Evaluating the influence of machine-specific DRR parameters on the accuracy of X-ray simulation and orthopaedic 2D-3D reconstruction

Cornelius Johannes Frederik Reyneke

Supervisor: Assoc. Prof. T. E. Mutsvangwa

Co-supervisors: Prof. T. S. Douglas
Prof. T. Vetter (Univ. Basel)

Department of Human Biology
Division of Biomedical Engineering

This dissertation is submitted for the degree of
Doctor of Philosophy

South Africa February 2023

To my wife Stefanie and my daughter Livia

Declaration

I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.

Signed by candidate

Cornelius Johannes Frederik Reyneke

February 2023

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s):

1. Reyneke CJF, Luthi M, Burdin V, Douglas TS, Vetter T, Mutsvangwa TEM. Review of 2-D/3-D Reconstruction Using Statistical Shape and Intensity Models and X-Ray Image Synthesis: Toward a Unified Framework. IEEE Rev Biomed Eng. 2019;12:269-286.

Signature: _____ Date: 16.02.2023

Student Name: Cornelius Johannes Frederik Reyneke Student Number: RYNCOR001

Acknowledgements

First, I would like to thank my supervisor, Assoc. Prof. Tinashe Mutsvangwa for diligently guiding me through my PhD and always motivating me to improve the quality of my work.
I would like to acknowledge my co-supervisor, Prof. Tania Douglas. I will always remember her calm disposition, her wisdom, and her keen eye for grammar.
I would also like to acknowledge Prof. Thomas Vetter of the University of Basel for his constructive criticism, expert advice, and for making me feel very welcome, both in his group and in Basel, which I now also call my home.
I am grateful to Prof. Valérie Burdin at IMT Atlantique for helping me make sense of the X-ray imaging protocols and mathematics I encountered during my PhD. I also thank her for graciously hosting me during my time in Brest.
I would like to acknowledge the South African Research Chair (NRF-SARChi) in Biomedical Engineering and Innovation and the Swiss National Science Foundation (SNSF) for the financial support they provided.
I would also like to acknowledge Dr. Marcel Lüthi, Dr. Andreas Forster, Dr. Thomas Gerig, and Ghazi Bouabene for their time and patience while teaching me Scala, helping me hunt down bugs in my code, and answering my many questions during the time I spent at the Graphics and Vision Research Group (GraVis).
I am grateful to Prof. Thomas Ertl for hosting me during my time at the University of Stuttgart Visualization Research Center (VISUS), and to Valentin Bruder for making me feel welcome, and helping me to create my own volume renderer.
Thank you to Dr. Sebastian Keller for his friendship, support, and valuable advice on the organisation of my work.
Also, sincere thanks to my colleagues and staff at the UCT Division of Biomedical Engineering, especially Xolisile Thusini, Dr. Jean-Raissaire Fouefack, Yvonne Karanja, and Hervé Nbonsou, for their friendly advice and assistance during the time we spent together in the Medical Image Inferencing & Distributed Diagnostics (Mi2D2) group.
Thank you to my parents, Cor and Aléta, and my friends, David, Stephen, Liam, and Chris, for their continued encouragement along the way.
Finally, and most of all, I would like to thank my wife, Stefanie, and my daughter, Livia, for all their love, patience, encouragement, and support, without which this work would never have been completed.

Abstract

In orthopaedics, two-dimensional-to-three-dimensional (2D-3D) reconstruction allows 3D bone structures, conventionally derived from 3D modalities such as computed tomography (CT), to be derived from 2D modalities such as X-ray imaging. Thus, clinical interventions such as implant design and postoperative evaluation can be made more accessible, less expensive and, in some cases, the dose of ionising radiation to the patient can be reduced. State-of-the-art approaches iteratively warp a deformable 3D model of the bone, simulate an X-ray projection image from it, and compare the result to a real X-ray target image, with the goal of minimising the disparity between the two. The X-ray simulation method includes implicit X-ray machine-specific calibration settings, which affect both the resulting geometry and intensity profile of simulated X-ray images. However, the importance of correct projection calibration is not adequately discussed in literature, despite the fact that X-ray machines vary significantly with regard to their imaging setup. Furthermore, the reconstruction inaccuracies resulting from projection miscalibration are not well understood.

