NAVIGATION

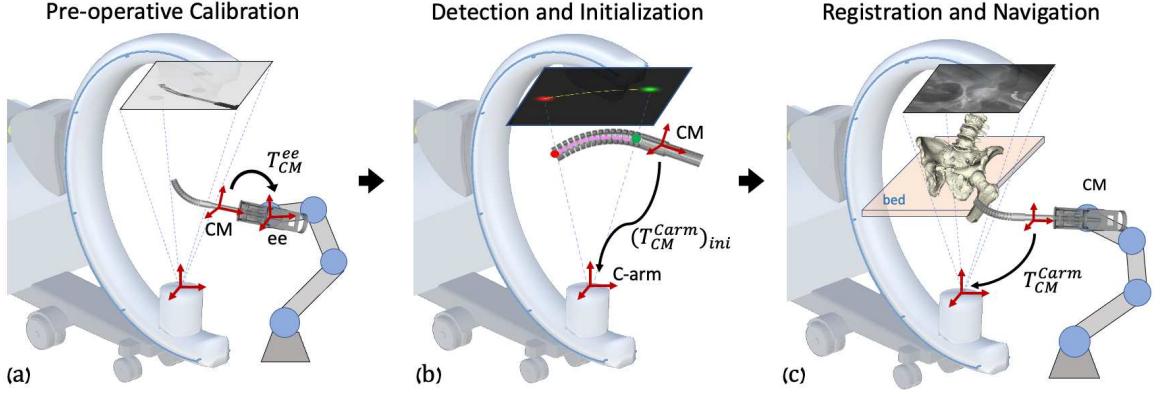
readings [135, 136]. The CM, tool, and integrated sensor are shown in Fig. 6.5 (a).

The CM kinematics configuration is determined by the notch joint angles. Following previous work on a kinematic modeling of this CM [131], we assume that the joint angle changes smoothly from one joint to the next. Angles are parameterized as a cubic spline of  $n$  equally distributed control points,  $\tau_i$ , along the central axis of the CM ( $n = 5$ ). The model of the CM used in the simulation is built from component volumetric models that are aligned according to the CM kinematic model. The origin of the CM model base reference frame is at the center of the structure between the first notch and the base. The y-z plane defines the CM bending plane and the x-axis is perpendicular to the bending plane. Fig. 6.5 (b)-(e) illustrates how the CM model is built and defines the reference frame.

### 6.4.2 Navigation Methods

Fig. 6.6 presents the overall pipeline of fluoroscopic image-based navigation of the CM. It consists of three stages: 1) *Pre-operative calibration*: The transformation from the CM model reference frame to the rigid-link robot end-effector frame is calibrated using X-ray image-based hand-eye calibration (Fig. 6.6 (a)). 2) *Detection and Initialization*: Distinct features of the CM are automatically detected in 2D X-ray images. An initial 3D pose of the CM ( $(T_{CM}^{C_{arm}})_{ini}$ ) is esti-

## CHAPTER 6. SURGICAL TOOL NAVIGATION



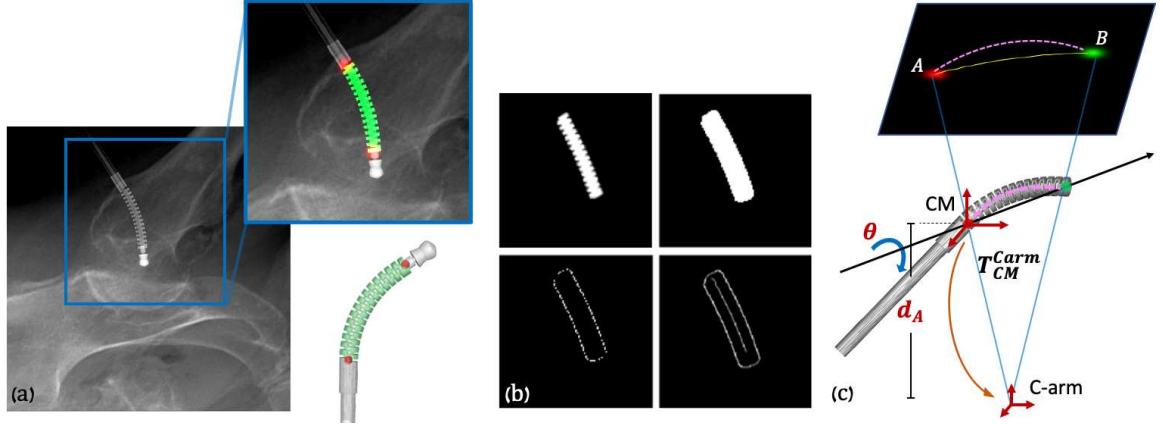
**Figure 6.6:** Illustration of the proposed image-based navigation for the CM. Key frames are shown with red cross arrows. (a) Pre-operative hand-eye calibration of the CM. The hand-eye matrix  $T_{CM}^{ee}$  is represented by the curved arrow. (b) CM feature detection and centerline-based initialization. 2D landmark heatmaps and corresponding 3D landmarks are shown in red and green. 2D and 3D centerlines are shown in yellow and pink, respectively. The CM initial pose estimation  $(T_{CM}^{Carm})_{ini}$  is marked with a black arrow. (c) Scheme of intensity-based 2D/3D registration of the CM. The registration transformations are described in Section II-C in detail.

mated using centerline-based 2D/3D registration (Fig. 6.6(b)). 3) *Registration and Navigation:* An accurate 3D pose and shape estimation of the CM ( $T_{CM}^{Carm}$ ) is achieved by intensity-based 2D/3D registration using X-ray images and 3D CM model (Fig. 6.6(c)).

### 6.4.2.1 CM Detection and Pose Initialization

The small size, symmetric structure, and dexterity of the CM make it particularly challenging to estimate its pose and shape using a fluoroscopic image. To this end, we propose to exploit semantic prior information on the imaged object, the CM, to heavily constrain rigid initialization and the deformable reg-

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.7:** (a) 3D segmentation label and landmark positions on an example CM model. (b) An example simulation image using DeepDRR. (c) Overlay of the 2D segmentation mask and landmark heatmaps on the cropped image in (b). (d) Extraction of the 2D centerline from a segmentation mask consists of dilation, edge, and distance map computation, and finally, shortest path extraction. (e) A simplified illustration of the centerline-based 2D/3D registration geometry.

istration. We train a convolutional neural network (ConvNet) to segment the CM in X-ray images and simultaneously detect its start and end point using multi-task learning [53, 59, 137], yielding an estimate of the projection domain centerline of the CM. Our segmentation target region covers the 27 alternating notches which discern the CM from other surgical tools (Fig. 6.7(a)). The two landmarks are defined as 1) the origin of the CM base reference frame and 2) the center of the distal plane of the last notch, start and end point of the CM centerline (Fig. 6.7(a)). Inspired by [138], the 2D centerline is extracted using morphological processing of the segmentation mask (Fig. 6.7(b)). Fig. 6.7 illustrates these features in an example 2D image and 3D model.

We have introduced this CM detection task in Chapter 3, surgical robotic

## CHAPTER 6. SURGICAL TOOL NAVIGATION

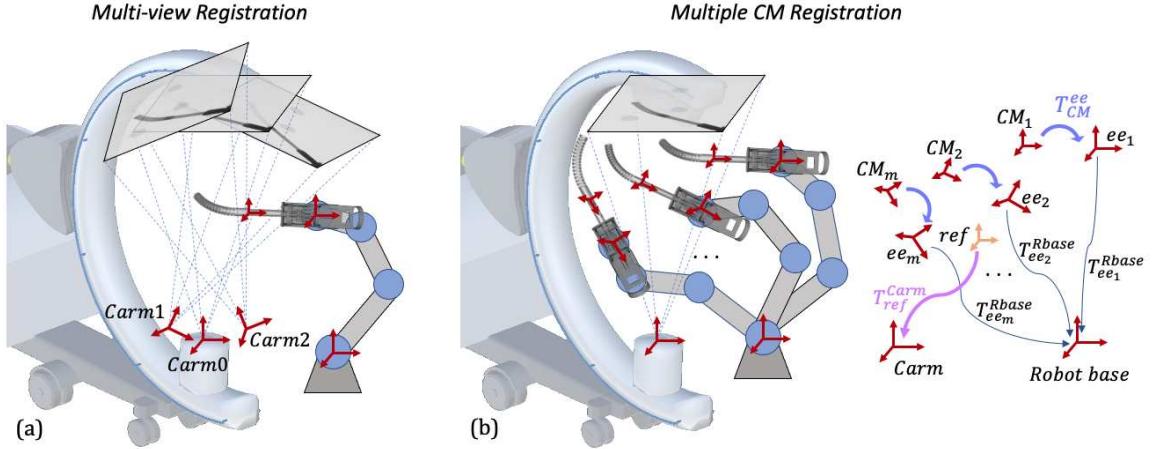
tool detection. Details of the network architecture, training, and data generation are present in Chapter. 3.4.2.

The detected semantic features combined with the known C-arm projection geometry and the CM curvature estimation from FBG readings enable an initial 3D pose estimation of the CM using centerline-based 2D/3D registration. Since two 2D landmarks only determine 4 of the required 6 DOF of the CM base rigid transformation, the other two DOFs can be decomposed as the depth distance of the 3D CM base landmark ( $d_A$ ) and the rotation angle ( $\theta$ ) about the axis passing through the two 3D CM landmarks (Fig. 6.7 (c)). We then perform an optimization search of this two variables ( $d_A, \theta$ ) by minimizing the distance between the extracted centerline and the reprojection of the 3D model centerline. Once the global optimal ( $d_A, \theta$ ) is found, the rigid CM poses in the C-arm frame ( $T_{CM}^{Carm}$ ) can be determined, which is used as the initialization of the intensity-based registration. We present more details of this algorithm in Appendix. B. Since the centerline feature is approximate, we expect to see some ambiguity in depth and rotation which is going to be resolved using intensity-based registration.

### 6.4.2.2 Image-based 2D/3D Registration

Image-based 2D/3D registration of the CM estimates both the rigid pose of the CM relative to the C-arm source frame ( $T_{CM}^{Carm} \in SE(3)$ ) and the deformable

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.8:** (a): Illustration of multi-view registration. Three example C-arm view source frames are noted with red cross arrows. (b): Illustration of multiple CM registration. The transformations are marked separately on the right and described in Section II-C.

kinematics configuration ( $\tau_i, i \in \{1..5\}$ ). The initial estimation of  $T_{CM}^{Carm}$  is from the centerline-based registration, and the initial  $\tau_i$  is from the FBG readings. Our proposed intensity-based registration of the CM is performed by optimizing an image similarity score, combined with a landmark reprojection penalty. DRRs are created by calculating ray casting line integrals through the CM model onto a 2D image plane following the C-arm projection geometry. The similarity score is calculated between each DRR and the intraoperative image  $I$ . The landmark reprojection penalty is defined as the sum of  $l_2$  distance between each detected landmark position from image  $I$  and the projected landmark position from the 3D model.

Given  $K$  images ( $I_k$ ) as input, using  $J$  CM model segment volumes ( $V_j, j \in \{1..J\}, J = 27$ ), a DRR operator ( $\mathcal{P}$ ), similarity metric ( $\mathcal{S}$ ), and regularizer over

## CHAPTER 6. SURGICAL TOOL NAVIGATION

plausible poses ( $\mathcal{R}$ ), the registration recovers the CM rigid pose ( $T_{CM}^{Carm}$ ) and deformation control points ( $\tau_i$ ) by solving the following optimization problem:

$$\begin{aligned} \min_{T_{CM}^{Carm} \in SE(3), \tau_i} & \sum_{k=1}^K \mathcal{S}\left(I_k, \mathcal{P}\left(\sum_{j=1}^J V_j; T_{CM}^{Carm}, \tau_i\right)\right) \\ & + \mathcal{R}(T_{CM}^{Carm}, \tau_i), i \in \{1..5\} \end{aligned} \quad (6.3)$$

We use patch-based normalized gradient cross correlation as the similarity score [92]. The 2D image is first cropped to a  $500 \times 500$  pixel region of interest (ROI) using the two landmark locations and downsampled 4 times in each dimension. We use CMAES optimization strategy [95]. The two detected landmarks are incorporated into a reprojection regularizer defined as

$$\begin{aligned} \mathcal{R}(T_{CM}^{Carm}, \tau_i) = & \frac{1}{2\sigma_l^2} \sum_{k=1}^K \|\mathcal{P}(p_{3D}^{(k,l)}; T_{CM}^{Carm}, \tau_i) - p_{2D}^{(k,l)}\|^2, \\ & i \in \{1..5\}, l \in \{A, B\} \end{aligned} \quad (6.4)$$

The registration produces the rigid pose of the CM in the C-arm frame ( $T_{CM}^{Carm}$ )<sub>regi</sub> and the deformation control point values ( $\tau_i$ ).

Since the CM object is deformable and symmetric about its central spine, single-view 2D/3D registration of the CM suffers from depth ambiguity and axial rotation ambiguity. We also investigate multiple view 2D/3D registration of the CM to remedy the single-view ambiguity. There are two different ways to introduce multiple projection geometries: One is moving the C-arm to multiple

## CHAPTER 6. SURGICAL TOOL NAVIGATION

views to register a static CM pose, which we phrase as *multi-view registration* (Fig. 6.8 (a)). The other one is moving the CM to multiple poses under a single C-arm view, which we phrase as *multiple CM registration* (Fig. 6.8 (b)). One key challenge of using multiple images is the relative geometry estimation among these poses, which is essentially a calibration problem. We have presented methods to estimate multi-view C-arm geometry using single-view registration of a “fiducial” anatomy, the pelvis, in Chapter 4. However, to keep the pelvis in the C-arm field of view, the multiple C-arm view geometries need to be close in rotation, limiting the differences in image appearance. In multiple CM registration, however, the CM poses can be more diverse by moving the rigid-link robot. Relative CM pose geometries can be calculated using the rigid-link robot forward kinematics ( $T_{ee}^{Rbase}$ ) and the hand-eye matrix ( $T_{CM}^{ee}$ ). Details of the hand-eye calibration methods are presented in Chapter 7.4.1.

As shown in Fig. 6.8(b), given  $M$  static rigid-link robot forward kinematics  $T_{ee_1}^{Rbase}, T_{ee_2}^{Rbase}, \dots, T_{ee_M}^{Rbase}$  and a hand-eye matrix  $T_{CM}^{ee}$ , the CM poses in the static rigid-link robot base frame are  $T_{CM_m}^{Rbase} = T_{ee}^{Rbase} \cdot T_{CM_m}^{ee}, m \in \{1..M\}$  with the fact that all configurations share the same hand-eye matrix. We then define a central reference frame with respect to the robot base frame. The reference frame is chosen to be the average position of multiple CM base origins. Its orientation is identity and its translation is the mean translation of all CM poses. The transformation from the reference frame to the C-arm frame is

## CHAPTER 6. SURGICAL TOOL NAVIGATION

noted as  $T_{ref}^{Carm}$ . Each CM poses with respect to the reference frame ( $T_{CM_m}^{ref}$ ) can be derived by the reference frame definition. We assume the CM shape is fixed during registration. Since the C-arm frame is static,  $T_{ref}^{Carm}$  determines all the CM rigid poses in the C-arm frame. Following the formulation of (6.3), the multiple CM registration can be defined using:

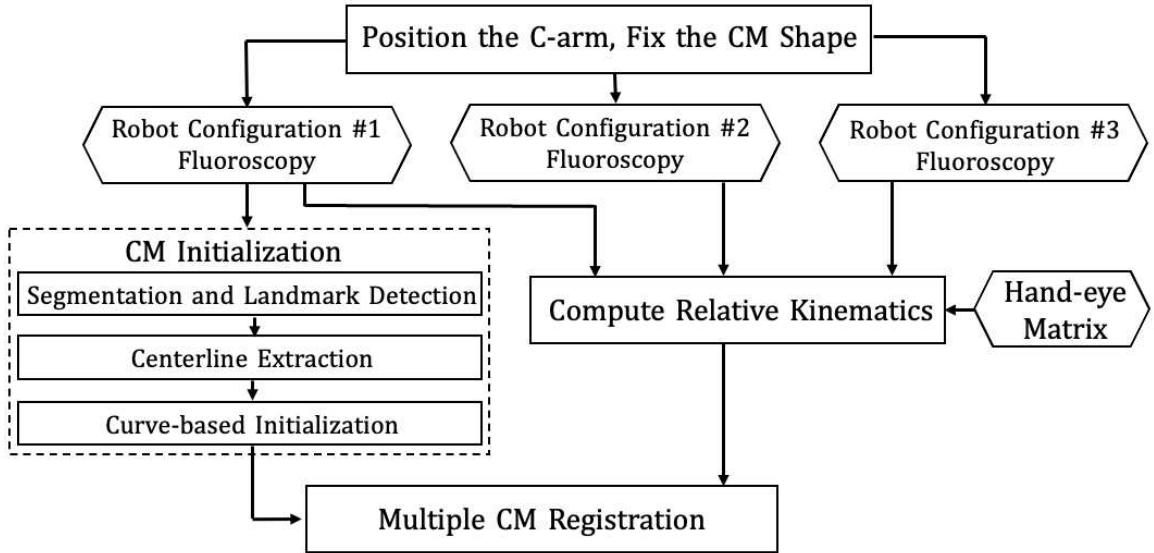
$$\begin{aligned} \min_{T_{ref}^{Carm} \in SE(3), \tau_i} & \sum_{m=1}^M \mathcal{S} \left( I_m, \mathcal{P} \left( \sum_{j=1}^J V_j; T_{ref}^{Carm}, \tau_i \right) \right) \\ & + \mathcal{R}(T_{ref}^{Carm}, \tau_i), i \in \{1..5\} \end{aligned} \quad (6.5)$$

The similarity score, image cropping and downsampling, and optimization strategy are the same as multi-view registration. The landmarks of all CMs are reprojected and incorporated to the regularizer, which can be described as

$$\begin{aligned} \mathcal{R}(T_{ref}^{Carm}, \tau_i) &= \frac{1}{2\sigma_l^2} \sum_{m=1}^M \left| \left| \mathcal{P}(p_{3D}^{(m,l)}; \theta_C, \tau_i) - p_{2D}^{(m,l)} \right| \right|^2, \\ & i \in \{1..5\}, l \in \{A, B\} \end{aligned} \quad (6.6)$$

The registration produces the pose of the reference frame ( $T_{ref}^{Carm}$ )<sub>regi</sub>. Each individual pose of the CM in the C-arm frame can be calculated using  $(T_{CM_i}^{Carm})_{regi} = (T_{ref}^{Carm})_{regi} \cdot T_{CM_m}^{ref}, m \in \{1..M\}$ . We present a step-by-step CM registration workflow in Fig. 6.9.

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.9:** Workflow of continuum manipulator registration.

### 6.4.3 Experiments and Results

We verified our fluoroscopic image-based CM navigation pipeline with a series of simulation studies. Lower torso CT scan images of a female cadaveric specimen were acquired for fluoroscopic simulation and anatomy registration. The CT voxel spacing is  $1.0 \times 1.0 \times 0.5$  mm with dimensions  $512 \times 512 \times 1056$ . Pelvis and femur volumes were segmented and pelvis anatomical landmarks were annotated using the method described in [3]. We manually annotated a drilling/injection entry point on the greater trochanter surface based on the biomechanical analysis. The simulation environment was set up to approximate a Siemens CIOS Fusion C-Arm, which has image dimensions of  $1536 \times 1536$ , isotropic pixel spacing of 0.194 mm/pixel, a source-to-detector distance of 1020 mm, and a principal point at the center of the image.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

### 6.4.3.1 Image-based Registration Simulation Study

We tested a series of CM 2D/3D registration workflows, including centerline-based 2D/3D registration, single-view 2D/3D registration, multi-view 2D/3D registration, and multiple CM 2D/3D registration. We performed 1,000 simulation studies with randomized geometries for each registration workflow and report the registration accuracy based on our simulated ground truth poses of the objects and the registration results.

The single-view image is approximately anterior/posterior (AP). Multi-view images included a perturbed AP view and two views at random rotations about the C-arm orbit with a mean and STD of  $+20 \pm 3^\circ$  and  $-15 \pm 3^\circ$ . Random movements of the pelvis were sampled uniformly to simulate patient pose variations, including translations from 0 to 10 mm and rotations from -10 to 10 degrees with respect to a randomly assigned unit vector in the pelvis volume center. Rotations of the femur were sampled from random rotations with respect to the center of the femoral head (FH). The axis of rotation was sampled uniformly between -15 and 15 degrees. The CM was simulated at random positions above the specimen femur head. Perturbed movements of the CM with respect to the C-arm included translations from -30 to 30 mm and rotations from -30 to 30 in degrees with respect to a randomly assigned unit vector in the CM base reference frame. To test the joint registration of multiple CM poses using a static C-arm view, we randomly sampled four CM poses with constant curvature in

## CHAPTER 6. SURGICAL TOOL NAVIGATION

the AP view including translations from -10 to 10 mm and rotations between -30 and 30 degrees with respect to randomly assigned unit vector in the CM reference frame for each registration.