In this thesis, first, an extensive literature review of the 2D/3D reconstruction problem is conducted and a unified mathematical formulation is proposed. Next, the development of a digitally reconstructed radiograph (DRR) renderer for simulating X-ray images is reported. The DRR renderer was developed using a standard volume rendering framework, adapted to the unique requirements of X-ray imaging. The renderer can be calibrated to machine-specific parameters, of which two were the focus in-depth experimental investigations: the distance from the X-ray source to the imaged object (S2O distance), and modelled energy intensity of the X-ray source (incident energy). For each correctly calibrated parameter of a nominal value, a set of perturbed values were generated and used to render corresponding simulated X-ray images. The resulting rendering errors were then measured through comparison to a ground-truth X-ray image of an object with known geometry and density. The results demonstrated a 2D Dice overlap error of 0.4% for every 1 cm with which the S2O distance was offset, and a pixel intensity error of 1% for every 0.1 point offset of the incident energy coefficient. In a subsequent experimental investigation, the same two machine-specific parameters were studied to determine the influence of machine-specific parameters on 2D-3D reconstruction accuracy using a single image. First, having generated a set of random DRRs, each was

used as a simulated registration target while offsetting the correctly calibrated parameters in the same fashion as for the first experiment. A 2D Dice overlap error of at least 0.3% was estimated for every 1 cm with which the S2O distance was offset. Similar experiments were performed on real X-ray images of dry-bone femurs as registration targets, leading to a 2D Dice overlap error of at least 0.4% for every 1 cm with which the S2O distance was offset. The incident energy parameter had no discernible impact on the resulting registration accuracy for either the simulated or real-world targets. As a whole, the experimental results demonstrate that machine-specific calibration has a noticeable impact on the accuracy of simulated X-ray images and, in the case of the S2O distance, also on the accuracy of 2D-3D reconstruction. An in-depth commentary on the work done, together with suggestions for future research, concludes the thesis. In summary, the presented work provides valuable insight into previously overlooked facets of orthopaedic 2D-3D reconstruction, and suggests that machine-specific calibration should be carefully considered when performing 2D-3D reconstruction.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Volumetric models of patient-specific bone structures enable orthopaedic interventions such as surgical planning and implant design. Such models are typically derived from three-dimensional (3D) imaging modalities, including computed tomography (CT) and magnetic resonance imaging (MRI). However, such 3D imaging technologies are expensive, have long image acquisition times, and are inaccessible to hospitals in resource–limited and remote settings [181, 28, 193, 78, 136]. The goal of orthopaedic two-dimensional-to-three-dimensional (2D-3D) reconstruction is to alleviate these problems through reliably reproducing three-dimensional (3D) bone structures from two-dimensional (2D) modalities such as X-ray imaging. While patient–specific reconstructions obtained in this manner are generally less accurate representations of actual patient 3D morphology, doing so significantly reduces both the imaging cost and - in the case of CT - the dose of ionising radiation a patient receives [136, 9, 105].

Researchers have framed the task of 2D-3D reconstruction as an inverse problem, where the parameters of a model of the 3D structure are inferred from 2D observations on the target X-ray image. It is also an ill-posed problem, since more than one 3D appearance can account for the same 2D projection. Moreover, the projected geometric structures are distorted in the case of perspective or fan-beam projection, as illustrated in Figure 1.1. Additional complexity is introduced by the fact that other structures such as soft tissues and metallic implants may be present in an X-ray image, introducing noise and imaging artefacts, while bone pathologies are difficult to model.

The most common method for 2D-3D reconstruction is to continuously project a deformable volumetric model into 2D space as a simulated X-ray image, while the model parameters are iteratively adjusted. Then, once the simulated X-ray image - known as a

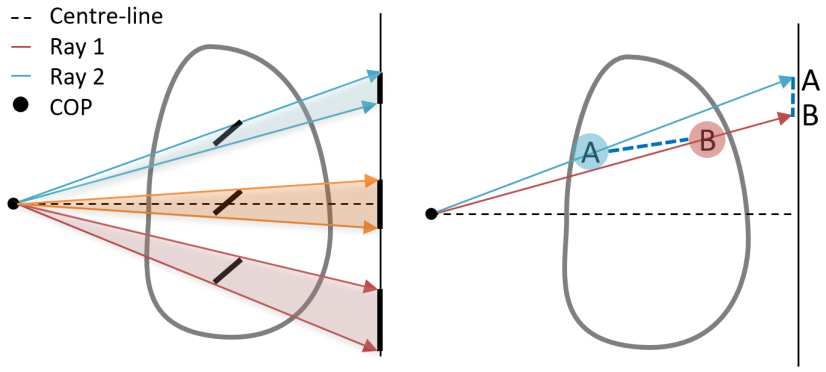


Figure 1.1 An illustration of depth–dependent distortions which are apparent in a perspective projection scheme. Tilted features of the same size, while situated at the same distance from the X–ray source, appear to be smaller or larger than one another (shown on the left). Furthermore, the relative distance between two features is distorted in proportion to their distance from the centre–line/ray (shown on the right) [98]. Fan-beam projection can be considered as a special case, having perspective projection in one direction, and parallel projection in the other.

digitally reconstructed radiograph (DRR) - is sufficiently similar to the target X-ray image, the model configuration from which it is projected is accepted as the patient-specific bone structure (see Figure 1.2).

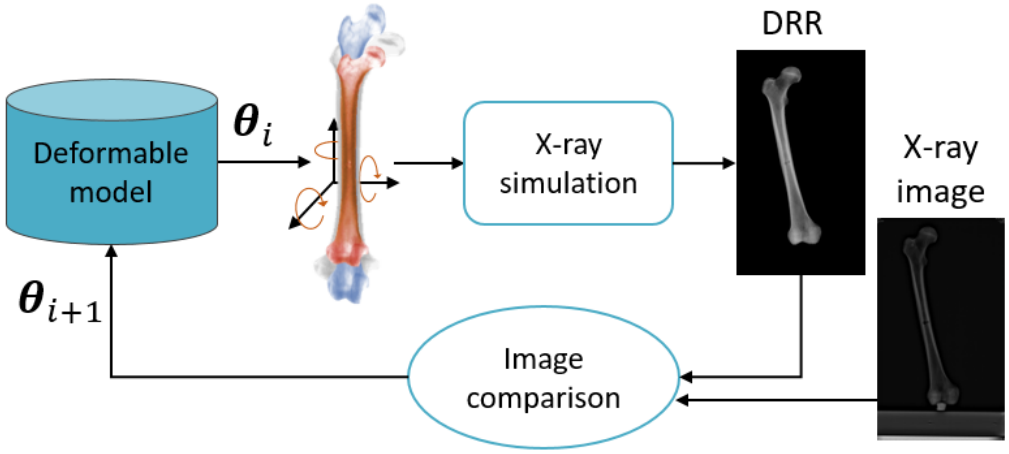


Figure 1.2 An overview of the DRR-based approach to 2D-3D reconstruction, where the set of deformable model parameters is shown as θ_i for the current iteration, and θ_{i+1} for the subsequent iteration.

The DRR-based approach, however, has two limitations. First, the algorithm suffers from a computational bottleneck because a few thousand DRRs need to be rendered in order

for the parameter space to be sufficiently explored such that an accurate patient-specific reconstruction can be obtained. The time that this process typically requires is impractical for clinicians. Researchers have addressed this limitation by parallelising the algorithm in order to leverage performance increase afforded by graphics processing units (GPUs). However, as will be explained in this thesis, the computational bottleneck now resides with the model deformation step. Second, the DRR of the initial model configuration must already be somewhat similar to the patient's X-ray image in order for the set of optimal parameters to be reliably determined without converging on a local optima (see Figure 1.3) [107, 106, 101]. A pose initialisation step such as landmark annotation is typically introduced to overcome this limitation [107, 15, 166].