Registration accuracy was assessed based on the the following metrics: 1) CM tip position in the C-arm frame ( $p_{Carm}^{CM_t}$ ); 2) entry point position in the C-arm frame ( $p_{Carm}^{ent}$ ); relative distance between the CM tip and the entry point in the C-arm frame ( $p_{Carm}^{ent} - p_{Carm}^{CM_t}$ ); 3) CM base rigid registration transformation in the C-arm frame ( $T_{Carm}^{CM_{base}}$ ). We computed the  $l_2$  distance error for items 1-3 above compared to the ground truth point position in the C-arm frame. CM base registration error was reported in the CM base reference frame using  $\delta T_{CM_{base}} = (T_{Carm}^{CM_{base}})_{regi} \cdot (T_{Carm}^{CM_{base}})_{gt}^{-1}$ . We also computed the CM notch (joint) rotation error compared to ground truth notch rotation angles.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

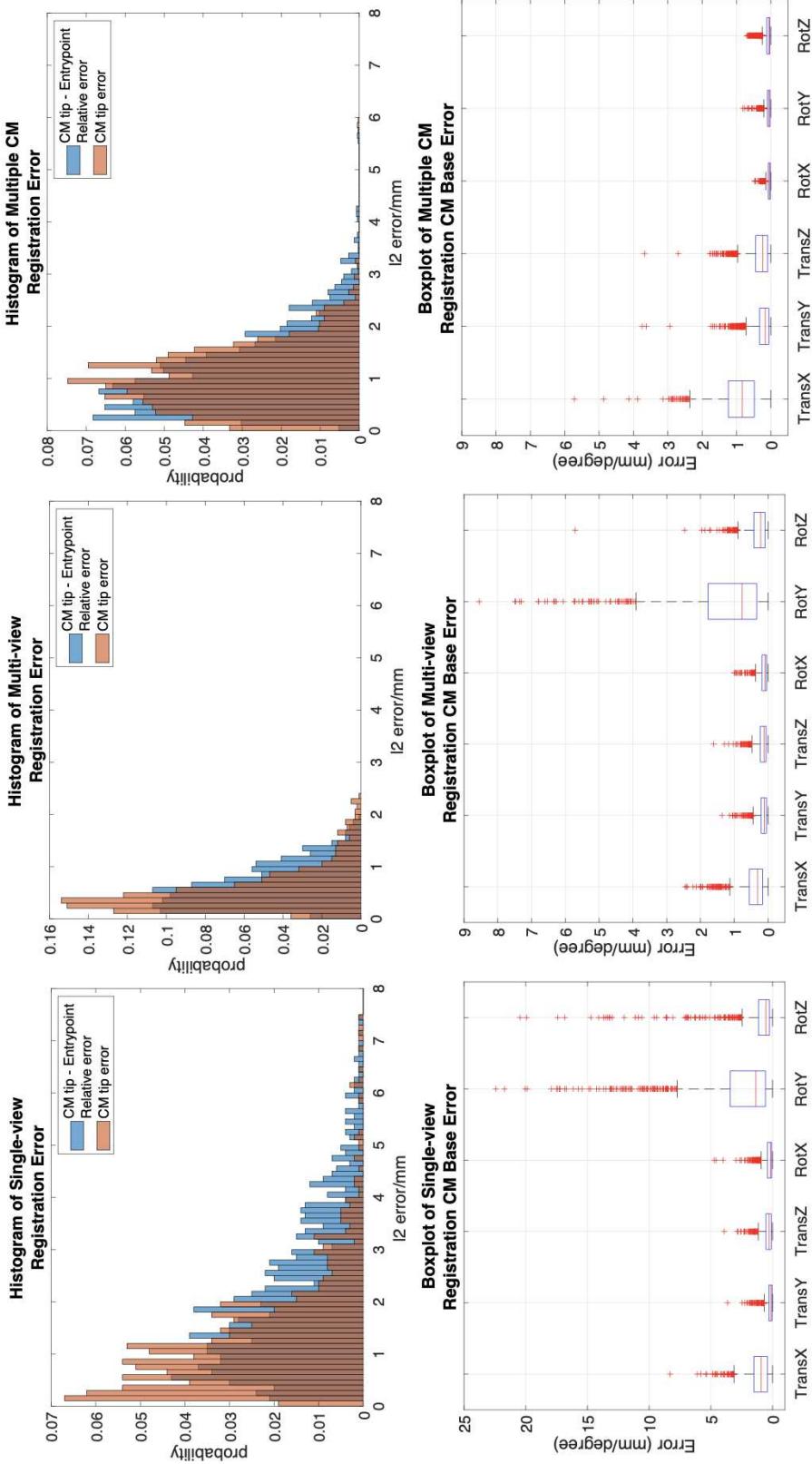
**Table 6.2:** Simulation Results of Registration Errors

		CM tip (mm)	Bone entry point (mm)	Relative (mm)	CM base translation (mm)	CM base rotation (degrees)	CM notch rotation (degrees)
Single-view	mean	1.32 ± 1.21	1.65 ± 1.13	2.07 ± 1.52	1.24 ± 1.05	3.32 ± 4.17	0.23 ± 0.26
	median	<b>1.04</b>	<b>1.33</b>	<b>1.73</b>	<b>1.03</b>	<b>1.68</b>	<b>0.16</b>
	P value	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001
Multi-view	mean	0.54 ± 0.42	0.59 ± 0.37	0.63 ± 0.39	0.52 ± 0.43	1.38 ± 1.32	0.22 ± 0.23
	median	<b>0.42</b>	<b>0.55</b>	<b>0.55</b>	<b>0.40</b>	<b>0.88</b>	<b>0.15</b>
	P value	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001
Multiple CM	mean	0.99 ± 0.59	0.63 ± 0.43	1.10 ± 0.72	1.00 ± 0.60	0.14 ± 0.10	0.18 ± 0.21
	median	<b>0.94</b>	<b>0.51</b>	<b>0.95</b>	<b>0.95</b>	<b>0.12</b>	<b>0.12</b>
	P value	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001

The p value is reported from T-Test score compared to single-view CM registration results.

The result is significant at  $p < 0.05$ .

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.10:** CM registration simulation results. Top: Histogram plot of CM tip error (in orange) and CM tip to femur entry point relative error (in blue) in the AP view C-arm frame. Bottom: Box plot of CM base registration error in translation and rotation of all axes.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

### 6.4.3.1.1 CENTERLINE-BASED INITIALIZATION

We used the CM detection results to test our centerline-based 2D/3D registration algorithm on the AP view simulation images. We assumed the initial curvature readings were added with a uniformly distributed random noise of  $[-0.5^\circ, 0.5^\circ]$  to the ground truth control point values. We uniformly sampled 1,000 points in the search space of  $d_A$  and  $\theta$ , and chose the lowest point as a global minimum for CM pose estimation. We compared the output with the ground truth values.  $d_A$  had a mean error of  $29.27 \pm 21.02$  mm, and  $\theta$  with  $30.78^\circ \pm 22.13^\circ$ .

### 6.4.3.1.2 SINGLE-VIEW REGISTRATION

We first performed a single-view pelvis registration and then the femur registration was initialized using the pelvis registration outcome as described in Chapter 4. We used the centerline-based registration result in the above section to initialize the intensity-based 2D/3D registration of the CM. The CM curvature noise model is the same as the centerline-based registration settings. Thus the registration started at a deflected shape estimation and optimized the  $n+6$  ( $n = 5$ ) DoF until convergence. We achieved a mean error of  $1.17 \pm 0.98$  mm for the CM tip position and  $1.88 \pm 1.39$  for the relative distance between the CM tip and femur bone entry point.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

### 6.4.3.1.3 MULTI-VIEW REGISTRATION

Multiple C-arm pose geometry was estimated using the pelvis registration result of each individual view. The femur registration was initialized by pelvis registration and jointly optimized using multi-view images. The CM registration was initialized using the AP view centerline-based registration result and also jointly optimized with multi-view images. The mean CM tip position error was  $0.64 \pm 0.45$  mm and the mean relative error was  $0.62 \pm 0.39$  mm.

### 6.4.3.1.4 MULTIPLE CM REGISTRATION

The femur registration was performed using the same method as in multi-view registration. However, the CM registration was performed using multiple CM pose configurations with the static AP view. We assumed that the relative CM poses were known and the CM curvature was constant among multiple poses. We set the reference frame of multiple CM poses using the method described in Section II-E. The registration optimized the reference frame pose and the CM control points resulting in  $n + 6$  ( $n = 5$ ) DoF until convergence. We achieved a mean CM tip position error of  $0.99 \pm 0.55$  mm and a mean relative error of  $1.29 \pm 0.78$  mm.

Registration errors are presented in Table 6.2. A histogram of the error metrics is shown in Fig. ???. These results are further discussed in the discussion

## CHAPTER 6. SURGICAL TOOL NAVIGATION

section.

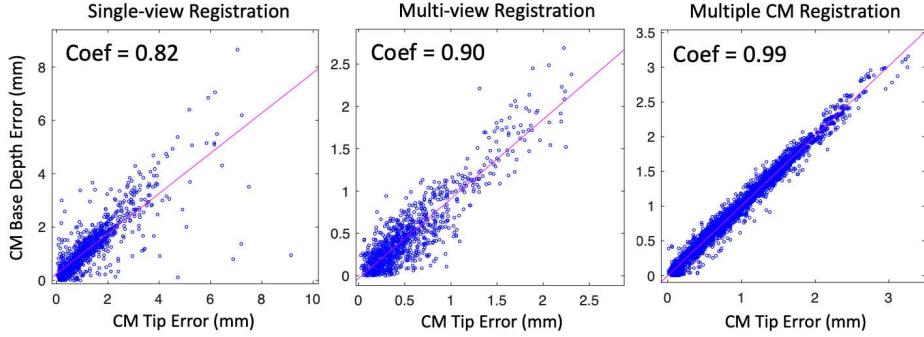
### 6.4.4 Discussion

Our proposed CM fluoroscopic navigation system performs multi-view registrations for both the CM and the bone. Our simulation study shows sub-millimeter accuracy in determining the relative translation error between the CM tip position and the target femur entry point. The mean CM base rotation error and the mean CM notch rotation error are less than 0.2 degrees.

Estimating CM pose and curvature using 2D/3D registration is a challenging problem. Our centerline feature-based registration yields a large mean error of  $30.78^\circ \pm 22.13^\circ$  in  $\theta$  (Fig. 6.7(c)), because the centerline feature has a strong symmetric property which leads to ambiguity when doing rotations. Due to its dexterity, the CM projection image itself has a distinct appearance, which is suitable for *intensity-based* 2D/3D registration. However, the problem of ambiguity along the projection line direction still exists in single-view intensity-based 2D/3D registration, which has large TransX (depth translation) and RotY (axial rotation) errors in simulation studies (Fig. ??).

Multi-view 2D/3D registration, however, combines the information of CM appearance under multiple projection geometries, which fundamentally improves single-view ambiguity. We compared two different methods of introducing multiple views: *multi-view registration* and *multiple CM registration*.

## CHAPTER 6. SURGICAL TOOL NAVIGATION



**Figure 6.11:** Scatter plots of the correlation matrix between CM tip position error and CM base TransX error of three registration methods. Correlation coefficients are marked on the upper left of each plot.

In simulation, although the mean translation error of multiple CM registration (1.00 mm) is higher than multi-view registration (0.52 mm), the mean axial rotation error is significantly smaller ( $0.12^\circ$  compared to  $0.88^\circ$ ) and the mean CM notch (joint) rotation error is also smaller ( $0.18^\circ$  compared to  $0.22^\circ$ ). Thus, multiple CM registration is preferable over multi-view registration for the CM because rotation accuracy is more important to navigating the positioning robot. Another advantage of the multiple CM registration is that motion of the CM can be automated as opposed to the C-arm motion. Multi-view registration is useful when the CM is inserted inside the bone and cannot move freely. Multi-view registration can provide an accurate CM tip position estimation with respect to the bone, which is essential for orthopedic applications.

To further demonstrate the relationship between the CM tip error and the registration ambiguity, we create correlation plots of the CM tip position error, ( $\| (p_{Carm}^{CM_t})_{gt} - (p_{Carm}^{CM_t})_{est} \|_2$ ), and the CM base origin depth error in the C-arm

## CHAPTER 6. SURGICAL TOOL NAVIGATION

frame, ( $(||(y_{Carm}^{CM_{base}})_{gt} - (y_{Carm}^{CM_{base}})_{est}||_2)$ ). All three methods show strong correlation between the CM tip error and the CM base depth error (correlation coefficient  $> 0.8$ ) (Fig. 6.11). Single-view registration has the lowest correlation of 0.82, which suggests part of the CM tip error comes from the rotation ambiguity. Multiple CM registration has the highest correlation coefficient of 0.99. This is because the joint registration of multiple CM poses balances the rotational ambiguity of registering a single CM, but the depth ambiguity still arises from using a single view.

The CM initialization process, including network detection, centerline extraction and curve-based initialization takes about 10 seconds. The multiple CM registration takes around 1 minute running on an NVIDIA GeForce RTX 2080 graphics processor. This registration is time consuming because we aim to estimate both the CM pose and curvature. The computation time can be significantly decreased if we only optimize for the shape deformation within a local search space.

The custom-designed flexible continuum manipulator has the potential to advance various orthopedic applications, such as core decompression of the hip, treatment of osteolysis, spinal surgeries, etc. We performed system-level calibration and cadaveric femur experiments to validate the CM navigation approach. The experiments and results are shown in Chapter. 7.

## CHAPTER 6. SURGICAL TOOL NAVIGATION

### 6.4.5 Conclusion

We present a fluoroscopic navigation system for a surgical robotic system including a continuum manipulator, which automatically detects the CM in a fluoroscopic image and uses purely image-based 2D/3D registration to estimate the pose and shape of the CM. We evaluated our methods through intensive simulation studies. The results showed the potential to apply the proposed CM navigation in robot-assisted orthopedic applications.

## 6.5 Acknowledgement

Sincere thanks to Dr. Amirhossein Farvardin and Mrs. Mahsan Bakhtiarinejad for their efforts in designing the injection device. I express my thanks to Dr. Shahriar Sefati for his work in developing and testing the robotic systems including the continuum manipulator. I want to thank Mr. Henry Phalen and Mr. Justin Ma for their help in the method design and experiments of the continuum manipulator registration. I also thank Drs. Mehran Armand, Mathias Unberath, Russell Taylor, Amit Jain, and Robert Grupp for their remarkable advice and suggestions.

# **Chapter 7**

## **System-Level Validation and Clinical Applications**

### **7.1 Introduction**

In previous chapters, we introduced critical component techniques for fluoroscopic navigation for robot-assisted orthopedic surgery. We have presented 2D/3D registration pipelines for bone anatomy and robotic surgical tools. The results of simulation studies suggest that the methods have the potential to be used for clinical applications. However, each clinical application has its unique challenge and requirement. Therefore, the navigation systems for robot-assisted orthopedic surgeries are custom-designed for each application. In this chapter, we present our efforts on system-level validation and cadaveric experiments

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

for clinical applications, including femoroplasty, core decompression of the hip, and transforaminal lumbar epidural injections. In the following sections, we introduce the overall system design, experiments, and results of each clinical application.

## 7.2 Contributions

The contributions of this chapter are as follows:

- We present hand-eye calibration methods of the robotic injection device and continuum manipulator, and robot system calibration results.
- We performed navigation system integration for the investigated orthopedic surgical applications, including femoroplasty, core decompression of the hip, and transforaminal lumbar epidural injections.
- We evaluated each application’s navigation system with cadaveric experiments using the robotic platform.

This chapter presents our efforts in performing system integration of the registration methods present in previous chapters and conducting *ex vivo* cadaveric experiments for each clinical application. We isolated and analyzed the source of errors of individual system components and the integrated performance under the clinical evaluation metrics. The results of the three applications sug-

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

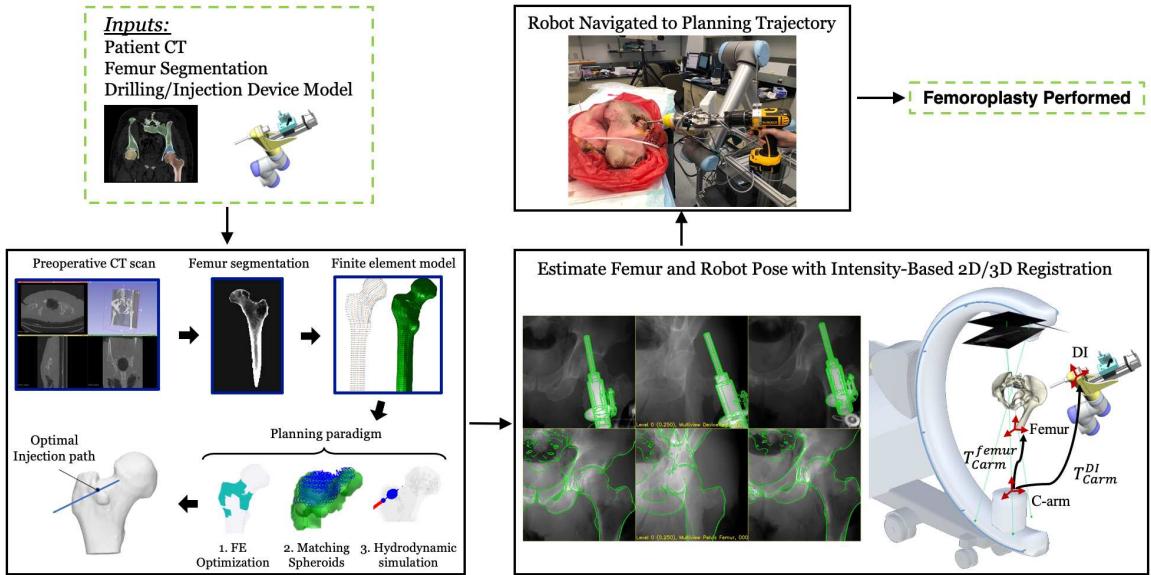
gest the feasibility of using the proposed fluoroscopic navigation methods for robot-assisted orthopedic applications.

The navigation system for femoroplasty is published in the July 2020 issue of IEEE Transactions on Medical Robotics and Bionics [3]. The femoroplasty cadaveric experiments and results are submitted to the journal of IEEE Transactions on Medical Robotics and Bionics under review. The experiments and results of core decompression are published in the July 2021 issue of IEEE Transactions on Biomedical Engineering [32]. The experiments and results of transforaminal lumbar epidural injections are published in the journal of IEEE Transactions on Medical Robotics and Bionics [30].

### 7.3 Femoroplasty

In Chapter 1, we present the clinical background of the proposed femoroplasty procedure. Patient-specific femoroplasty requires precise intra-operative navigation of the bone injection device according to a planned trajectory. Previous studies have presented a navigation system utilizing an image-based 2D/3D registration framework with intraoperative X-ray images and fiducial-based C-arm tracking [40]. This navigation system utilizes an optically-tracked fluoroscope fiducial and custom-designed fluoroscopy tracking fiducial (FTRAC) [139] in order to register the anatomy with respect to a hand-held injection device.

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

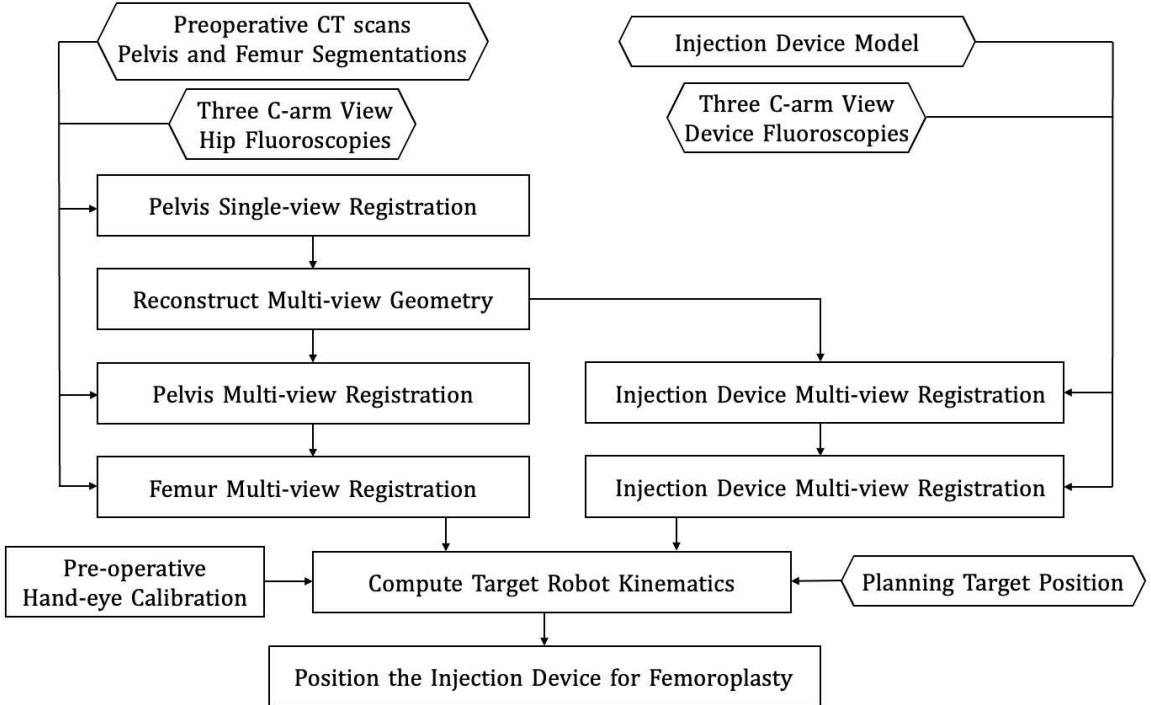


**Figure 7.1:** Fluoroscopic navigation pipeline for robot-assisted femoroplasty.

However, attaching external pins to the bone may not be an ideal option for patients with severe osteoporosis. Furthermore, an optical tracking system requires a clear line of sight. In addition, simultaneous viewing of the X-ray fiducial and the anatomy of interest may become challenging given the limited field of view of fluoroscopy. Hand-held manipulation of the injection device requires interaction with the surgeon, which makes the procedure complicated.

In this dissertation, we present the proposed fiducial-free fluoroscopic navigation pipeline for robot-assisted femoroplasty. We used a custom-designed drilling and injection device guide to the end effector of a 6 DOF positioning robot, UR-10 (Universal Robot, Odense, Denmark) for a more precise injection device positioning. The robotic injection device is introduced in Chapter 6.3. Navigation of the robotic injection device to the planned trajectory

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS



**Figure 7.2:** Workflow for femoroplasty.

is achieved using fiducial-free fluoroscopic image-based 2D/3D registration.

Fig. 7.1 presents an overview of the proposed navigation pipeline. The inputs include patient CT scans, automatically segmented pelvis and femur volumes, and the injection device model. The CT data acquisition, preprocessing and automatic segmentation are introduced in Chapter 4.4.1. A patient-specific injection path is planned following the preoperative planning paradigm fully described in [1]. The intra-operative poses of the target femur and robot injection device are estimated using the image intensity-based 2D/3D registration methods described in Chapter 4.4.2 and Chapter 6.3. After registration, the robotic injection device guide is navigated and positioned at the planned trajectory

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

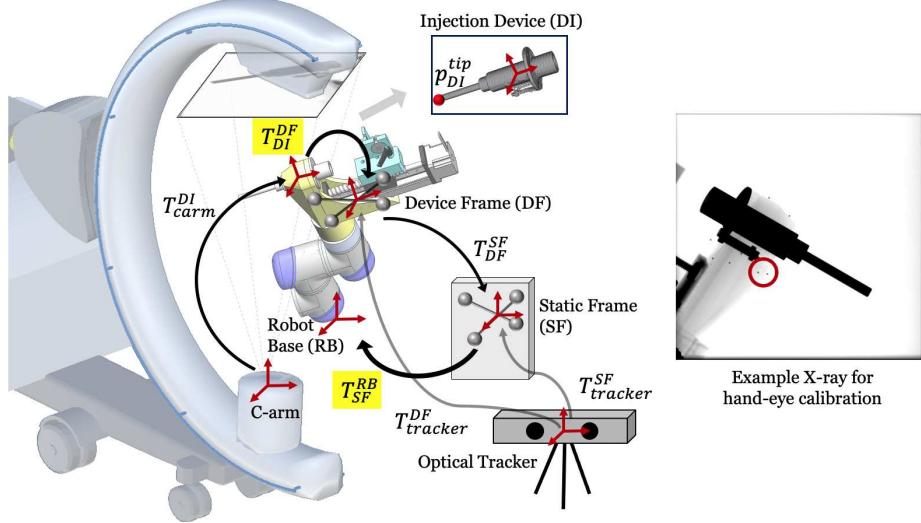
entry position. Femoroplasty is then performed through drilling followed by injection of the Polymethylmethacrylate (PMMA) is performed. We present a workflow chart for the navigation pipeline in Fig. 7.2. In the following subsections, we present the methods of calibrating the injection device to the robotic system and femoroplasty cadaveric experiments and results.

### 7.3.1 Injection Device Calibration

Our robotic system uses closed-loop navigation by optimizing the real-time pose of an optical marker attached to the injection device. Two calibrations are required: hand-eye calibration of the optical frame, and hand-eye calibration of the injection device (Fig. 7.3).

#### 7.3.1.1 Hand-eye Calibration of the Device Frame

A hand-eye calibration was performed to determine the location of the optical tracker body on the injector unit ( $DF$ ) relative to the robot's base coordinate frame ( $RB$ ). This allows for real-time estimation of the manipulator Jacobian  $J_m$  associated with the movement of the injector attached to the base robot. The calibration was performed using the well-established method of moving the robot to 60 configurations within the region of its workspace in which the injections would occur, recording the robot's forward kinematics and the optical tracker body location ( $T_{DF}^{SF}$ ), and using these measurements to solve an



**Figure 7.3:** Hand-eye calibration scheme. Coordinate frames are marked as red cross arrows. Key transformations are shown in black arrows. The hand-eye calibration transformations are highlighted. The 3D injection device model for 2D/3D registration is illustrated on top. An example X-ray image used for hand-eye calibration is shown on the right. Example BBs are marked in a red circle.

$AX = XB$  problem to find  $T_{SF}^{RB}$  highlighted in Fig. 7.3.

### 7.3.1.2 Hand-eye Calibration of the Injection Device

Another hand-eye calibration was conducted to compute the transformation of the injection device model frame (*DI*) to the optical tracker unit (*DF*). This is necessary to integrate the registration pose estimation into the closed-loop control. At different robot configurations, X-ray images of the injection device were acquired. 2D BB locations are easily detected on the images and were manually annotated (Fig. 7.3). The rigid pose of the injection device ( $T_{Carm}^{DI}$ ) was estimated by solving a PnP problem. This also results in an  $AX = XB$  problem

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

to find  $T_{DI}^{DF}$  in Fig. 7.3(a). These two hand-eye calibration processes only need to occur if the injector is removed and reattached to the robot.

We moved the robotic injection device to 30 variant configurations for injection device hand-eye calibration, while the C-arm was fixed static. At each configuration, we took an X-ray image and solved the injection device pose  $T_{Carm}^{DI}$ . After solving the  $AX = XB$  problem to find  $T_{DI}^{DF}$ , we reported the calibration accuracy by calculating the injection device tip position difference between the PnP estimation ( $(T_{Carm}^{DI})_i^{pnp}$ ) and estimation using the chain of calibration transformations:

$$(T_{Carm}^{DI})_i^{cali} = (T_{DI}^{DF})^{-1} \cdot (T_{DF}^{SF})_i^{-1} \cdot T_{DF}^{SF} \cdot T_{DI}^{DF} \cdot (T_{Carm}^{DI})_0, \quad (7.1)$$

where  $i$  is the index of the calibration frame and  $(T_{Carm}^{DI})_0$  is our reference pose corresponding to the first calibration frame. The hand-eye calibration error was calculated as the mean  $l_2$  difference of the estimated injection guide tip in the injector model ( $p_{DI}^{tip}$ ) between these two pose estimations:  $\|((T_{Carm}^{DI})_i^{cali} - (T_{Carm}^{DI})_i^{pnp}) \cdot p_{DI}^{tip}\|$ . The mean error was  $2.49 \pm 1.55$  mm.

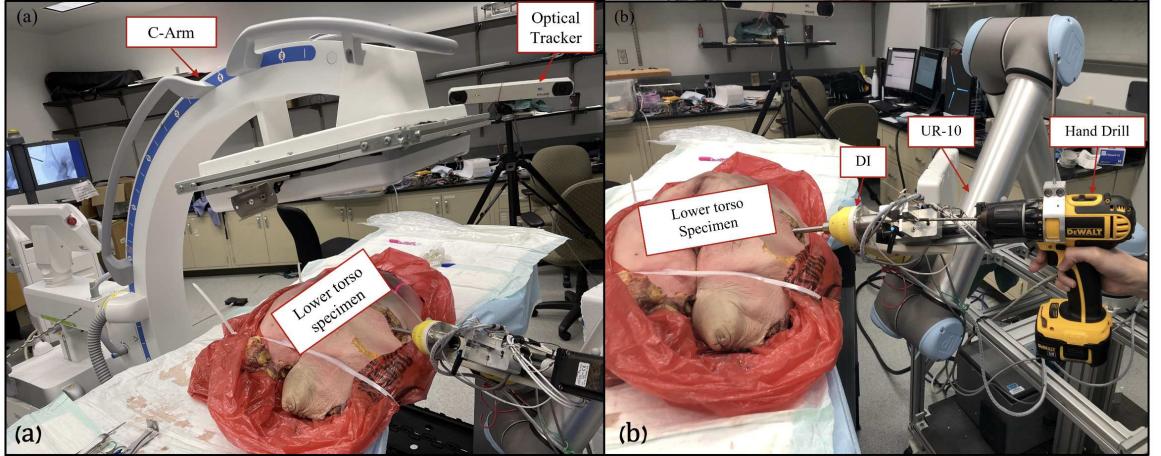
### 7.3.2 Cadaveric Experiments and Results

We performed a cadaveric drilling experiment on a male specimen including the lower torso, pelvis, and femurs to further evaluate the performance of the robot-assisted femoroplasty with the use of fluoroscopic image-based 2D/3D registration. Fig. 7.4 shows the experimental setup. Based on the biomechanical planning, the drill path was determined on the segmented 3D model of the right femur. We commanded the manipulator to position the injection device guide 1-2 mm from the insertion point. After drilling, we placed a 4 mm drill bit inside the trajectory and tool a lower torso post-op CT scan. An example of multi-view registration overlay images is illustrated in Fig. 7.5 (c). The precise match of the DRR edges and the X-rays in multi-view geometries indicates an accurate pose estimation in 3D.

We took post-operative CT scans to evaluate the image-guided drilling performance. The insertion and target points along the drill trajectory were marked on the post-op CT. We performed a 3D/3D registration between the post-operative and pre-operative CT scans. A 3D point cloud was manually annotated on the femoral head in the post-op CT scan. Next, we performed an ICP registration from the 3D point cloud to the pre-op segmented femur surface to register the post-op CT to pre-op CT coordinates. The marked insertion and target points were subsequently transformed to the pre-op CT coordinates.

We illustrate the comparison between measured and planned trajectories in

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS



**Figure 7.4:** (a) Femoroplasty cadaver experiment setup. The C-arm, optical tracker and lower torso specimen are marked in the image. (b) Picture of guided drilling using robotic injection device after fluoroscopic navigation.

the cadaver experiment in Fig. 7.5. We report  $l_2$  distance error and an orientation error of the insertion and target points compared to the planning using

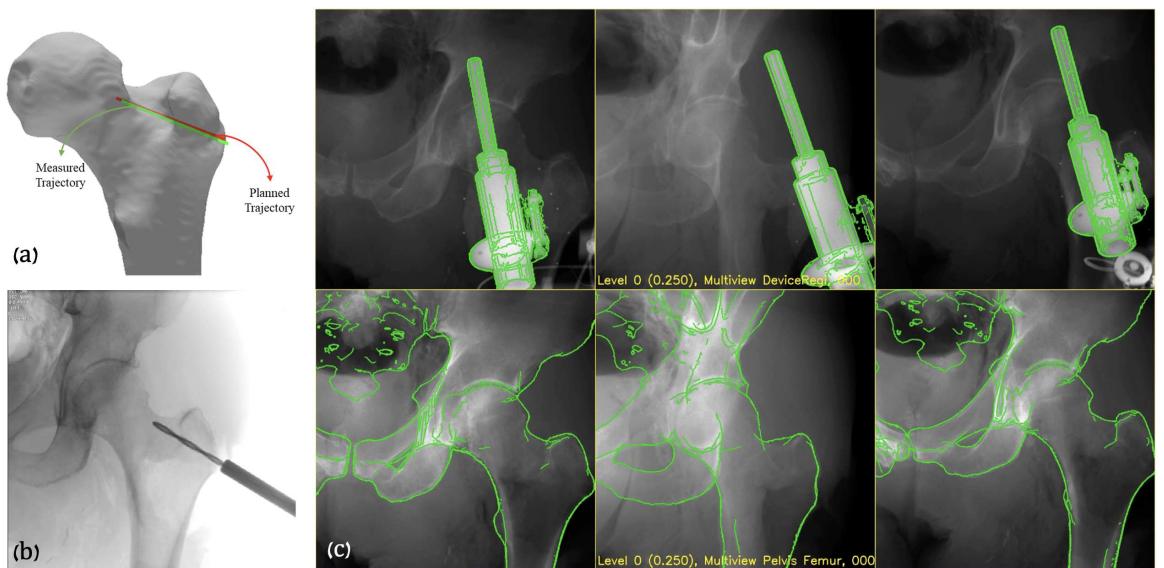
$$\delta x = \|x_{plan}^i - x_{postop}^i\|_2, i \in \{insertion, target\} \quad (7.2)$$

$$p = x^{target} - x^{insertion} \quad (7.3)$$

$$\delta\theta = \arccos\left(\frac{\mathbf{p}_{plan} \cdot \mathbf{p}_{postop}}{\|\mathbf{p}_{plan}\| \cdot \|\mathbf{p}_{postop}\|}\right) \quad (7.4)$$

The distance errors between planned and measured insertion and target points in the lower torso specimen are calculated as 3.28 mm and 2.64 mm, respectively and the drill line orientation error is 2.30°. The biomechanical

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS



**Figure 7.5:** (a) Measured and planned trajectory on the cadaver femur model. (b) X-ray image of the guided drilling using an injection device in the cadaveric experiment. (c) Illustrations of registration convergence in the cadaveric study. Multi-view cadaver X-ray images are placed as background and DRR-derived edges are overlaid in green. The top row shows injection device registration, and the bottom row shows the pelvis and femur registration.

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

simulation predicts a 33% increase of yield load for the right femur of the lower torso specimen with the FE-optimized cement injection pattern, simulating a side-way fall on the greater trochanter. Based on the measured target point of the drill trajectory from the post-op CT scan, we shift the cement injection blobs inferior and perform the SPH simulation. FE simulation predicts a 26% increase in the yield load with the injection pattern of the post-op trajectory.

### 7.3.3 Discussion

In this cadaveric study, we successfully integrated the image-based 2D/3D registration pipeline into the surgical robotic system and evaluated it in a cadaver experiment. The entry point error (3.28 mm) is slightly higher than the cadaveric registration tip error reported in [3] (2.64 mm). The potential reasons include systematic errors introduced during hand-eye calibration; optical tracker jig frame detection error; the difference between the simulated DRR and real X-ray images caused by imperfect bone CT segmentation, and X-ray spectrum and exposure simulations. On the other hand, the target point error is slightly lower (2.64 mm). The orientation error ( $2.30^\circ$ ) between planned and measured drill lines is likely due to the femur registration error. Because the femur surface is smooth and featureless, it is challenging to accurately estimate the axial rotation of the femur.

To evaluate the impact of the drilling accuracy on the fracture-related biome-

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

chanical outcome of femoroplasty, we perform a hydrodynamics simulation on the shifted cement profile based on the post-op trajectory measurements. As a result, the estimation of the yield load improvement has been reduced by only 7% compared to the originally planned trajectory, which demonstrates the sufficient accuracy of the robot-assisted drilling and injection. In addition, improving the registration accuracy will consequently enhance the post-operative biomechanical outcome of femoroplasty.

The current study focuses on assessing the feasibility of the bone drilling component of our robot-assisted bone augmentation surgical workstation through sawbone and cadaver experiments. The results of the robot-assisted drilling demonstrate a promising improvement over the navigated user drilling with visual feedback. In addition, in the study of Farvadin et al., it was demonstrated that utilizing curved patterns for cementation reduces the volume of cement injection to an average of 7.2 ml while substantially increasing the yield load of the femur by 69% [140].

Our robot-assisted injection system using image-based navigation positions the injection guide to the planned femur entry point without creating large incisions, which is less invasive than a surgeon's free hand operation. Compared to a previous study using hand-held injection unit with optical fiducial-based navigation [40], the robot system is more stable and the bone pin markers are not required. Our registration pipeline is developed based on X-ray imaging

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

because the C-arm X-ray machine is a common device in all the orthopedic operating rooms. Therefore, using intraoperative X-ray does not add additional imaging hardware for the procedure. Although our method requires radiation exposure on the patient, using X-rays in orthopedic applications is common. Taking six X-rays is usually not excessive as compared to other orthopedic applications. The full intra-operative registration pipeline including anatomy and injection device registrations takes about 5 minutes to run on an NVIDIA GeForce RTX 2080 graphics processor. Considering the data transfer, injection device X-ray landmark annotation, and registration result verification, the whole process takes around 15 - 20 minutes to finish. In the future, the pipeline can be further automated and optimized to speed up the registration processing time.

## 7.4 Core Decompression of the Hip

Core decompression of the hip is the procedure that treats osteonecrosis in the femoral head. The continuum manipulator (CM) introduced in Chapter 6 is custom-designed for drilling a curvature through the femoral neck to better reach the osteonecrosis lesion. We have presented methods to precisely estimate the shape and pose of the CM using fluoroscopic images and 2D/3D registration. In this chapter, we present a modified hand-eye calibration method

that calibrates the CM to the robotic system and the efforts of cadaveric experiments to validate the navigation system.

### 7.4.1 CM Hand-eye Calibration

In order to navigate the CM using the surgical robotic system, the transformation from the CM model frame to the rigid-link robot end-effector frame,  $T_{CM}^{ee}$ , needs to be calibrated, which is essentially a hand-eye calibration problem. Fig. 7.6 illustrates the hand-eye calibration transformations. The conventional hand-eye calibration method collects the calibration data at each individual frame, building matrices  $A$  from the robot kinematics and  $B$  from sensor readings. If A system of linear equations  $AX = XB$  can be defined where  $X$  is the hand-eye matrix. However, in this application,  $X$  will be inaccurate because  $B$  is ambiguous due to severe single-view 2D/3D registration ambiguity. Following the multiple CM registration methods as described in Section 6.4.2.2, we propose a modified hand-eye calibration method that numerically optimizes the hand-eye matrix during the multiple CM registration. Specifically, the hand-eye matrix  $T_{CM}^{ee}$  is now included as a target variable in the optimization problem defined in (6.5). The intuition is that the collection of multiple calibration CM poses can be regarded as a CM object group, then  $T_{ref}^{Carm}$  controls the central pose of the group, and  $T_{CM}^{ee}$  controls the relative poses between the CM objects within the group.  $T_{ref}^{Carm}$  and  $T_{CM}^{ee}$  are sufficient to de-

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

termine all the CM rigid poses in the static C-arm frame. The hand-eye calibration can be formulated as

$$\begin{aligned} \min_{T_{ref}^{Carm}, T_{CM}^{ee} \in SE(3), \tau_i} & \sum_{m=1}^M \mathcal{S}\left(I_k, \mathcal{P}\left(\sum_{j=1}^J V_j; T_{ref}^{Carm}, T_{CM}^{ee}, \tau_i\right)\right) \\ & + \mathcal{R}(T_{ref}^{Carm}, T_{CM}^{ee}, \tau_i), i \in \{1..5\} \end{aligned} \quad (7.5)$$

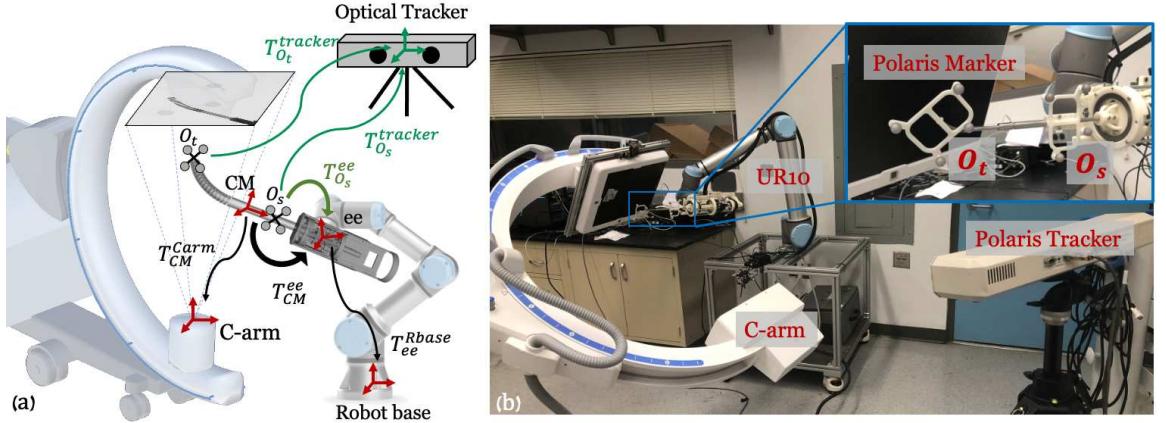
and the regularizer is defined as

$$\begin{aligned} \mathcal{R}(T_{ref}^{Carm}, T_{CM}^{ee}, \tau_i) = & \frac{1}{2\sigma_l^2} \sum_{m=1}^M \|\mathcal{P}(p_{3D}^{(m,l)}; \theta_C, \tau_i) - p_{2D}^{(m,l)}\|^2, \\ & i \in \{1..5\}, l \in \{A, B\} \end{aligned} \quad (7.6)$$

The registration directly produces the hand-eye matrix  $T_{CM}^{ee}$ , which integrates the CM model frame to the rigid-link robot end-effector frame. There is no need for a fiducial marker in the system as the CM itself is functioning as a fiducial via the model-to-image registration that connects the C-arm camera frame to the robot kinematics.

In order to *validate* the accuracy of the hand-eye calibration, we put two rigid bodies with reflective markers on the CM tip segment and shaft. The optical tracker simultaneously tracks these two rigid bodies for validation. At each static configuration, we also take the optical tracker readings. Thus, we

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS



**Figure 7.6:** (a) Illustration of the hand-eye calibration transformations. The image-based hand-eye calibration loop is shown in black. The image-based hand-eye transformation is  $T_{CM}^{ee}$ . The optical tracker-based hand-eye calibration loop is shown in green, and  $T_{O_s}^{ee}$  is the optical tracker-based hand-eye matrix. (b) Hand-eye calibration data collection setup with UR-10, C-arm Polaris tracker, and markers.

can perform an additional standard hand-eye calibration using the accurate optical tracking result. The data collection setup is shown in Fig. 7.6 (b). The shaft marker pose in the optical tracker frame,  $T_{O_s}^{tracker}$ , is used to define the  $A$  matrix, and the robot kinematics  $T_{ee}^{Rbase}$  is used to define the  $B$  matrix.

We collected X-ray images using the surgical robotic system for hand-eye calibration. Two rigid bodies with 4 optical markers each were used for validation. Marker ( $O_s$ ) was fixed rigidly to the shaft of the CM. Marker ( $O_t$ ) was rigidly attached to the last segment of CM. The two rigid bodies were simultaneously tracked by an NDI Polaris (Northern Digital Inc., Waterloo, Ontario, Canada) system. We used a 6-DOF UR-10 (Universal Robots, Odense, Denmark) as the rigid-link robot. We manually commanded the UR-10 to 60

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

different configurations. The C-arm was static during the calibration data collection procedure. At each configuration, we took an X-ray image using a 30 cm flat panel detector Siemens CIOS Fusion C-Arm, recorded the UR-10 forward kinematics ( $T_{ee}^{Rbase}$ ) and two Polaris marker positions ( $T_{O_t}^{tracker}, T_{O_s}^{tracker}$ ). The CM curvature was kept constant during the calibration data collection.

We performed the proposed multi-view hand-eye calibration method and found  $T_{CM}^{ee}$ . We solved another hand-eye calibration using the optical tracking data of marker ( $O_s$ ). The relative frame transformations  $(T_{O_t}^{tracker})_i \cdot (T_{O_t}^{tracker})_j$  and  $(T_{ee}^{Rbase})_i \cdot (T_{ee}^{Rbase})_j$  are the  $A$  and  $B$  matrices of the hand-eye equation  $AX = YB$ , and  $T_{O_s}^{ee}$  is the  $X$  hand-eye matrix. We then transformed the CM tip position into the robot base frame using both the optical tracking and image-based registration hand-eye results. This can be formulated as:

$$p_{Rbase}^{O_t} = T_{ee}^{Rbase} \cdot T_{O_s}^{ee} \cdot (T_{O_s}^{tracker})^{-1} \cdot T_{O_t}^{tracker} \cdot p_{O_t}^{O_t} \quad (7.7)$$

$$p_{Rbase}^{CM_t} = T_{ee}^{Rbase} \cdot T_{CM}^{ee} \cdot (T_{CM}^{Carm})^{-1} \cdot T_{CM_t}^{Carm} \cdot p_{CM_t}^{CM_t} \quad (7.8)$$

where  $p_{O_t}^{O_t}$  is the origin point of the  $O_t$  marker and  $p_{CM_t}^{CM_t}$  is the tip point of the CM model. We then computed the  $l_2$  distance error between  $p_{Rbase}^{O_t}$  and  $p_{Rbase}^{CM_t}$ . We achieved a mean error of  $2.15 \pm 0.50$  mm.

## 7.4.2 Cadaveric Experiments and Results

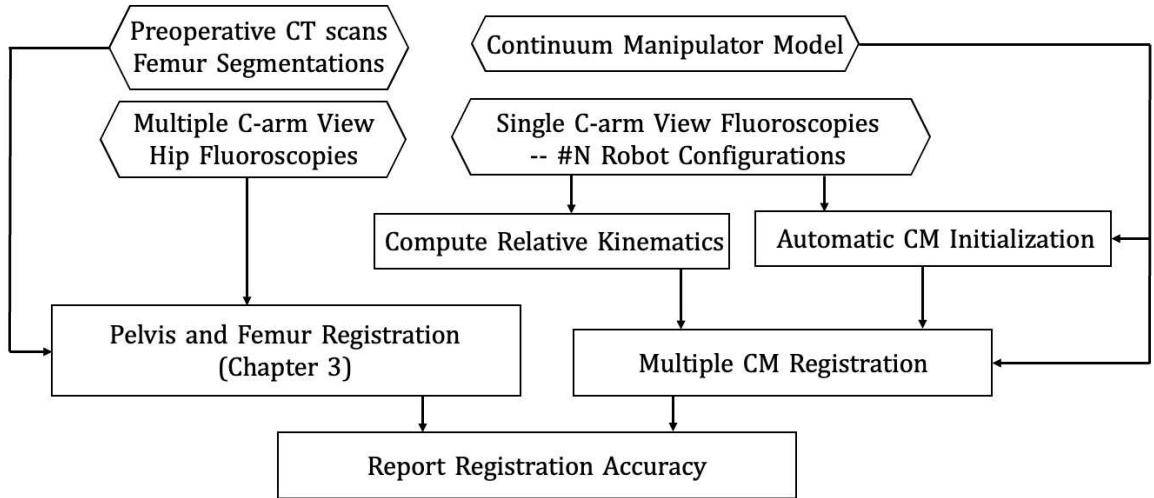
To validate the proposed CM navigation pipeline, We performed cadaveric experiments using two cadaveric specimens. A female Specimen1, including lower torso, pelvis, and femurs, was used for testing registration and robot positioning. A right femur of Specimen2 was used for testing registration when the CM was inserted into the femoral head. To obtain the ground truth poses for the femur anatomy, metallic BBs were implanted into the femoral head as shown in Fig. 7.9. For Specimen1, the BBs were implanted closer to the trochanter and the femoral head center region in order to accurately estimate the femoral head pose. For Specimen2, the BBs were evenly distributed around the femoral head. We took CT scans of both Specimen1 and Specimen2. The 3D locations of the BBs were manually labeled in the CT scans of the specimens. The 2D BB locations were manually annotated in the X-ray images. The registration workflows used in the cadaveric experiments for Specimen1 and Specimen2 are presented in Fig. 7.7 and Fig. 7.8, respectively.

**Table 7.1:** Cadaver Specimen1 Results of Error Metrics

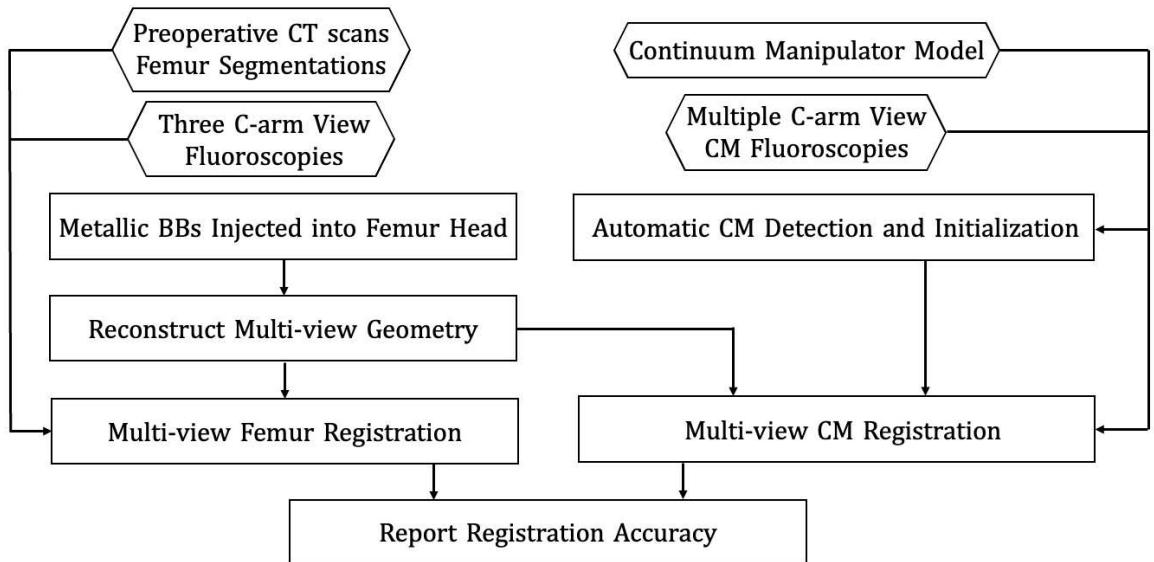
Trial ID	I	II	III	IV	V
CM Tip Position (mm)	2.73	2.44	2.77	4.20	2.09
Femur Entry Point (mm)	1.21	2.23	1.91	2.35	2.04
Relative (mm)	2.95	3.31	3.65	4.10	1.83

We performed the registration workflow five times using Specimen1 with varying C-arm geometries and specimen poses. Each registration workflow

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

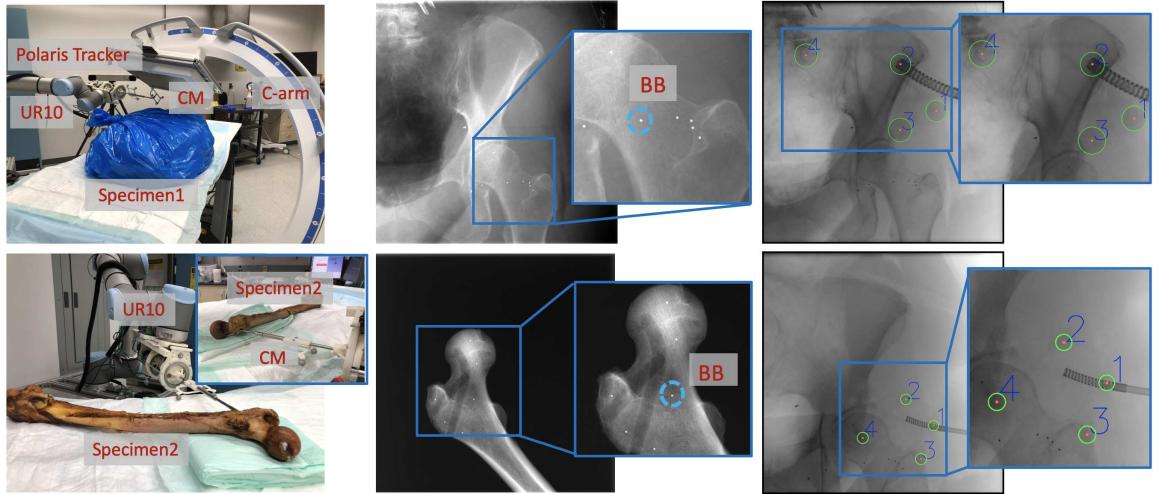


**Figure 7.7:** Workflow of continuum manipulator cadaveric experiment for Specimen1.

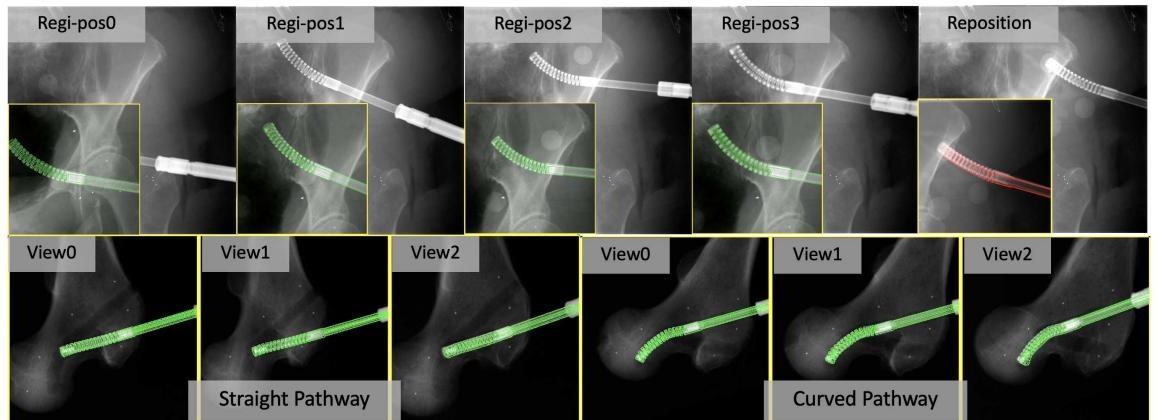


**Figure 7.8:** Workflow of continuum manipulator cadaveric experiment for Specimen2.

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS



**Figure 7.9:** Upper Left: Specimen1 experiment setup with C-arm, UR-10, Polaris tracker, and CM. Upper Right: Specimen2 experiment setup with UR10 and CM inserted into the femur head. Middle: Examples of BB locations. Bottom: Examples of Polaris fiducial detection results.



**Figure 7.10:** 2D overlay examples of fluoroscopic images (background) and DRR-derived edges in green when registration is converged. Top: Cadaveric Specimen1 CM registration results and position overlay after reposition (in red). The original full-size image is placed in the background and the cropped CM registration image is placed in the foreground lower-left corner. Bottom: Cadaveric Specimen2 CM registration results of each C-arm view for the straight and curved pathway insertions.

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

had three images for femur registration and four images for CM registration. After each registration, we commanded the UR-10 to another configuration and took another X-ray image to check the CM reposition accuracy. The optical fiducials were detected in the X-ray image using Hough transform circle detection. The optical marker ( $O_t$ ) was attached to the tip segment of the CM. Thus, the ground truth CM tip position,  $(p_{Carm}^{CM_t})_{gt}$ , in the C-arm frame was calculated by solving the PnP problem using corresponding fiducials in X-ray and marker configurations. The CM tip position error was reported using the  $l_2$  distance between estimated tip position from registration  $(p_{Carm}^{CM_t})_{est}$  and  $(p_{Carm}^{CM_t})_{gt}$ . Ground truth femur entry-point position,  $(p_{Carm}^{FEM_{ent}})_{gt}$ , was calculated from the PnP solution using the corresponding BBs.  $l_2$  distance between estimated  $((p_{Carm}^{FEM_{ent}})_{est})$  and ground truth  $((p_{Carm}^{FEM_{ent}})_{gt})$  femur entry point was reported as femur entry point error. We also calculated the relative error between the entry point and the CM tip using

$$||((p_{Carm}^{FEM_{ent}})_{gt} - (p_{Carm}^{CM_t})_{gt}) - ((p_{Carm}^{FEM_{ent}})_{est} - (p_{Carm}^{CM_t})_{est})||_2 \quad (7.9)$$

We achieved a mean CM tip position error of  $2.86 \pm 0.80$  mm and a mean relative error between CM tip and femur entry point of  $3.17 \pm 0.86$  mm. Table. 7.1 presents the error metrics of each individual trial.

For Specimen2, we drilled two pathways inside the femoral head using the

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

CM. One pathway was curved and the other one was straight. The CT scan was taken after the pathways were drilled. Three multiple-view X-ray images were taken at several positions along with the insertion of the CM into the end of the bone pathways (Fig. 7.10). We then performed the registration of the CM and the femur. We manually annotated both the straight and the curved pathway endpoints in the CT scan. The registration accuracy was reported using the  $l_2$  distance between CM tip positions and the CT pathway endpoints in the CT coordinate frame. We achieved a tip position error of 2.88 mm of the curved pathway registration and 2.65 mm of the straight pathway registration.

Fig. 7.10 illustrates overlay images of the registration.

### 7.4.3 Discussion

Our studies suggest the feasibility of applying purely fluoroscopic image-based registration for CM navigation. This is of interest because fluoroscopy is the most common imaging modality for orthopedic applications. Our navigation system automates the detection of distinct CM features in fluoroscopic images, which provides an initial CM pose estimation in 3D. Accurate pose estimation is achieved using intensity-based 2D/3D registration of the fluoroscopic image to the CM model. To navigate robotic interventions, the model-based registration result is integrated into the rigid-link robot kinematics configuration using a modified hand-eye calibration method. The cadaver study results sug-

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

gest a mean translation error of 3.17 mm and a maximum error of 4.10 mm between the two points, which is sufficient for osteonecrosis lesion removal (less than 1 cm), and is feasible to guide the CM for orthopedic applications.

The accuracy of the cadaveric experiments is less than the simulation study. There are several potential reasons: 1) the CM appearance in real fluoroscopic images is different from simulated DRR images due to spectrum and exposure; 2) the UR kinematics introduces joint configuration errors due to the robot going out of calibration; 3) the gravity parameter of the CM and the actuation unit not compensated in the UR kinematics; 4) the BB injection, Polaris fiducial detection, manual annotation, and segmentation are likely to introduce errors.

One drawback of fluoroscopic navigation is the radiation exposure to the patient. Our approach considers both rigid CM pose tracking and shape estimation, which requires six to seven X-rays to register the patient and the CM. This is usually not excessive as compared to other orthopedic applications. The multiple CM registration takes about 1 minute to finish. However, the pose of patient anatomy can be different after the registration due to tool/bone interactions, such as drilling/milling or injection. In this case, additional X-rays are required to correct the registration and update the robot navigation planning. If needed, additional registrations can keep track of the tool pose during a surgical operation, which can account for movement of the anatomy, or unmodeled interaction behavior. The process of re-registration will require only

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

2-3 additional X-ray shots. The registration method can be adapted to only estimate the CM deformation, which limits the optimization parameter search space and will be faster to converge. Future studies can involve such direction to decrease the registration computation time. Unlike other navigation systems which can perform real-time tracking (e.g. optical tracking), fluoroscopic images cannot be acquired at high frequency in real-time for safety reasons. The internal sensing units, like the FBG sensors, can provide real-time CM curvature sensing with high-frequency (100 Hz) [141], but the measurements may be inaccurate, and the tool to tissue relationship can not be directly visualized. Thus, to better control the CM, the image-based navigation needs to be combined with the use of internal sensing units.

The feasibility of fusing data from FBG sensors and overhead cameras to control the CM motion was previously investigated [142]. The future work includes the study of combining the FBG sensor data with our fluoroscopic image-based registration to perform intra-operative CM control.

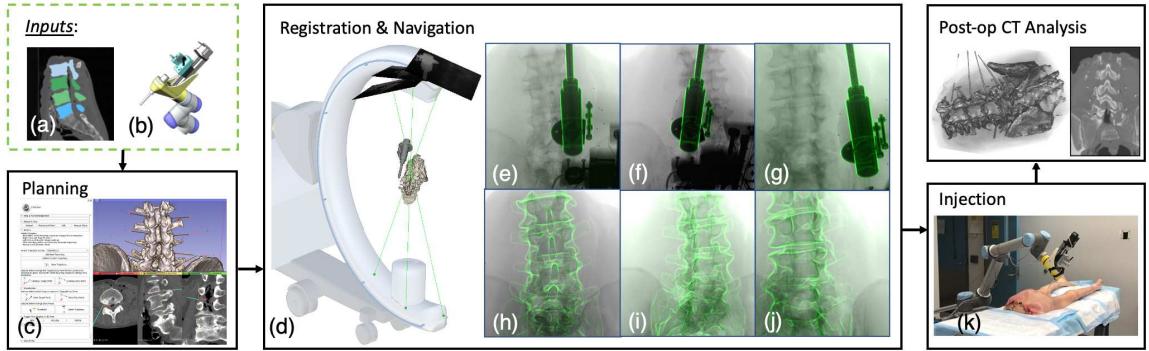
## 7.5 Transforaminal Lumbar Epidural Injections

Transforaminal epidural steroid injection in the lumbar spine (TLESI) treats lower back pain or sciatica by injecting a steroid with needle placement. Robotic

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

systems have been considered in the literature as a tool to perform these injections. Various imaging technologies have been used for guidance of these robotic systems including MRI [143–146], ultrasound [147,148], and cone-beam CT [149]. However, MRI and CT machines are expensive and are not commonly available in orthopedic operating rooms. Furthermore, these 3D imaging modalities - MRI in particular - can greatly prolong the surgical procedure. Ultrasound data are often noisy and it can be complicated to extract contextual information. Thus, ultrasound-guided needle injection requires a longer scanning time and is limited in reconstruction accuracy [147]. Often additional sensing modalities are needed along with ultrasound, such as force sensing [148]. Fluoroscopic guided needle placement has been studied [150–152]. These approaches either require custom-designed markers to calibrate the robot end effector to the patient’s anatomy or the surgeon’s supervision to verify the needle placement accuracy.

Our proposed fiducial-free fluoroscopic navigation does not put markers on the patient’s body, which is less invasive and simplifies the procedures. The robotic injection system performs planning, registration and navigation, automatic injection, and post-operative analysis. Fig. 7.11 presents the overall pipeline of the robotic injection system. In Chapters 4 and 6, we have presented pose estimation methods of the injection device and spine vertebrae. In this chapter, we present the planning module and system calibration, and our



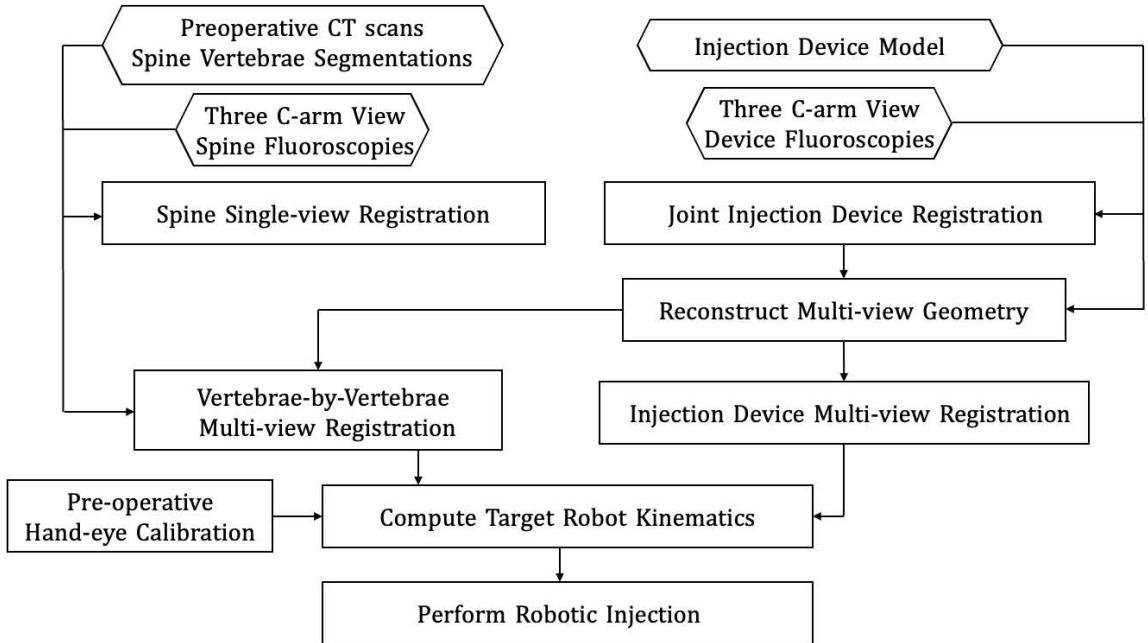
**Figure 7.11:** Overall pipeline of our robotic needle injection system. Inputs include (a) patient-specific CT scan and spine vertebrae segmentation, and (b) an injection device model. The planning module shows (c) the surgeon’s interface to annotate needle injection trajectories and an example display of the planned trajectories on the CT segmentation. Multi-view registration presents: (d) multi-view C-arm X-ray projection geometries. The source-to-detector center projection line is rendered in green and the detector planes are rendered as squares. The needle injector guide and the spine anatomy are rendered using the registration pose. (e)(f)(g) Registration overlay images of the needle injector guide. The outlines of the reprojected injection device are overlaid in green. (h)(i)(j) Registration overlay images of the cadaveric spine vertebrae. We present an actual cadaveric needle injection image in (k).

clinical experiments and results. A step-by-step navigation workflow is presented in Fig. 7.12.

### 7.5.1 Pre-operative Injection Planning

Needle targets and trajectories were planned in a custom-designed module in 3D Slicer [153]. Pre-procedure lower torso CT scans were acquired. The CT images were rendered in the module with the standard coronal, sagittal, and transverse slice views as well as a 3D volume of the bony anatomy, segmented automatically by Slicer’s built-in volume renderer. Needle target and entry

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS



**Figure 7.12:** Workflow for transforaminal lumbar epidural injections.

points could be picked on any of the four views. A model needle was rendered in the 3D view according to the trajectory defined by the mentioned points and the needle projection was displayed on each slice view. Users had the option to switch to a “down-trajectory” view where the coronal view was replaced with a view perpendicular to the needle trajectory and the other two slice views were reformatted to provide views orthogonal to the down-trajectory view. These views, together with 3D rendering, provided opportunities to determine the amount of clearance between the planned needle trajectory and bone outline. An example screenshot of the surgeon’s interface is presented in Fig. 7.11(c).

## 7.5.2 Post-op Evaluation

We took postoperative CT scans and manually annotated the needle tip and base positions from the CT images. We reported the metrics of target point error, needle orientation error, and needle tip position relative to the safety zone of this application. Considering the spine shape mismatch of the post-operative and pre-operative CT scans, we performed a 3D/3D registration of each vertebra from post-op to pre-op CT. The annotated needlepoint positions were transformed into the pre-operative CT frame for comparison.

The annotation of the safety zone was performed on pre-operative CT scans under the instruction of experienced surgeons. We followed the definitions of both [51] and [154] manually annotate the Kambin triangle safety zone. In [51], the safety triangle is the inferior aspect of the pedicle and above the traversing nerve root. In [154], the Kambin triangle is defined by the hypotenuse, base, and height. The hypotenuse is the exiting nerve; the base is the caudad vertebral body, and the height is the traversing nerve root. Our safety zone is a combination of both these definitions. The safety zone for each injection trajectory target was manually segmented in 3D Slicer. We checked the needle tip positions relative to these safety zones in the postoperative CT scans as part of the evaluation.

### 7.5.3 System Setup and Calibration Method

The robotic system's end effector consisted of a custom-designed automated injection unit, attached to a 6-DOF UR-10 (Universal Robots, Odense, Denmark). The injection device is introduced in Chapter. 6.3 designed for femoro-plasty. In this work, we mainly used this device as a syringe holder. A custom-designed attachment between the syringe and needle was constructed to allow for the robotic system to leave a needle behind after placement with minimal perturbation and to allow for repeatable reloading of needles with minimal positional deviation. (Fig. 7.13). The syringe mount consisted of a plug with a female Luer lock and a receptacle with a male Luer lock, for which the receptacle was screwed onto the syringe and the needle was screwed into the plug. The mating tapers on each Luer lock connection ensured concentricity between needles, while the linear degree of freedom between the plug and receptacle, when unlocked, allowing for precise adjustment of the needles' axial position, to ensure that the length from the tip of the needle to the base of the injection device was consistent between trials. The forward kinematic accuracy of the robot is insufficient for this task. This insufficiency is further amplified by the weight of the injection unit and the long operating reach needed to perform injections on both sides of the specimen spine from L2 to the sacrum from a single position at the bedside. To ameliorate these inaccuracies, and NDI Polaris (Northern Digital Inc., Waterloo, Ontario, Canada) system was used to

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

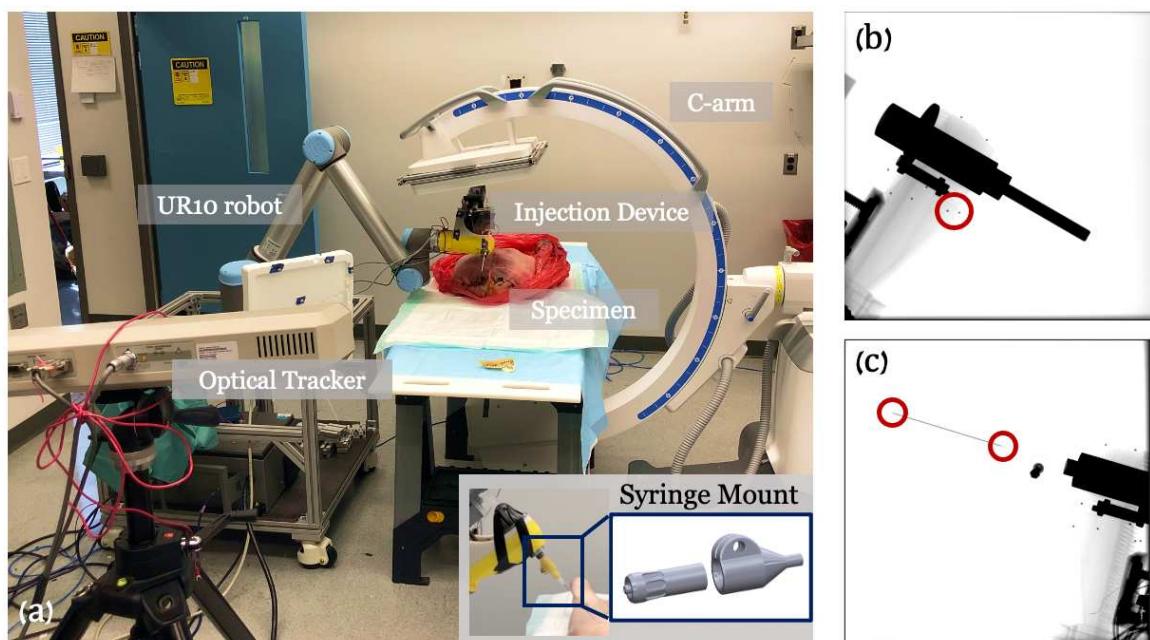
achieve closed-loop position control of the robotic system.

Our robotic injection system was navigated using pose estimations from X-ray image-based 2D/3D registration. Accurate calibration of the device registration model to the robot kinematic chain is required for automatic robot positioning and injection. To achieve closed-loop navigation, several calibrations were required: hand-eye calibration of the optical frame, hand-eye calibration of the injection device, and needle calibration. The two hand-eye calibrations of the injection device are introduced in Section 7.3.1. A one-time calibration was completed to determine the location and direction of the needle tip relative to the marker body on the injector. Ten X-ray images were taken with the injector and the needle in the view of the image. The needle tip and BB markers attached to the surface of the injector were annotated in each image (Fig.7.13(c)). These annotations were used when solving the optimization of the 3D location of the needle tip relative to the injector's coordinate frame.

### 7.5.4 System Calibration Results

We pre-operatively calibrated the needle base and tip positions in the injection device model frame using an example needle attached to the syringe mount. Six X-ray images were taken with variant C-arm poses. 2D needle tip, base ( $x_k^{tip}, x_k^{base}, k \in \{1, \dots, 6\}$ ) and metallic BB positions were manually annotated in each X-ray image. The C-arm pose ( $(T_{Carm}^D)_k^{pnp}, k \in \{1, \dots, 6\}$ ) was

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS



**Figure 7.13:** (a) Picture of robotic injection system setup including C-arm, UR-10, optical tracker, injection device, and a cadaveric specimen. Bottom right: Picture of the syringe mount. (b) An example X-ray image used for hand-eye calibration. Example BBs are marked in a red circle. (c) An example X-ray image used for needle calibration. The needle tip and base points are marked with red circles.

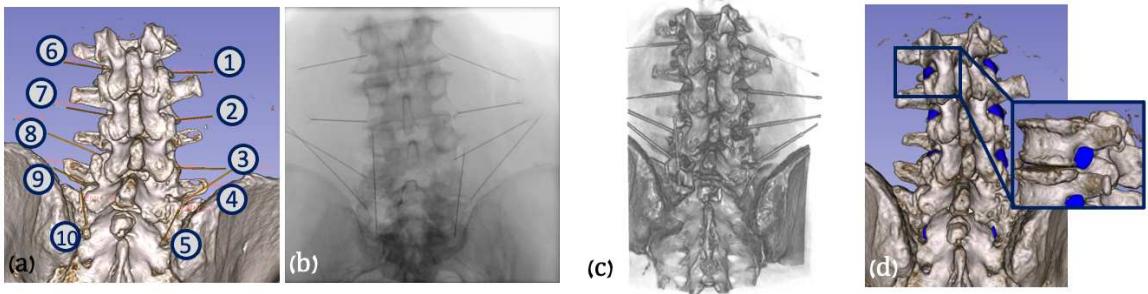
## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

estimated by solving the PnP problem using corresponding 2D and 3D BBs on the injection device. Using the projector operator ( $\mathcal{P}$ ), the 3D needle tip and base positions ( $p_D^{tip}, p_D^{base}$ ) were estimated by solving the following optimization:

$$\min_{p_D^{tip,base} \in \mathbb{R}^3} \sum_{k=1}^K \left( x_k^{tip,base}, \mathcal{P}(p_D^{tip,base}, (T_{Carm}^D)_k^{pnp}) \right). \quad (7.10)$$

The optimization was performed using brute force local search starting from a manual initialization point. We report the residual 2D error by calculating the  $l_2$  difference of the annotated needle tip and base points ( $x^{tip,base}$ ) and the reprojected points ( $\mathcal{P}(p_D^{tip,base}, (T_{Carm}^D)_k^{pnp})$ ) on each X-ray image. the mean 2D needle tip and base point errors were  $0.64 \pm 0.53$  mm and  $0.57 \pm 0.42$  mm, respectively.

In order to quantify the precision of the injection module, we performed a system testing experiment by adding weight load to the tip of the injection device and monitoring its shape deflection. Two optical fiducial markers were used for this testing: one was attached to the injection device tip, and the other one was the injection device marker itself. We applied increasing weight loads on the injection module tip uniformly from 0 to 2 kg. An NDI Polaris system was used to monitor the pose transformation of the two fiducial markers. We calculated the positional deviation of the tip marker relative to the injection device marker. We observed a mean relative deviation of  $0.14 \pm 0.08$  mm, and a maximum relative deviation of 0.29 mm.



**Figure 7.14:** (a) Screenshot of planning trajectories. (b) An example X-ray image was taken after the robotic needle injections. (c) Rendering of the Post-operative CT scans. (d) Illustration of the manually labeled safety zones.

### 7.5.5 Cadaveric Experiments and Results

Lower torso CT scan images of a male cadaveric specimen were acquired for fluoroscopic simulation and spine vertebrae registration. Vertebrae S1, L2, L3, L4, and L5 were segmented. An injection plan on this specimen was made by an expert clinician who also performed the procedure according to this plan, allowing for a comparison of performance to the robotic injection (Fig. 7.14 (a)). Ten injections were simulated via needle placement at five targets on each side of the specimen. Targets were the epidural spaces L2/3, L3/4, L4/5, L5/S1, and the first sacral foramen on each side. The target points were planned at the center of each safety zone.

We performed needle injections with our robotic system according to this plan under X-ray image-based navigation. The registration workflow was initialized using the PnP solutions from eight corresponding 2D and 3D anatomical landmarks. 3D landmarks were annotated pre-operatively on the CT scans.

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

2D landmarks were annotated intra-operatively after taking the registration X-rays. For the purpose of needle placement validation in this study, a small deviation from the proposed clinical workflow was performed in which needles were left within the specimen after placement. This allowed for the acquisition of a post-procedure CT to directly evaluate the needle targeting performance relative to the cadaveric anatomy with high fidelity. After the postoperative CT scan was taken, needles were removed and the needle placement was repeated by the expert clinician as his normal operation, using fluoroscopy as needed and another post-procedure CT was taken for evaluation. Fig. 7.14 presents a rendering of the postoperative CT scan and an X-ray image taken after the robotic injection.

We reported the needle injection performance using three metrics: needle tip error, needle orientation error, and safety zone. The needle tip error is calculated as the  $l_2$  distance between the planned trajectory target point and the injected needle tip point after registering vertebrae from post-operative CT to pre-operative CT. The orientation error was measured as the angle between trajectory vectors pointing along the long axis of the needle in its measured and planned positions. The results are summarized in Table. 7.2 and plotted in Fig. 7.15. Our robotic needle injection achieved a mean needle tip error of  $5.09 \pm 2.36$  mm and a mean orientation error of  $3.61 \pm 1.93$  degrees, compared to the clinical expert's performance of  $7.58 \pm 2.80$  mm and  $9.90 \pm 4.73$  mm, re-

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

spectively. The manually annotated safety zones in the postoperative CT scans are illustrated in Fig. 7.14 (d). All the injected needle tips, including both the robotic and clinician's injections, were within the safety zones.

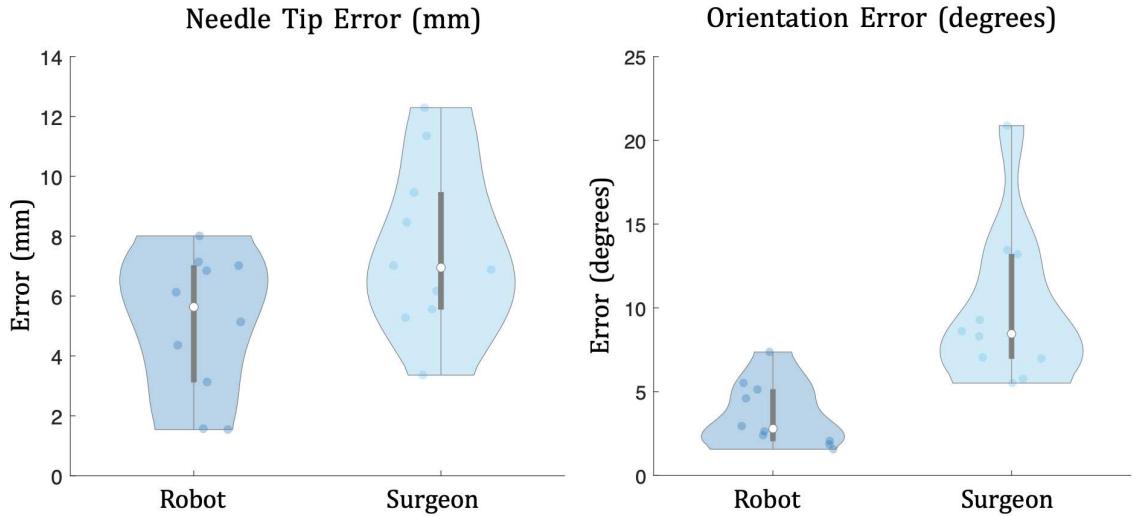
**Table 7.2:** Cadaveric Needle Injection Accuracy

ID	Needle Tip Error (mm)			Orientation Error (degrees)		
	Robot	Surgeon	Diff	Robot	Surgeon	Diff
1	3.13	9.46	6.33	5.13	5.77	0.64
2	6.13	11.35	5.22	1.85	8.30	6.45
3	7.02	6.17	-0.85	2.40	13.20	10.80
4	7.14	12.29	5.15	4.60	6.98	2.38
5	4.36	6.88	2.52	2.06	9.28	6.68
6	1.54	8.46	6.92	1.56	8.61	7.05
7	5.14	3.36	-1.78	2.63	7.05	4.42
8	8.01	7.02	-0.99	5.52	20.88	15.36
9	1.57	5.28	3.71	2.95	13.46	10.51
10	6.85	5.56	-1.29	7.37	5.51	-1.86
Mean	$5.09 \pm 2.36$	$7.58 \pm 2.80$	-	$3.61 \pm 1.93$	$9.90 \pm 4.73$	-
P value	0.02232			0.00049		

Diff is computed as the error magnitude difference between the surgeon's and robot's. The p value is reported from T-Test score. The result is significant at  $p < 0.05$ .

### 7.5.6 Discussion

Our robotic needle injection system is fiducial-free, using purely image information to close the registration loop, automatically position the needle injector to the planned trajectory and execute the injection. The robotic needle injection was navigated using multi-view X-ray 2D/3D registration. Our cadaver study experiments show the feasibility of using our system for trans-



**Figure 7.15:** Scatter violin plot of cadaver needle injection accuracy comparison between robot and surgeon.

foraminal lumbar epidural injections. Our comparison study with an expert clinician's manual injection using the same plan presents clear improvements in both translation and orientation accuracy: mean needle tip translation error of  $5.09 \pm 2.36$  mm and  $7.58 \pm 2.80$ , mean needle orientation error of  $3.61 \pm 1.93$  degrees and  $9.90 \pm 4.73$  degrees, corresponding to the robot and clinician's performance, respectively. The maximum needle tip translation and needle orientation errors of robot injection are 8.01 mm and 7.37 degrees, respectively. The experienced clinician's performance showed corresponding maximum errors of 12.29 mm and 20.88 degrees, respectively. The standard deviation of both translation and orientation errors of robot injection is lower than those of the clinician's injection. The P values suggest that the results are significant at  $p < 0.05$ . We also evaluated the performance using the defined safety zone for

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

this application. Both the robotic and clinician's injected needle tips were laid inside the safety zones. Although the expert clinician's injection tip error and orientation error are larger, this manual injection's accuracy is still sufficient for this application. However, the robotic performance of higher accuracy and stability demonstrates a potential reduction of the risk of violating the safety zone.

We also looked at the individual contributions of errors due to hand-eye calibration and registration. The needle tip error due to registration as compared to planning was  $2.82 \pm 2.61$  mm. The needle tip error resulting from hand-eye calibration was  $2.49 \pm 1.55$  mm (Section. 7.3.1). Our system testing result shows that the injection module tip precision is within the magnitude of 0.3 mm. This suggests that the potential needle tip deflection due to the relatively large distance between the tip and injection module is not significant. The other factor affecting the overall error is that we performed calibration only for one needle and did not repeat for successive injections with different needles. Calibrations after changing each needle may also help to reduce the reported translation error. The above system error investigation suggests that this non-purpose-built injection unit does not introduce significant system error due to its mechanical design. An improved injection device with a smaller body, lighter mass, and more rigid material would perform better.

One common concern of the fluoroscopic navigation system is the excessive

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

radiation exposure to the patient. Our approach requires ten to twelve X-rays to register the patient to our injection device. Considering X-rays are commonly used in the clinician’s manual injections to checking the needle position, this amount of radiation is acceptable for this procedure. Since our pipeline is designed to be fully automated and the injections are delivered by the robot, the clinicians can supervise the procedure remotely and thus are not affected by the radiation exposure. The X-ray data acquisition including various robot configuration positioning took about 15 minutes in our cadaver study. The full registration pipeline takes around 5 minutes. Our current implementation required a few manual annotations from the clinician to initialize the registration. The overall data collection and registration time was about 30 minutes. Future work may consider automating the data collection and intra-operative landmark detection to further simplify the registration workflow, similar to our work reported in [59, 67]. However, once the system is registered and positioned, the automated robotic single injection takes less than 10 seconds. The time saving is more significant if more injections are planned on the same patient.

In this study, needle steering was neglected. This is a widely studied topic, and such functionality could be added in future work and may improve results. The decision to not consider needle steering was made as 1) the focus of this work was on the novel application of the registration techniques used to the

## CHAPTER 7. SYSTEM INTEGRATION AND CLINICAL APPLICATIONS

spine and correction via needle steering could mask inaccuracies of the registration. 2) The relatively large injection target area does not necessitate sub-millimeter accuracy. 3) The use of stiff low gauge needles in this application limits bending in soft tissue, reducing both the need for and the effect of needle steering. We are aware that there are some studies on the clinical effects of spinal screw fixation [155] and needle steering [156, 157]. We want to emphasize that the clinical requirement of TLESI is positioning the needle tip within the safety triangle zone. Requirements for screw placement are more strict compared to our application. Based on our experiment results, the deflection of the needle does not affect our accuracy to go outside the safety zone.

The current methodology design requires the patient to be stationary during the registration procedure without dealing with the potential patient's motion before or during the needle injection. In the future, we plan to take more intermediate fluoroscopic images during the needle injection phase to verify the registration pose estimation. If the trajectory is detected to be off with respect to the plan, the needle can be moved out and redo the injection using the updated registration.

## 7.6 Conclusion

In this chapter, we present the methods and results of calibrating the injection device and continuum manipulator to the robotic system. Overviews of the proposed navigation systems are shown in each application section. Our cadaveric experiment results showed the feasibility to apply the proposed navigation system for robot assisted orthopedic applications.

## 7.7 Acknowledgement

Sincere thanks to Mr. Henry Phalen for developing the closed-loop robot control software module and the planning module in the lumbar epidural injection experiment. Sincere thanks to Mrs. Mahsan Bakhtiarinejad for her work in the biomechanical planning in the femoroplasty experiment. I want to thank Dr. Shahriar Sefati and Mr. Justin Ma for their help in system calibration testing. I express my great appreciation to Dr. Amirhossein Farvardin, Mrs. Mahsan Bakhtiarinejad, Mr. Henry Phalen, Mr. Adam Margalit, Dr. Liuhong Ma, and Ms. Mareike Thies for their contributions in the cadaveric experiments. Special thanks to Dr. David Cohen and Akhil Chhatre for their help in performing manual cadaveric epidural lumbar injections. Finally, I thank Drs. Mehran Armand, Mathias Unberath, Russell Taylor, Amit Jain, and Julius Oni for their insightful comments and feedback.

# **Chapter 8**

## **Conclusion and Future Work**

### **8.1 Conclusion**

In this thesis, we have presented fiducial-free fluoroscopic navigation pipelines for three robot-assisted orthopedic applications. Our efforts can be divided into two categories, one using traditional intensity-based 2D/3D registration methods as a backbone to design the navigation pipeline for accurate surgical tool-to-tissue pose estimation, and one using advanced machine learning techniques for automating the navigation process such as the initialization or addressing the challenges of conventional 2D/3D registration approaches such as the limited capture range. We try to demonstrate the effectiveness of the proposed methods in clinical usage by integrating them into a robotic surgical system and performing *ex vivo* cadaveric experiments.

## CHAPTER 8. CONCLUSION AND FUTURE WORK

The contributions of this thesis can be summarized as follows:

- In Chapter 2, we introduced SyntheX, a proposed synthetic training data generation pipeline from human models for machine learning X-ray image analysis tasks. We used SyntheX to train feature detection deep neural networks for extracting anatomical landmarks from pelvic X-rays and distinct features from the continuum manipulator X-rays. These features were used to automatically initialize the pelvis and continuum manipulator 2D/3D registration pipelines, respectively.
- In Chapter 3, we presented multi-view intensity-based 2D/3D registration methods for pose estimation of femur and spine vertebrae. The methods are fiducial-free and produce accurate intra-operative pose transformations of the bone anatomies with respect to the C-arm source frame.
- In Chapter 4, we presented a novel projective spatial transformer module called ProST. Using the differentiable property of ProST, we proposed an end-to-end deep neural network architecture that learns to approximate a convex shape similarity function for iterative 2D/3D registration. We demonstrated this as a promising solution to largely extend the capture of conventional intensity-based registration methods.
- In Chapter 5, we introduced 2D/3D registration algorithms for two types of surgical tools: a rigid injection device and a flexible continuum manip-

## CHAPTER 8. CONCLUSION AND FUTURE WORK

ulator. The algorithms estimate pose and shape of the surgical devices.

Combining pose estimations of the bone anatomies, we computed the critical tool-to-tissue relationship using purely fluoroscopic images.

- In Chapter 6, we presented our efforts on system-level validation and *ex vivo* clinical experiments. We described the system calibration methods and results. We also analyzed the navigation system performance regarding clinical requirements and metrics.

## 8.2 Future Work

We have demonstrated the benefits of image-based navigation for robot-assisted orthopedic interventions. To bring such benefits from laboratory experiments to the operating room, several directions can be considered for future work.

Registration initialization still presents the bottleneck of the navigation pipeline that requires human supervision to ensure success, which interrupts the procedure. Chapters 3 and 5 present learning-based landmark detection methods for automatic registration initialization of the pelvis and continuum manipulator. We did not investigate such automatic initialization strategies for the spine and injection device. These registrations are initialized using manual landmark annotations. In Chapter 4, the work of ProST presents an

## CHAPTER 8. CONCLUSION AND FUTURE WORK

encouraging alternative approach by learning a convex similarity function in a data-driven fashion. However, this approach appears to be most effective when the object pose is far away from ground truth, but becomes ambiguous when the object is close. Future directions can include improving the learning-based similarity function to be robust on both large offsets and small local differences.

Our multi-view 2D/3D registration geometry was manually defined to approximate the views in common clinician's practices, such as an AP view of the anatomy. However, this view may not be optimal to register the surgical tool when it moves and rotates. For example, as the continuum manipulator drills and deforms its shape, its X-ray projection appearance changes drastically with respect to the C-arm view geometry. Future studies can research on finding the optimal C-arm geometry where the registration has the highest chance to succeed. Our navigation system does not estimate the confidence or uncertainty of the registration result. The success or failure of the registration is determined by interpreting the edge overlay of the DRR image and target X-ray image. In Chapter 4, we have discussed the correlation between the ProST learned similarity score and registration errors, which suggests being an indicator of uncertainty. Future work can include research on modeling the registration uncertainty to reflect the registration results.

Our proposed navigation pipeline assumes the patient to be stationary during registration and robot positioning. We did not account for the potential

## CHAPTER 8. CONCLUSION AND FUTURE WORK

patient motion. In Chapter 5, we have discussed that we can use more verification X-ray images to rerun the registration after the surgical tool is positioned. Future work must investigate methods to detect patient motion and strategies to rerun the verification registration. For the application of continuum manipulator, which involves continuous drilling, it would be helpful to design algorithms for “real-time” shape estimation using consecutive fluoroscopic images taken with a specific frequency, or using imaging modality to verify and improve the estimations of other shape sensing units.

From a broader point of view, the back end registration and navigation technology has the potential to advance intra-operative surgical visualizations with augmented reality (AR) [158] and mixed reality (MR) [159]. The fiducial-free tool to tissue relationship estimation provides a non-invasive solution to overlay the pre-operative data to the intra-operative patient. After taking a few X-ray images using a mobile C-arm, the navigation system guides the robotic tool to an entry position. The current 3D positional visualizations and the following surgical operations such as drilling and injection can be simulated and displayed to the surgeon, which helps the surgeon to review the plan. The future of fluoroscopic navigation lies in blending the computational X-ray image information to robot and human interaction with the goal of improving patient outcomes.

# **Appendix A**

## **Supplementary Material for Chapter 3**

We introduce further details on the domain randomization and domain adaptation methods applied in our SyntheX controlled studies. We then provide additional information on experimental setup and network training details of the clinical tasks and benchmark investigations.

### **A.0.1 Domain Randomization**

Domain randomization effects were applied to the input images during network training. We studied two domain randomization levels: regular and strong domain randomization. Regular domain randomization included the most commonly used data augmentation schemes. For strong domain random-

## APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

ization, we included more drastic effects and combined them together. We use  $x$  to denote a training image sample. The domain randomization techniques we introduced are as follows:

Regular domain randomization included: 1) *Gaussian noise injection*:  $x + N(0, \sigma)$ , where  $\sigma$  was uniformly chosen from the interval  $(0.005, 0.1)$  multiplied by the image intensity range. 2) *Gamma transform*:  $\text{norm}(x)^\gamma$ , where  $x$  was normalized by its maximum and minimum value and  $\gamma$  was uniformly selected from the interval  $(0.7, 1.3)$ . 3) *Random crop*  $x$  was cropped at random locations using a square shape which has the dimension of 90%  $x$  size. Regular domain randomization methods were applied to every training iteration.

Strong domain randomization included: 1) *Inverting*:  $\max(x) - x$ , where the maximum intensity value was subtracted from all image pixels. 2) *Impulse/Pepper/Salt noise injection*: 10% of pixels in  $x$  were replaced with one type of noise including impulse, pepper and salt. 3) *Affine transform*: a random 2D affine warp including translation, rotation, shear and scale factors was applied. 4) *Contrast*:  $x$  was processed with one type of the contrast manipulations including linear contrast, log contrast and sigmoid contrast. 5) *Blurring*:  $x$  was processed with a blurring method including Gaussian blur  $N(\mu = 0, \sigma = 3.0)$  and average blur (kernel size between  $2 \times 2$  and  $7 \times 7$ ). 6) *Box corruption*: a random number of box regions were corrupted with large noise. 7) *Dropout*: Either randomly dropped 1-10% of pixels in  $x$  to zero, or dropped them in a

## APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

rectangular region with 2-5% of the image size. 8) *Sharpening and embossing*: Sharpen  $x$  blended the original image with a sharpened version with an alpha between 0 and 1 (no and full sharpening effect). Embossing added the sharpened version rather than blending it. 9) One of the pooling methods was applied to  $x$ : average pooling, max pooling, min pooling and median pooling. All of the pooling kernel sizes were between  $2 \times 2$  and  $4 \times 4$ . 10) *Multiply*: Either changed brightness or multiplied  $x$  element wise with 50-150% of original value. 11) *Distort*: Distorted local areas of  $x$  with a random piece-wise affine transformation. For each image, we still applied basic domain randomization but only randomly concatenated up to two strong domain randomization methods during each training iteration to avoid too heavy augmentation.

### A.0.2 Domain Adaptation

We select the two most commonly used domain adaptation approaches for our comparison study, which are CycleGAN [68] and adversarial discriminative domain adaptation (ADDA) [69]. CycleGAN was trained using unpaired synthetic and real images prior to task network training. All synthetic images were then processed with trained CycleGAN generators, to alter their appearance to match real data. We strictly enforced the data split used during task-model training so that images from the test set were excluded during both CycleGAN and task network training. ADDA introduced an adversarial

## APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

discriminator branch as an additional loss to discriminate between features derived from synthetic and real images. We followed the design of Haq et al. to build the discriminator for ADDA training on the task of semantic segmentation [69]. Both CycleGAN and ADDA models were tested using realistic and naïve simulation images.

### A.0.2.1 CycleGAN

CycleGAN was applied to learn mapping functions between two image domains  $X$  and  $Y$  given training samples  $\{x_i\}_{i=1}^N$  where  $x_i \in X$  and  $\{y_j\}_{j=1}^M$  where  $y_j \in Y$ . The model includes two mapping functions  $G : X \rightarrow Y$  and  $F : Y \rightarrow X$ , and two adversarial discriminators  $D_X$  and  $D_Y$ . The objective contains two terms: *adversarial loss* to match the distribution between generated and target image domain; and *cycle consistency loss* to ensure learned mapping functions are cycle-consistent.

For one mapping function  $G : X \rightarrow Y$  with its discriminator  $D_Y$ , the first term, *adversarial loss*, can be expressed as:

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) &= \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ &\quad + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))] \end{aligned} \tag{A.1}$$

## APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

where  $G$  generates images  $G(x)$  with an appearance similar to images from domain  $Y$ , while  $D_Y$  tries to distinguish between translated samples  $G(x)$  and real samples  $y$ . Overall,  $G$  aims to minimize this objective against an adversary  $D$  that tries to maximize it. Similarly, there is an *adversarial loss* for the mapping function  $F : Y \rightarrow X$  with its discriminator  $D_X$ .

The second term, *cycle consistency loss*, can be expressed as:

$$\begin{aligned} \mathcal{L}_{\text{cyc}}(G, F) = & \mathbb{E}_{x \sim p_{\text{data}}(x)} [F(G(x)) - x_1] \\ & + \mathbb{E}_{y \sim p_{\text{data}}(y)} [G(F(y)) - y_1], \end{aligned} \quad (\text{A.2})$$

where for each image  $x$  from domain  $X$ ,  $x$  should be recovered after one translation cycle, i.e.,  $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ . Similarly, each image  $y$  from domain  $Y$  should be recovered as well. Zhu et al. [68] argued that learned mapping functions should be cycle-consistent to further reduce the space of possible mapping functions. The above formulation using domain discrimination and cycle consistency enables unpaired image translation, i.e., learning the mappings  $G(x)$  and  $F(y)$  without corresponding images.

## APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

The overall objective for CycleGAN training is expressed as:

$$\begin{aligned} \mathcal{L}(G, F, D_X, D_Y) = & \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ & + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ & + \lambda \mathcal{L}_{\text{cyc}}(G, F), \end{aligned} \quad (\text{A.3})$$

where  $\lambda$  controls the relative importance of cycle consistency loss, aiming to solve:

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y). \quad (\text{A.4})$$

For the generator network, 6 blocks for  $128 \times 128$  images and 9 blocks for  $256 \times 256$  and higher-resolution training images were used with instance normalization. For the discriminator network, a  $70 \times 70$  PatchGAN was used.

### A.0.2.2 Adversarial Discriminative Domain Adaptation

We applied the idea of Haq et al. on our pelvis segmentation and landmark localization task [69]. The architecture consists of three components, including *Segmentation and Localization network*, *Decoder* and *Discriminator*. The input to *Segmentation and Localization network* is image ( $x$ ) and the output prediction feature is  $z$ . The loss is  $L_{\text{seg}}$  and  $L_{\text{ld}}$  as introduced in Section ???. The *Decoder* shared the same U-Net architecture, takes  $z$  as input and the output is the reconstruction  $R(z)$ . The reconstruction loss,  $L_{\text{recons}}$ , is the mean squared

## APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

error between  $x$  and  $z$ . The *Discriminator* was trained using an adversarial loss:

$$L_{dis}(z) = -\frac{1}{H \times W} \sum_{h,w} s \cdot \log(D(z)) + (1-s) \cdot \log(1 - D(z)), \quad (\text{A.5})$$

where  $H$  and  $W$  are the feature dimension of the discriminator output,  $s = 0$  when  $D$  takes target domain prediction ( $Y_t$ ) as input, and  $s = 1$  when taking source domain prediction ( $Y_s$ ) as input. The *Discriminator* contributed an adversarial loss during training in order to bring in domain transfer knowledge.

The adversarial loss is defined as:

$$L_{adv}(x_t) = -\frac{1}{H \times W} \sum_{h,w} \log(D(z_t)). \quad (\text{A.6})$$

Thus, the total training loss can be written as:

$$L_t(x_s, x_t) = L_{seg}(x_s) + L_{ld}(x_s) + \lambda_{adv} L_{adv}(x_t) + \lambda_{recons} L_{recons}(x_t), \quad (\text{A.7})$$

where  $\lambda_{adv}$  and  $\lambda_{recons}$  are weight hyperparameters, which are empirically chosen to be 0.001 and 0.01, as suggested by Haq et al. [69].

## APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

### A.0.3 Network Training Details

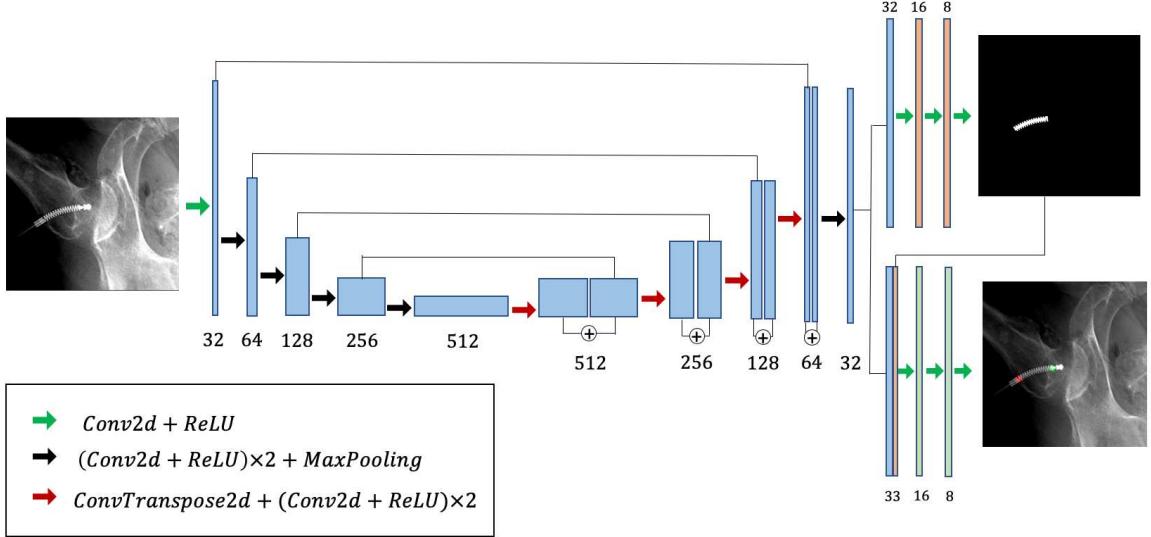
We used Pytorch for all implementations and trained the networks from scratch, using stochastic gradient descent with a learning rate of 0.0002, Nesterov momentum of 0.9, weight decay of 0.00001, and a constant batch size of five images. The multi-task network training loss is equally weighted between landmark detection loss and segmentation loss. For all ablation experiments in the precisely controlled study, we kept the same learning rate for the first 100 epochs and linearly decayed the rate to zero over the next 100 epochs.

## **Appendix B**

# **Supplementary Materials for Chapter 6**

Inspired by the work of [137], we design a ConvNet-based auto-encoder like architecture with skip connections, and split the connection from the last feature layer to perform two tasks concurrently, segmentation and landmark detection. Fig. B.1 illustrates the ConvNet architecture used here. In the decoder, we repeat the connection of 2D convolutional layer and maxpooling layer four times to abstract a feature representation with 512 channels. In the decoder part, we concatenate the upsampled features and features from the same level in the encoding stage. The final decoded 32 channel feature layer is shared across the segmentation path and the localization path. The final output of the segmentation mask is backward concatenated with this shared feature to boost

the localization task.



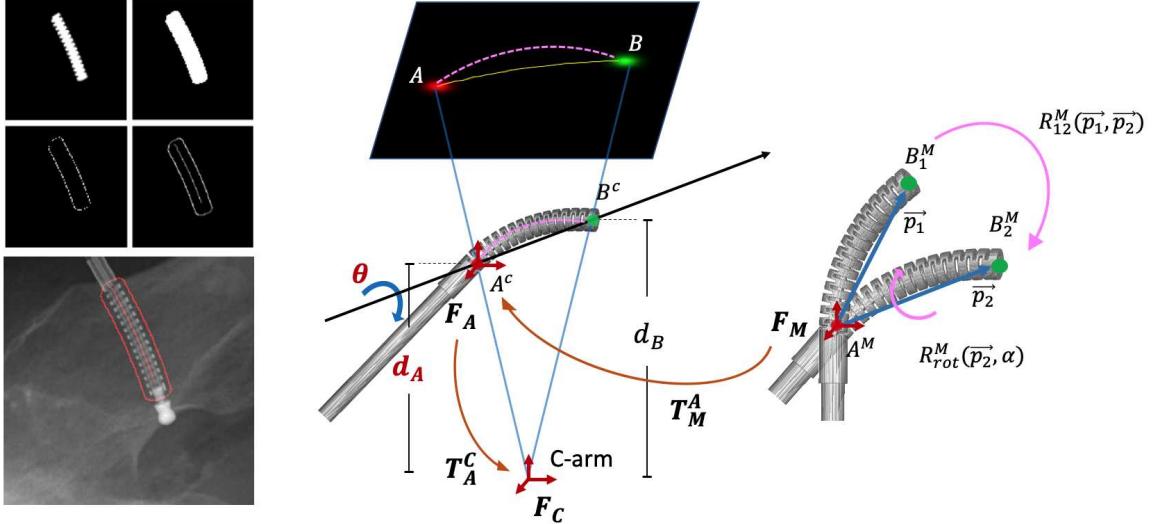
**Figure B.1:** Network architecture used for concurrent segmentation and landmark detection.

## B.1 Centerline-based 2D/3D Registration

### B.1.1 Centerline Extraction

The centerline is extracted from the predicted segmentation mask. We first use a spherical ball ( $diameter = 5$ ) to dilate the original mask and smooth out the notches. We then compute the mask outline via differentiation to calculate a distance map that stores, for every point in the mask, the distance to the closest point on the edge. Using the two corresponding landmarks, we run Dijkstra's algorithm [160] to find the optimal path from start to end landmark

## APPENDIX B. SUPPLEMENTARY MATERIALS FOR CHAPTER 6



**Figure B.2:** Centerline-based 2D/3D registration. (a): Extraction of the 2D centerline from a segmentation mask consists of dilation, edge and distance map computation, and finally, shortest path extraction. (b): Geometric relations of curve-based 2D/3D registration. (c): Coordinate transformation from world frame  $F_M$  to camera frame  $F_C$ . 3D positions of  $A^C$  and  $B^C$  in  $F_C$  are defined as  $\vec{p}_A = (x_A^C, y_A^C, d_A)$  and  $\vec{p}_B = (x_B^C, y_B^C, d_B)$ .  $\vec{p}_{AB} = \vec{p}_B - \vec{p}_A$ . In  $F_M$ , we define vector  $\vec{p}_2$ , which is parallel to  $\vec{p}_{AB}$  and has the same length, starting at the origin, which defines as  $A^W$  (the first landmark), ending at  $B_2^W$  (the second landmark).

which defines the 2D centerline of the CM,  $l_{cen}^{img}$ . Fig. B.2 illustrates centerline extraction process.

### B.1.2 Centerline-based 2D/3D Registration

$A$  and  $B$  in the 2D projection image are the landmarks detected in the X-ray image, which correspond to  $A^C$  and  $B^C$  on the 3D model. In the 2D image domain, we denote the coordinates of  $A$  and  $B$  as  $(u_A, v_A)$  and  $(u_B, v_B)$  while in the 3D camera coordinate space, we denote the coordinates of  $A^C$  and  $B^C$  as

## APPENDIX B. SUPPLEMENTARY MATERIALS FOR CHAPTER 6

$(x_A^C, y_A^C, d_A)$  and  $(x_B^C, y_B^C, d_B)$ . Then, from the X-ray camera projection model, we have

$$\frac{x_A^C}{u_A s_p} = \frac{y_A^C}{v_A s_p} = \frac{d_A}{D}, \quad (\text{B.1})$$

$$\frac{x_B^C}{u_B s_p} = \frac{y_B^C}{v_B s_p} = \frac{d_B}{D}, \quad (\text{B.2})$$

where  $s_p$  is the pixel size on detector, and  $D$  is the source-detector distance.  $L$  is the distance between  $A^C$  and  $B^C$  which can be calculated using:

$$(x_A^C - x_B^C)^2 + (y_A^C - y_B^C)^2 + (d_A - d_B)^2 = L^2. \quad (\text{B.3})$$

Combining (B.1)(B.2) with (B.3), we have

$$\begin{aligned} & \left( \frac{u_A s_p}{D} d_A - \frac{u_B s_p}{D} d_B \right)^2 + \left( \frac{v_A s_p}{D} d_A - \frac{v_B s_p}{D} d_B \right)^2 \\ & + (d_A - d_B)^2 = L^2. \end{aligned} \quad (\text{B.4})$$

Eq. B.4 builds the relationship between  $d_A$  and  $d_B$ .  $d_A$  is the 1 DoF variable to determine the 3D positions of  $A^C$  and  $B^C$ . Then, the only remaining DoF is the rotation angle  $\theta$  w.r.t. the axis connection defined by  $A^C$  and  $B^C$ . Fig. B.2 shows that the rigid transformation from the CM model frame ( $F_M$ ) to C-arm source frame ( $F_C$ ) can be uniquely determined by  $\theta$  and  $d_A$  using:

$$T_M^A = R_{rot}^M(\vec{p}_{AB}, \theta) \cdot R_{12}^M(\vec{p}_1, \vec{p}_2), \quad (\text{B.5})$$

## APPENDIX B. SUPPLEMENTARY MATERIALS FOR CHAPTER 6

where  $R_{rot}^M(\vec{p}_{AB}, \theta)$  is the rotation matrix determined by the rotation of angle  $\theta$  about axis  $\vec{p}_{AB}$ , and  $R_{12}^M(\vec{p}_1, \vec{p}_2)$  is the rotation matrix from vector  $\vec{p}_1$  to  $\vec{p}_2$ . Transformation from  $F_A$  to  $F_C$  can be find with  $\vec{p}_A$  as  $T_A^C(\vec{p}_A)$ . Thus, the extrinsic transformation matrix is calculated as

$$T_M^C = T_A^C(\vec{p}_A)R_{rot}^M(\vec{p}_{AB}, \theta)R_{12}^M(\vec{p}_1, \vec{p}_2), \quad (\text{B.6})$$

where  $R_{rot}^M(\vec{p}_2, \theta)$  is the same as  $R_{rot}^C(\vec{p}_{AB}, \theta)$ .

The values of  $d_A$  and  $\theta$  are determined using a search strategy. The search space of  $\theta$  is  $[0, 2\pi]$ . The search space of  $d_A$  covers valid solutions that suffice (B.3). For each pair of  $(\theta, d_A)$ , we uniformly sample 100 points along the 3D centerline and project the points to the 2D image plane using the projection geometry. The loss is the mean  $l_2$  distance from the projected points to the extracted centerline  $l_{cen}^{img}$ . The optimal values of  $\theta$  and  $d_A$  correspond to the minimal loss within the search space.

# Bibliography

- [1] A. Farvardin, E. Basafa, M. Bakhtiarinejad, and M. Armand, “Significance of preoperative planning for prophylactic augmentation of osteoporotic hip: A computational modeling study,” *Journal of biomechanics*, 2019.
- [2] J. C. Mandell, G. J. Czuczman, G. C. Gaviola, V. Ghazikhanian, and C. H. Cho, “The lumbar neural foramen and transforaminal epidural steroid injections: an anatomic review with key safety considerations in planning the percutaneous approach,” *American Journal of Roentgenology*, vol. 209, no. 1, pp. W26–W35, 2017.
- [3] C. Gao, A. Farvardin, R. B. Grupp, M. Bakhtiarinejad, L. Ma, M. Thies, M. Unberath, R. H. Taylor, and M. Armand, “Fiducial-free 2d/3d registration for robot-assisted femoroplasty,” *IEEE transactions on medical robotics and bionics*, vol. 2, no. 3, pp. 437–446, 2020.
- [4] C. Nikou, B. Jaramaz, A. M. DiGioia, and T. J. Levison, “Description of

## BIBLIOGRAPHY

- anatomic coordinate systems and rationale for use in an image-guided total hip replacement system,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2000, pp. 1188–1194.
- [5] S. Sefati, R. Hegeman, F. Alambeigi, I. Iordachita, P. Kazanzides, H. Khanuja, R. H. Taylor, and M. Armand, “A surgical robotic system for treatment of pelvic osteolysis using an fbg-equipped continuum manipulator and flexible instruments,” *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 1, pp. 369–380, 2020.
- [6] R. H. Taylor, B. D. Mittelstadt, H. A. Paul, W. Hanson, P. Kazanzides, J. F. Zuhars, B. Williamson, B. L. Musits, E. Glassman, and W. L. Bargar, “An image-directed robotic system for precise orthopaedic surgery,” *IEEE Transactions on Robotics and Automation*, vol. 10, no. 3, pp. 261–275, 1994.
- [7] W. Bargar, A. DiGioia, R. Turner, J. Taylor, J. McCarthy, and D. Mears, “Robodoc multi-center trial: an interim report,” in *Proc. 2nd Int. Symp. on Medical Robotics and Computer Assisted Surgery*, 1995, pp. 208–214.
- [8] W. L. Bargar, A. Bauer, and M. Börner, “Primary and revision total hip replacement using the robodoc (r) system.” *Clinical Orthopaedics and Related Research (1976-2007)*, vol. 354, pp. 82–91, 1998.

## BIBLIOGRAPHY

- [9] J. Lang, S. Mannava, A. Floyd, M. Goddard, B. Smith, A. Mofidi, T. M. Seyler, and R. Jinnah, “Robotic systems in orthopaedic surgery,” *The Journal of bone and joint surgery. British volume*, vol. 93, no. 10, pp. 1296–1299, 2011.
- [10] B. Innocenti and E. Bori, “Robotics in orthopaedic surgery: why, what and how?” *Archives of orthopaedic and trauma surgery*, vol. 141, no. 12, pp. 2035–2042, 2021.
- [11] C. H. Ewurum, Y. Guo, S. Pagnha, Z. Feng, and X. Luo, “Surgical navigation in orthopedics: workflow and system review,” *Intelligent Orthopaedics*, pp. 47–63, 2018.
- [12] M. Rednam and V. Tiwari, “Fluoroscopy orthopedic assessment, protocols, and interpretation,” 2021.
- [13] L. Kausch, S. Thomas, H. Kunze, M. Privalov, S. Vetter, J. Franke, A. H. Mahnken, L. Maier-Hein, and K. Maier-Hein, “Toward automatic c-arm positioning for standard projections in orthopedic surgery,” *International journal of computer assisted radiology and surgery*, vol. 15, no. 7, pp. 1095–1105, 2020.
- [14] E. Lungu and T. P. Moser, “A practical guide for performing arthrography under fluoroscopic or ultrasound guidance,” *Insights into imaging*, vol. 6, no. 6, pp. 601–610, 2015.

## BIBLIOGRAPHY

- [15] R. K. Chaudhary, S. Acharya, R. S. Chahal, and K. L. Kalra, “Fluoroscopy guided percutaneous transpedicular biopsy of vertebral body lesion,” *Journal of Nepal Health Research Council*, vol. 17, no. 2, pp. 163–167, 2019.
- [16] Y. Fan, J. P. Du, J. J. Liu, J. N. Zhang, H. H. Qiao, S. C. Liu, and D. J. Hao, “Accuracy of pedicle screw placement comparing robot-assisted technology and the free-hand with fluoroscopy-guided method in spine surgery: an updated meta-analysis,” *Medicine*, vol. 97, no. 22, 2018.
- [17] N. Sarkalkan, H. Weinans, and A. A. Zadpoor, “Statistical shape and appearance models of bones,” *Bone*, vol. 60, pp. 129–140, 2014.
- [18] R. B. Grupp, R. A. Hegeman, R. J. Murphy, C. P. Alexander, Y. Otake, B. A. McArthur, M. Armand, and R. H. Taylor, “Pose estimation of periacetabular osteotomy fragments with intraoperative x-ray navigation,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 441–452, 2019.
- [19] D. C. Liu and J. Nocedal, “On the limited memory bfgs method for large scale optimization,” *Mathematical programming*, vol. 45, no. 1, pp. 503–528, 1989.
- [20] M. Berger, K. Müller, A. Aichert, M. Unberath, J. Thies, J.-H. Choi, R. Fahrig, and A. Maier, “Marker-free motion correction in weight-

## BIBLIOGRAPHY

- bearing cone-beam ct of the knee joint," *Medical physics*, vol. 43, no. 3, pp. 1235–1248, 2016.
- [21] N. Hansen, S. D. Müller, and P. Koumoutsakos, "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es)," *Evolutionary computation*, vol. 11, no. 1, pp. 1–18, 2003.
- [22] M. J. Powell, "The bobyqa algorithm for bound constrained optimization without derivatives," *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, pp. 26–46, 2009.
- [23] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, p. 155, 2009.
- [24] M. Unberath, C. Gao, Y. Hu, M. Judish, R. H. Taylor, M. Armand, and R. Grupp, "The impact of machine learning on 2d/3d registration for image-guided interventions: A systematic review and perspective," *Frontiers in Robotics and AI*, p. 260, 2021.
- [25] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

## BIBLIOGRAPHY

- [26] A. Uneri, Y. Otake, A. Wang, G. Kleinszig, S. Vogt, A. J. Khanna, and J. Siewerdsen, “3d–2d registration for surgical guidance: effect of projection view angles on registration accuracy,” *Physics in Medicine & Biology*, vol. 59, no. 2, p. 271, 2013.
- [27] Y. Otake, A. S. Wang, J. W. Stayman, A. Uneri, G. Kleinszig, S. Vogt, A. J. Khanna, Z. L. Gokaslan, and J. H. Siewerdsen, “Robust 3d–2d image registration: application to spine interventions and vertebral labeling in the presence of anatomical deformation,” *Physics in Medicine & Biology*, vol. 58, no. 23, p. 8535, 2013.
- [28] C. Gao, B. D. Killeen, Y. Hu, R. B. Grupp, R. H. Taylor, M. Armand, and M. Unberath, “Synthex: Scaling up learning-based x-ray image analysis through in silico experiments,” *arXiv preprint arXiv:2206.06127*, 2022.
- [29] C. Gao, R. B. Grupp, M. Unberath, R. H. Taylor, and M. Armand, “Fiducial-free 2d/3d registration of the proximal femur for robot-assisted femoroplasty,” in *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11315. International Society for Optics and Photonics, 2020, p. 113151C.
- [30] C. Gao, H. Phalen, M. Adam, M. H. Justin, K. Ping-Cheng, M. Unberath, R. H. Taylor, J. Amit, and M. Armand, “Fluoroscopy-guided robotic sys-

## BIBLIOGRAPHY

- tem for transforaminal lumbar epidural injections,” *IEEE Transactions on Medical Robotics and Bionics*, no. doi: 10.1109/TMRB.2022.3196321.
- [31] C. Gao, X. Liu, W. Gu, B. Killeen, M. Armand, R. Taylor, and M. Unberath, “Generalizing spatial transformers to projective geometry with applications to 2d/3d registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 329–339.
- [32] C. Gao, H. Phalen, S. Sefati, J. Ma, R. H. Taylor, M. Unberath, and M. Armand, “Fluoroscopic navigation for a surgical robotic system including a continuum manipulator,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 453–464, 2021.
- [33] C. Gao, A. Farvardin, R. B. Grupp, M. Bakhtiarinejad, L. Ma, M. Thies, M. Unberath, R. H. Taylor, and M. Armand, “Fiducial-free 2d/3d registration for robot-assisted femoroplasty,” *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 437–446, 2020.
- [34] M. J. Goldacre, S. E. Roberts, and D. Yeates, “Mortality after admission to hospital with fractured neck of femur: database study,” *Bmj*, vol. 325, no. 7369, pp. 868–869, 2002.
- [35] G. G. Teng *et al.*, “Mortality and osteoporotic fractures: is the link causal,

## BIBLIOGRAPHY

- and is it modifiable?" *Clinical and experimental rheumatology*, vol. 26, no. 5 0 51, p. S125, 2008.
- [36] A. Dinah, "Sequential hip fractures in elderly patients," *Injury*, vol. 33, no. 5, pp. 393–394, 2002.
- [37] J. Land, L. Russell, and S. Khan, "Osteoporosis," *Clinical Orthopaedics and Related Research*, vol. 372, pp. 139–150, 2000.
- [38] E. Basafa and M. Armand, "Subject-specific planning of femoroplasty: a combined evolutionary optimization and particle diffusion model approach," *Journal of biomechanics*, vol. 47, no. 10, pp. 2237–2243, 2014.
- [39] E. Basafa, R. J. Murphy, Y. Otake, M. D. Kutzer, S. M. Belkoff, S. C. Mears, and M. Armand, "Subject-specific planning of femoroplasty: an experimental verification study," *Journal of biomechanics*, vol. 48, no. 1, pp. 59–64, 2015.
- [40] Y. Otake, M. Armand, O. Sadowsky, R. S. Armiger, M. D. Kutzer, S. C. Mears, P. Kazanzides, and R. H. Taylor, "An image-guided femoroplasty system: development and initial cadaver studies," in *Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling*, vol. 7625. International Society for Optics and Photonics, 2010, p. 76250P.

## BIBLIOGRAPHY

- [41] A. Cohen-Rosenblum and Q. Cui, “Osteonecrosis of the femoral head,” *Orthopedic Clinics of North America*, vol. 50, no. 2, pp. 139–149, 2019.
- [42] J. R. Lieberman, D. J. Berry, M. A. Montv, R. K. Aaron, J. J. Callaghan, A. Rayadhyaksha, and J. R. Urbaniak, “Osteonecrosis of the hip: management in the twenty-first century,” *JBJS*, vol. 84, no. 5, pp. 834–853, 2002.
- [43] N. F. SooHoo, S. Vyas, J. Manunga, H. Sharifi, G. Kominski, and J. R. Lieberman, “Cost-effectiveness analysis of core decompression,” *The Journal of arthroplasty*, vol. 21, no. 5, pp. 670–681, 2006.
- [44] A. J. Johnson, M. A. Mont, A. K. Tsao, and L. C. Jones, “Treatment of femoral head osteonecrosis in the united states: 16-year analysis of the nationwide inpatient sample,” *Clinical Orthopaedics and Related Research®*, vol. 472, no. 2, pp. 617–623, 2014.
- [45] T. Nishii, N. Sugano, K. Ohzono, T. Sakai, Y. Sato, and H. Yoshikawa, “Significance of lesion size and location in the prediction of collapse of osteonecrosis of the femoral head: a new three-dimensional quantification using magnetic resonance imaging,” *Journal of orthopaedic research*, vol. 20, no. 1, pp. 130–136, 2002.
- [46] J. K. Freburger, G. M. Holmes, R. P. Agans, A. M. Jackman, J. D. Darter, A. S. Wallace, L. D. Castel, W. D. Kalsbeek, and T. S. Carey, “The rising

## BIBLIOGRAPHY

- prevalence of chronic low back pain,” *Archives of internal medicine*, vol. 169, no. 3, pp. 251–258, 2009.
- [47] B. Duthey, “Background paper 6.24 low back pain,” *Priority medicines for Europe and the world. Global Burden of Disease (2010), (March)*, pp. 1–29, 2013.
- [48] I. M. Wilkinson and S. P. Cohen, “Epidural steroid injections,” *Current pain and headache reports*, vol. 16, no. 1, pp. 50–59, 2012.
- [49] V. B. Vad, A. L. Bhat, G. E. Lutz, and F. Cammisa, “Transforaminal epidural steroid injections in lumbosacral radiculopathy: a prospective randomized study,” *Spine*, vol. 27, no. 1, pp. 11–15, 2002.
- [50] J. Bui and N. Bogduk, “A systematic review of the effectiveness of ct-guided, lumbar transforaminal injection of steroids,” *Pain Medicine*, vol. 14, no. 12, pp. 1860–1865, 2013.
- [51] C. Kim, C. J. Moon, H. E. Choi, and Y. Park, “Retrodiscal approach of lumbar epidural block,” *Annals of rehabilitation medicine*, vol. 35, no. 3, p. 418, 2011.
- [52] K. T. AbdeleRahman and G. Rakocevic, “Paraplegia following lumbosacral steroid epidural injections,” *Journal of Clinical Anesthesia*, vol. 26, no. 6, pp. 497–499, 2014.

## BIBLIOGRAPHY

- [53] C. Gao, M. Unberath, R. Taylor, and M. Armand, “Localizing dexterous surgical tools in x-ray for image-based navigation,” *arXiv preprint arXiv:1901.06672*, 2019.
- [54] N. Drenkow, N. Sani, I. Shpitser, and M. Unberath, “Robustness in deep learning for computer vision: Mind the gap?” *arXiv preprint arXiv:2112.00639*, 2021.
- [55] R. Taori, A. Dave, V. Shankar, N. Carlini, B. Recht, and L. Schmidt, “Measuring robustness to natural distribution shifts in image classification,” *arXiv preprint arXiv:2007.00644*, 2020.
- [56] F. Mahmood and N. J. Durr, “Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy,” *Medical image analysis*, vol. 48, pp. 230–243, 2018.
- [57] W. Gu, C. Gao, R. Grupp, J. Fotouhi, and M. Unberath, “Extended capture range of rigid 2d/3d registration by estimating riemannian pose gradients,” in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2020, pp. 281–291.
- [58] B. Bier, F. Goldmann, J.-N. Zaech, J. Fotouhi, R. Hegeman, R. Grupp, M. Armand, G. Osgood, N. Navab, A. Maier *et al.*, “Learning to detect anatomical landmarks of the pelvis in x-rays from arbitrary views,” *In-*

## BIBLIOGRAPHY

- ternational journal of computer assisted radiology and surgery*, vol. 14, no. 9, pp. 1463–1473, 2019.
- [59] R. B. Grupp, M. Unberath, C. Gao, R. A. Hegeman, R. J. Murphy, C. P. Alexander, Y. Otake, B. A. McArthur, M. Armand, and R. H. Taylor, “Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2d/3d registration,” *International journal of computer assisted radiology and surgery*, vol. 15, no. 5, pp. 759–769, 2020.
- [60] H. Edgar, S. Daneshvari Berry, E. Moes, N. Adolphi, P. Bridges, and K. Nolte, “New mexico decedent image database.” *Office of the Medical Investigator, University of New Mexico.*, 2020, doi.org/10.25827/5s8c-n515.
- [61] M. Krčah, G. Székely, and R. Blanc, “Fully automatic and fast segmentation of the femur bone from 3d-ct images with no shape prior,” in *2011 IEEE international symposium on biomedical imaging: from nano to macro*. IEEE, 2011, pp. 2087–2090.
- [62] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, “Vision-based and marker-less surgical tool detection and tracking: a review of the literature,” *Medical image analysis*, vol. 35, pp. 633–654, 2017.
- [63] F. Alambeigi, Y. Wang, S. Sefati, C. Gao, R. J. Murphy, I. Iordachita, R. H. Taylor, H. Khanuja, and M. Armand, “A curved-drilling approach in

## BIBLIOGRAPHY

- core decompression of the femoral head osteonecrosis using a continuum manipulator,” *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1480–1487, 2017.
- [64] M. Bakhtiarinejad, F. Alambeigi, A. Chamani, M. Unberath, H. Khanuja, and M. Armand, “A biomechanical study on the use of curved drilling technique for treatment of osteonecrosis of femoral,” *Computational Biomechanics for Medicine: Personalisation, Validation and Therapy*, p. 87, 2019.
- [65] C. Gao, H. Phalen, S. Sefati, J. Ma, R. H. Taylor, M. Unberath, and M. Armand, “Fluoroscopic navigation for a surgical robotic system including a continuum manipulator,” *IEEE Transactions on Biomedical Engineering*, 2021.
- [66] M. Unberath, J.-N. Zaech, S. C. Lee, B. Bier, J. Fotouhi, M. Armand, and N. Navab, “Deepdrr—a catalyst for machine learning in fluoroscopy-guided procedures,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 98–106.
- [67] M. Unberath, J.-N. Zaech, C. Gao, B. Bier, F. Goldmann, S. C. Lee, J. Fotouhi, R. Taylor, M. Armand, and N. Navab, “Enabling machine learning in x-ray-based procedures via realistic simulation of image formation,”

## BIBLIOGRAPHY

- International journal of computer assisted radiology and surgery*, vol. 14, no. 9, pp. 1517–1528, 2019.
- [68] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [69] M. M. Haq and J. Huang, “Adversarial domain adaptation for cell segmentation,” in *Medical Imaging with Deep Learning*, 2020.
- [70] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [71] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.
- [72] J. Park, D. K. Han, and H. Ko, “Adaptive weighted multi-discriminator cyclegan for underwater image enhancement,” *Journal of Marine Science and Engineering*, vol. 7, no. 7, p. 200, 2019.

## BIBLIOGRAPHY

- [73] M. Alber, A. Buganza Tepole, W. R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W. W. Lytton, P. Perdikaris, L. Petzold *et al.*, “Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences,” *NPJ digital medicine*, vol. 2, no. 1, pp. 1–11, 2019.
- [74] M. Viceconti, A. Henney, and E. Morley-Fletcher, “In silico clinical trials: how computer simulation will transform the biomedical industry,” *International Journal of Clinical Trials*, vol. 3, no. 2, pp. 37–46, 2016.
- [75] A. Badano, C. G. Graff, A. Badal, D. Sharma, R. Zeng, F. W. Samuelson, S. J. Glick, and K. J. Myers, “Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial,” *JAMA network open*, vol. 1, no. 7, pp. e185474–e185474, 2018.
- [76] B. Bier, M. Unberath, J.-N. Zaech, J. Fotouhi, M. Armand, G. Osgood, N. Navab, and A. Maier, “X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 55–63.

## BIBLIOGRAPHY

- [77] R. B. Grupp *et al.*, “Computer-assisted fluoroscopic navigation for orthopaedic surgery,” Ph.D. dissertation, Johns Hopkins University, 2020.
- [78] A. Margalit, H. Phalen, C. Gao, J. Ma, K. V. Suresh, P. Jain, A. Farvardin, R. H. Taylor, M. Armand, A. Chattere *et al.*, “Autonomous spinal robotic system for transforaminal lumbar epidural injections: A proof of concept of study,” *Global Spine Journal*, p. 21925682221096625, 2022.
- [79] G. Chintalapani, *Statistical atlases of bone anatomy and their applications*. The Johns Hopkins University, 2010.
- [80] L. M. Ellingsen, G. Chintalapani, R. H. Taylor, and J. L. Prince, “Robust deformable image registration using prior shape information for atlas to patient registration,” *Computerized Medical Imaging and Graphics*, vol. 34, no. 1, pp. 79–90, 2010.
- [81] G. Zheng, S. Gollmer, S. Schumann, X. Dong, T. Feilkas, and M. A. G. Ballester, “A 2d/3d correspondence building method for reconstruction of a patient-specific 3d bone surface model using point distribution models and calibrated x-ray images,” *Medical image analysis*, vol. 13, no. 6, pp. 883–899, 2009.
- [82] C. M. Cyr, A. F. Kamal, T. B. Sebastian, and B. B. Kimia, “2d-3d registration based on shape matching,” in *Proceedings IEEE Workshop on Math-*

## BIBLIOGRAPHY

- ematical Methods in Biomedical Image Analysis. MMBIA-2000 (Cat. No. PR00737).* IEEE, 2000, pp. 198–203.
- [83] A. Guéziec, P. Kazanzides, B. Williamson, and R. H. Taylor, “Anatomy-based registration of ct-scan and intraoperative x-ray images for guiding a surgical robot,” *IEEE Transactions on Medical Imaging*, vol. 17, no. 5, pp. 715–728, 1998.
- [84] Y. Chen, X. Ee, W. K. Leow, and T. S. Howe, “Automatic extraction of femur contours from hip x-ray images,” in *International Workshop on Computer Vision for Biomedical Image Applications*. Springer, 2005, pp. 200–209.
- [85] W. Yu and G. Zheng, “2d-3d regularized deformable b-spline registration: Application to the proximal femur,” in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 829–832.
- [86] X. Zhang, Y. Zhu, C. Li, J. Zhao, and G. Li, “Sift algorithm-based 3d pose estimation of femur,” *Bio-medical materials and engineering*, vol. 24, no. 6, pp. 2847–2855, 2014.
- [87] A. Zaman and S. Y. Ko, “Improving the accuracy of 2d-3d registration of femur bone for bone fracture reduction robot using particle swarm optimization,” in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, 2018, pp. 101–102.

## BIBLIOGRAPHY

- [88] B. Bier, K. Aschoff, C. Syben, M. Unberath, M. Levenston, G. Gold, R. Fahrig, and A. Maier, “Detecting anatomical landmarks for motion estimation in weight-bearing imaging of knees,” in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2018, pp. 83–90.
- [89] M. Grimm, J. Esteban, M. Unberath, and N. Navab, “Pose-dependent weights and domain randomization for fully automatic x-ray to ct registration,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2221–2232, 2021.
- [90] X. Zhang, G. Zheng, F. Langlotz, and L.-P. Nolte, “Assessment of spline-based 2d–3d registration for image-guided spine surgery,” *Minimally invasive therapy & allied technologies*, vol. 15, no. 3, pp. 193–199, 2006.
- [91] T. De Silva, A. Uneri, M. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, N. Aygun, S. Lo, J. Wolinsky, and J. Siewerdsen, “3d–2d image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch,” *Physics in Medicine & Biology*, vol. 61, no. 8, p. 3009, 2016.
- [92] R. B. Grupp, M. Armand, and R. H. Taylor, “Patch-based image similarity for intraoperative 2d/3d pelvis registration during periacetabular osteotomy,” in *OR 2.0 Context-Aware Operating Theaters, Computer*

## BIBLIOGRAPHY

*Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis.* Springer, 2018, pp. 153–163.

- [93] C. Payer, D. Štern, H. Bischof, and M. Urschler, “Coarse to fine vertebrae localization and segmentation with spatialconfiguration-net and u-net,” in *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP*, vol. 5, 2020, pp. 124–133.
- [94] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [95] N. Hansen and A. Ostermeier, “Completely derandomized self-adaptation in evolution strategies,” *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.
- [96] Y. Otake, M. Armand, R. S. Armiger, M. D. Kutzer, E. Basafa, P. Kazanzides, and R. H. Taylor, “Intraoperative image-based multiview 2d/3d registration for image-guided orthopaedic surgery: incorporation of fiducial-based c-arm tracking and gpu-acceleration,” *IEEE transactions on medical imaging*, vol. 31, no. 4, pp. 948–962, 2011.
- [97] C. Gao, X. Liu, W. Gu, B. Killeen, M. Armand, R. Taylor, and M. Unberath, “Generalizing spatial transformers to projective geometry with

## BIBLIOGRAPHY

- applications to 2d/3d registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 329–339.
- [98] A. Varnavas, T. Carrell, and G. Penney, “Fully automated 2d–3d registration and verification,” *Medical image analysis*, vol. 26, no. 1, pp. 108–119, 2015.
- [99] S. Miao, R. Liao, J. Lucas, and C. Chefd’hotel, “Toward accurate and robust 2-d/3-d registration of implant models to single-plane fluoroscopy,” in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*. Springer, 2013, pp. 97–106.
- [100] S. Miao, Z. J. Wang, and R. Liao, “A cnn regression approach for real-time 2d/3d registration,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.
- [101] Y. Li, G. Wang, X. Ji, Y. Xiang, and D. Fox, “Deepim: Deep iterative matching for 6d pose estimation,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 683–698.
- [102] S. Jaganathan, J. Wang, A. Borsdorf, K. Shetty, and A. Maier, “Deep iterative 2d/3d registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 383–392.

## BIBLIOGRAPHY

- [103] S. Miao, S. Piat, P. Fischer, A. Tuysuzoglu, P. Mewes, T. Mansi, and R. Liao, “Dilated fcn for multi-agent 2d/3d medical image registration,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [104] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, “Robust non-rigid registration through agent-based action learning,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 344–352.
- [105] R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D. Comaniciu, “An artificial agent for robust image registration,” in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [106] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier, “Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems,” *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1454–1463, 2018.
- [107] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, “Spatial transformer networks,” in *Advances in neural information processing systems*, 2015, pp. 2017–2025.
- [108] D. Kuang and T. Schmah, “Faim—a convnet method for unsupervised

## BIBLIOGRAPHY

- 3d medical image registration,” in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 646–654.
- [109] E. Ferrante, O. Oktay, B. Glocker, and D. H. Milone, “On the adaptability of unsupervised cnn-based deformable image registration to unseen image domains,” in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 294–302.
- [110] X. Yan, J. Yang, E. Yumer, Y. Guo, and H. Lee, “Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision,” in *Advances in neural information processing systems*, 2016, pp. 1696–1704.
- [111] S. S. M. Salehi, S. Khan, D. Erdoganmus, and A. Gholipour, “Real-time deep registration with geodesic loss,” *arXiv preprint arXiv:1803.05982*, 2018.
- [112] S. Mahendran, H. Ali, and R. Vidal, “3d pose regression using convolutional neural networks,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2174–2182.
- [113] N. Miolane, J. Mathe, C. Donnat, M. Jorda, and X. Pennec, “geomstats: a python package for riemannian geometry in machine learning,” *arXiv preprint arXiv:1805.08308*, 2018.
- [114] C.-F. R. Chen, Q. Fan, and R. Panda, “Crossvit: Cross-attention multi-

## BIBLIOGRAPHY

- scale vision transformer for image classification,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 357–366.
- [115] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.
- [116] C. Gao, R. B. Grupp, M. Unberath, R. H. Taylor, and M. Armand, “Fiducial-free 2d/3d registration of the proximal femur for robot-assisted femoroplasty,” in *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11315. International Society for Optics and Photonics, 2020, p. 113151C.
- [117] C. Payer, D. Stern, H. Bischof, and M. Urschler, “Coarse to fine vertebrae localization and segmentation with spatialconfiguration-net and u-net.” in *VISIGRAPP (5: VISAPP)*, 2020, pp. 124–133.
- [118] R. Grupp, M. Unberath, C. Gao, R. Hegeman, R. Murphy, C. Alexander, Y. Otake, B. McArthur, M. Armand, and R. Taylor, “Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2d/3d registration,” *arXiv preprint arXiv:1911.07042*, 2019.
- [119] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and

## BIBLIOGRAPHY

- K. Keutzer, “Densenet: Implementing efficient convnet descriptor pyramids,” *arXiv preprint arXiv:1404.1869*, 2014.
- [120] L. N. Smith, “Cyclical learning rates for training neural networks,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 464–472.
- [121] M. Bakhtiarinejad, C. Gao, A. Farvardin, G. Zhu, Y. Wang, J. K. Oni, R. H. Taylor, and M. Armand, “A surgical robotic system for osteoporotic hip augmentation: An early feasibility study,” *IEEE Transactions on Medical Robotics and Bionics*.
- [122] J. Burgner-Kahrs, D. C. Rucker, and H. Choset, “Continuum robots for medical applications: A survey,” *IEEE Transactions on Robotics*, vol. 31, no. 6, pp. 1261–1280, 2015.
- [123] A. H. Gosline, N. V. Vasilyev, E. J. Butler, C. Folk, A. Cohen, R. Chen, N. Lang, P. J. Del Nido, and P. E. Dupont, “Percutaneous intracardiac beating-heart surgery using metal mems tissue approximation tools,” *The International journal of robotics research*, vol. 31, no. 9, pp. 1081–1093, 2012.
- [124] G. Dwyer, F. Chadebecq, M. T. Amo, C. Bergeles, E. Maneas, V. Pawar, E. Vander Poorten, J. Deprest, S. Ourselin, P. De Coppi *et al.*, “A continuum robot and control interface for surgical assist in fetoscopic interven-

## BIBLIOGRAPHY

- tions,” *IEEE robotics and automation letters*, vol. 2, no. 3, pp. 1656–1663, 2017.
- [125] H. Wang, X. Wang, W. Yang, and Z. Du, “Design and kinematic modeling of a notch continuum manipulator for laryngeal surgery,” *INTERNATIONAL JOURNAL OF CONTROL AUTOMATION AND SYSTEMS*, 2020.
- [126] F. Campisano, A. A. Remirez, C. A. Landewe, S. Caló, K. L. Obstein, R. J. Webster III, and P. Valdastri, “Teleoperation and contact detection of a waterjet-actuated soft continuum manipulator for low-cost gastroscopy,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6427–6434, 2020.
- [127] N. Simaan, R. M. Yasin, and L. Wang, “Medical technologies and challenges of robot-assisted minimally invasive intervention and diagnostics,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 465–490, 2018.
- [128] M. Hwang and D.-S. Kwon, “Strong continuum manipulator for flexible endoscopic surgery,” *IEEE /ASME Transactions on Mechatronics*, vol. 24, no. 5, pp. 2193–2203, 2019.
- [129] X. Ma, C. Song, P. W. Chiu, and Z. Li, “Autonomous flexible endoscope for minimally invasive surgery with enhanced safety,” *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2607–2613, 2019.

## BIBLIOGRAPHY

- [130] M. Bakhtiarinejad, F. Alambeigi, A. Chamani, M. Unberath, H. Khanuja, and M. Armand, “A biomechanical study on the use of curved drilling technique for treatment of osteonecrosis of femoral head,” in *Computational Biomechanics for Medicine*. Springer, 2020, pp. 87–97.
- [131] Y. Otake, R. J. Murphy, M. Kutzer, R. H. Taylor, and M. Armand, “Piecewise-rigid 2d-3d registration for pose estimation of snake-like manipulator using an intraoperative x-ray projection,” in *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9036. International Society for Optics and Photonics, 2014, p. 90360Q.
- [132] M. D. Kutzer, S. M. Segreti, C. Y. Brown, M. Armand, R. H. Taylor, and S. C. Mears, “Design of a new cable-driven manipulator with a large open lumen: Preliminary applications in the minimally-invasive removal of osteolysis,” in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 2913–2920.
- [133] R. J. Murphy, M. D. Kutzer, S. M. Segreti, B. C. Lucas, and M. Armand, “Design and kinematic characterization of a surgical manipulator with a focus on treating osteolysis,” *Robotica*, vol. 32, no. 6, p. 835, 2014.
- [134] M. Armand, M. D. Kutzer, C. Y. Brown, R. H. Taylor, and E. Basafa,

## BIBLIOGRAPHY

- “Cable-driven morphable manipulator,” Aug. 22 2017, uS Patent 9,737,687.
- [135] S. Sefati, F. Alambeigi, I. Iordachita, M. Armand, and R. J. Murphy, “Fbg-based large deflection shape sensing of a continuum manipulator: Manufacturing optimization,” in *2016 IEEE SENSORS*. IEEE, 2016, pp. 1–3.
- [136] S. Sefati, C. Gao, I. Iordachita, R. H. Taylor, and M. Armand, “Data-driven shape sensing of a surgical continuum manipulator using an uncalibrated fiber bragg grating sensor,” *IEEE Sensors Journal*, 2020.
- [137] I. Laina, N. Rieke, C. Rupprecht, J. P. Vizcaíno, A. Eslami, F. Tombari, and N. Navab, “Concurrent segmentation and localization for tracking of surgical instruments,” in *International conference on medical image computing and computer-assisted intervention*. Springer, 2017, pp. 664–672.
- [138] M. Unberath, O. Taubmann, M. Hell, S. Achenbach, and A. Maier, “Symmetry, outliers, and geodesics in coronary artery centerline reconstruction from rotational angiography,” *Medical physics*, vol. 44, no. 11, pp. 5672–5685, 2017.
- [139] N. Kuo, J. Lee, A. Deguet, D. Song, E. C. Burdette, and J. Prince, “Automatic segmentation of seeds and fluoroscope tracking (ftrac) fiducial in

## BIBLIOGRAPHY

- prostate brachytherapy x-ray images,” in *Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling*, vol. 7625. International Society for Optics and Photonics, 2010, p. 76252T.
- [140] A. Farvardin and M. Armand, “Osteoporotic bone augmentation utilizing curved pattern of pmma injection: A combined finite element and optimization investigation,” in *ASME International Mechanical Engineering Congress and Exposition*, vol. 59407. American Society of Mechanical Engineers, 2019, p. V003T04A043.
- [141] S. Sefati, R. Hegeman, F. Alambeigi, I. Iordachita, and M. Armand, “Fbg-based position estimation of highly deformable continuum manipulators: Model-dependent vs. data-driven approaches,” in *2019 International Symposium on Medical Robotics (ISMR)*. IEEE, 2019, pp. 1–6.
- [142] F. Alambeigi, S. A. Pedram, J. L. Speyer, J. Rosen, I. Iordachita, R. H. Taylor, and M. Armand, “Scade: Simultaneous sensor calibration and deformation estimation of fbg-equipped unmodeled continuum manipulators,” *IEEE Transactions on Robotics*, vol. 36, no. 1, pp. 222–239, 2019.
- [143] G. Li, N. A. Patel, W. Liu, D. Wu, K. Sharma, K. Cleary, J. Fritz, and I. Iordachita, “A fully actuated body-mounted robotic assistant for mri-guided low back pain injection,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 5495–5501.

## BIBLIOGRAPHY

- [144] G. Li, N. A. Patel, J. Hagemeister, J. Yan, D. Wu, K. Sharma, K. Cleary, and I. Iordachita, “Body-mounted robotic assistant for mri-guided low back pain injection,” *International journal of computer assisted radiology and surgery*, vol. 15, no. 2, pp. 321–331, 2020.
- [145] R. Monfaredi, K. Cleary, and K. Sharma, “Mri robots for needle-based interventions: systems and technology,” *Annals of biomedical engineering*, vol. 46, no. 10, pp. 1479–1497, 2018.
- [146] A. Squires, J. N. Oshinski, N. M. Boulis, and Z. T. H. Tse, “Spinobot: an mri-guided needle positioning system for spinal cellular therapeutics,” *Annals of biomedical engineering*, vol. 46, no. 3, pp. 475–487, 2018.
- [147] J. Esteban, W. Simson, S. R. Witzig, A. Rienmüller, S. Virga, B. Frisch, O. Zettinig, D. Sakara, Y.-M. Ryang, N. Navab *et al.*, “Robotic ultrasound-guided facet joint insertion,” *International journal of computer assisted radiology and surgery*, vol. 13, no. 6, pp. 895–904, 2018.
- [148] M. Tirindelli, M. Victorova, J. Esteban, S. T. Kim, D. Navarro-Alarcon, Y. P. Zheng, and N. Navab, “Force-ultrasound fusion: Bringing spine robotic-us to the next “level”,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5661–5668, 2020.
- [149] S. Schafer, S. Nithiananthan, D. Mirota, A. Uneri, J. Stayman, W. Zbijewski, C. Schmidgunst, G. Kleinszig, A. Khanna, and J. Siewerdsen,

## BIBLIOGRAPHY

- “Mobile c-arm cone-beam ct for guidance of spine surgery: Image quality, radiation dose, and integration with interventional guidance,” *Medical physics*, vol. 38, no. 8, pp. 4563–4574, 2011.
- [150] S. Onogi, K. Morimoto, I. Sakuma, Y. Nakajima, T. Koyama, N. Sugano, Y. Tamura, S. Yonenobu, and Y. Momoi, “Development of the needle insertion robot for percutaneous vertebroplasty,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2005, pp. 105–113.
- [151] Z. Han, K. Yu, L. Hu, W. Li, H. Yang, M. Gan, N. Guo, B. Yang, H. Liu, and Y. Wang, “A targeting method for robot-assisted percutaneous needle placement under fluoroscopy guidance,” *Computer Assisted Surgery*, vol. 24, no. sup1, pp. 44–52, 2019.
- [152] G. Burström, M. Balicki, A. Patriciu, S. Kyne, A. Popovic, R. Holthuizen, R. Homan, H. Skulason, O. Persson, E. Edström *et al.*, “Feasibility and accuracy of a robotic guidance system for navigated spine surgery in a hybrid operating room: a cadaver study,” *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020.
- [153] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, “3d slicer

## BIBLIOGRAPHY

- as an image computing platform for the quantitative imaging network,” *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.
- [154] J. W. Park, H. S. Nam, S. K. Cho, H. J. Jung, B. J. Lee, and Y. Park, “Kambin’s triangle approach of lumbar transforaminal epidural injection with spinal stenosis,” *Annals of rehabilitation medicine*, vol. 35, no. 6, p. 833, 2011.
- [155] Z. A. Smith, K. Sugimoto, C. D. Lawton, and R. G. Fessler, “Incidence of lumbar spine pedicle breach after percutaneous screw fixation: a radiographic evaluation of 601 screws in 151 patients,” *Clinical Spine Surgery*, vol. 27, no. 7, pp. 358–363, 2014.
- [156] G. Li, N. A. Patel, A. Melzer, K. Sharma, I. Iordachita, and K. Cleary, “Mri-guided lumbar spinal injections with body-mounted robotic system: cadaver studies,” *Minimally Invasive Therapy & Allied Technologies*, pp. 1–9, 2020.
- [157] H. Li, Y. Wang, Y. Li, and J. Zhang, “A novel manipulator with needle insertion forces feedback for robot-assisted lumbar puncture,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 17, no. 2, p. e2226, 2021.
- [158] L. Jud, J. Fotouhi, O. Andronic, A. Aichmair, G. Osgood, N. Navab, and M. Farshad, “Applicability of augmented reality in orthopedic surgery—

## BIBLIOGRAPHY

- a systematic review,” *BMC musculoskeletal disorders*, vol. 21, no. 1, pp. 1–13, 2020.
- [159] J. T. Verhey, J. M. Haglin, E. M. Verhey, and D. E. Hartigan, “Virtual, augmented, and mixed reality applications in orthopedic surgery,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 16, no. 2, p. e2067, 2020.
- [160] E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

# Vita



Cong Gao received his bachelor's degree in Electronic Engineering from Tsinghua University in 2017 and a Master's degree in Computer Science from Johns Hopkins University in 2021. He enrolled in the Computer Science PhD program at Johns Hopkins University in 2017, and worked closely with Drs. Mehran Armand, Mathias Unberath and Russell Taylor in the Laboratory of Computational Sensing and Robotics (LCSR). He worked as an Applied Research Intern with Intuitive Surgical Inc. in 2021, leading a project to improve spatial sensing of the da Vinci Robot.