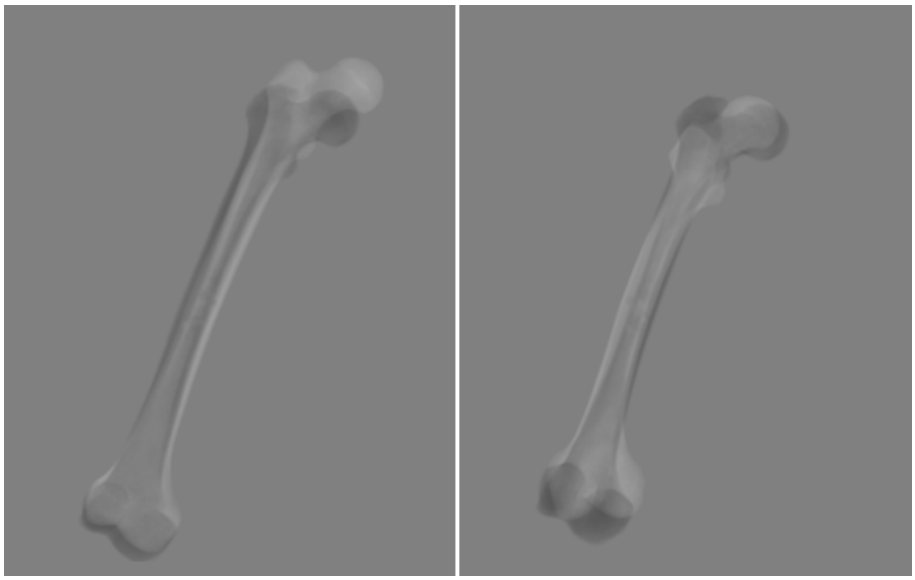


Figure 1.3 When the initial configuration of the deformable model is too dissimilar from the target X-ray image, the parameter search strategy can get caught in local minima, resulting in an insufficiently accurate patient-specific reconstruction.

Finally, the DRR rendering method includes implicit X-ray machine-specific calibration settings, which affect both the resulting geometry and intensity profile of the resulting simulated X-ray image. However, despite the fact that X-ray machines vary significantly with regard to their imaging setup[1], the importance of correct projection calibration is not discussed in literature. Furthermore, the extent to which projection mis-calibrations influence 2D-3D reconstruction accuracy is not well understood. Figure 1.4 demonstrates

[1]Parameters such as the distance from the X-ray source to the photon detector vary from one X-ray machine to another. Moreover, some X-ray machines have unusual ray geometry, which produce significantly different projections than the conventional "point-source" setup.

the discrepancies that may result from incorrect machine-specific calibration with respect to both DRR fidelity and 2D-3D reconstruction accuracy.

(a) 2D-3D alignment: incorrectly calibrated

(b) 2D-3D alignment: correctly calibrated

(c) 2D-3D reconstruction: incorrectly calibrated

(d) 2D-3D reconstruction: correctly calibrated

Figure 1.4 Two sets of images are shown, where each image represents the difference between a registration target and its estimation. Errors in machine-specific calibration - in this case the source-to-object distance parameter - may result in inaccurate DRRs (top), but also negatively influence the accuracy of volumes estimated using a 2D-3D reconstruction algorithm (bottom).

1.2 Objectives

The aim of this work was to develop a computationally efficient 2D-3D reconstruction algorithm for estimating patient-specific bone CT information using a single X-ray image, and to investigate the extent to which machine-specific calibration influences both DRR fidelity and 2D-3D reconstruction accuracy. In pursuit of these aims, the following objectives were defined:

- The first objective was to develop a X-ray simulation method which can be configured to the characteristics of a given X-ray machine (Chapter 3). In doing so, the target X-ray image of the algorithm can be approximated with greater accuracy with regard to both geometric and radiodensity information.
- The second objective was to investigate the extent to which machine-specific calibration influences DRR fidelity (Chapter 4).
- The third objective was to construct a deformable model of CT information for a specific bone structure, and reduce the time it takes to obtain a model deformation, thus easing the computational bottleneck of the 2D-3D reconstruction algorithm (Chapter 5).
- The fourth objective was to implement a parameter search strategy that is less susceptible to local optima, in order to avoid a manual pose-initialisation step, and thus further reduce the time it takes to obtain a patient-specific reconstruction (Chapter 5).
- The fifth objective was to parallelise aspects of the algorithm responsible for the model deformation step and parameter search strategy in order to further reduce the time it takes to obtain a patient-specific reconstruction using a hardware-based implementation (Chapter 5).
- The sixth objective was to investigate the extent to which machine-specific calibration influences 2D-3D reconstruction accuracy (Chapter 6).

1.3 Methods overview

A volume rendering approach, specifically "direct volume ray-casting" [170], was used for the task of X-ray simulation. This approach provided a framework for making the method adaptive to machine-specific parameters, thus making more accurate simulations possible.

A number of example CT volumes were used to to train a deformable model of a femur, using a voxel-based extension to "Gaussian Process Morphable Models" [95]. Gaussian processes make it possible to train a deformable model when constrained by computational resources, or the number of example volumes in the training set. Also, per-pixel parallelisation and GPU-based deformation processing enabled a considerable reduction in algorithm computation time.

A Markov Chain Monte Carlo (MCMC) approach [138] with random sampling was implemented as the parameter search strategy. The use of random sampling rendered the search strategy less susceptible to local optima, as is the case with conventional gradient-based approaches. By varying the magnitude of the random parameter updates, a broader range of parameters could be explored, while the MCMC approach allowed a joint search for pose and model parameters, thus avoided the need for a manual pose initialisation strategy.

Finally, the algorithm was validated with regard to reconstruction accuracy and computation time, using a dataset of corresponding pairs of X-ray-images and CT-volumes.

1.4 Contributions

The thesis has the following key contributions:

- An extensive literature review of the field of orthopaedic 2D/3D reconstruction was completed. Furthermore, a common conceptual framework and unified mathematical formulation of the 2D/3D reconstruction problem were proposed. These contributions were published in a leading journal (Impact Factor: 7, Number of citations: 40, at the time of submission) [130].

- A specialised volume renderer was developed for the simulation of X-ray images. The render can be automatically calibrated to machine-specific parameters, resulting in an increased accuracy of X-ray image simulations, as well as more accurate 2D-3D reconstructions.

- Gaussian Process Morphable Models were adapted for CT-based volumetric bone modeling, and implemented on the GPU, thus easing the algorithm's computational bottleneck, and reducing the time it takes to obtain a patient-specific reconstruction. This work was included in an IEEE conference publication [129].

- A novel application of MCMC with random sampling was made for the task of determining the patient-specific set of model parameters.

- The specialised renderer and 2D-3D reconstruction algorithm were both validated using real-world data. Moreover, a first investigation was made into the extent to which machine-specific calibration influences DRR fidelity - as well as 2D-3D reconstruction accuracy. Such an investigation using real-world data has not previously been reported in the literature.

1.5 Structure of dissertation

Chapter 2 presents an extensive literature review in the field of orthopaedic 2D-3D reconstruction. The chapter was published as a research paper in a peer-reviewed journal and has been reproduced verbatim.

Chapter 3 presents a volume rendering approach to X-ray simulation and the development of a configurable DRR renderer based on this approach (Objective 1).

Chapter 4 details the validation of the DRR renderer, and an experimental investigation into the influence of machine-specific parameters on DRR fidelity (Objective 2).

Chapter 5 presents a novel implementation of the 2D-3D reconstruction algorithm, including the development of a voxel-based deformable femur model (Objective 3), and a Bayesian inferencing parameter search strategy (Objective 4), as well as the parallelisation of the algorithm (Objective 5).

Chapter 6 describes the validation of the deformable femur model and an experimental investigation into the influence of machine-specific parameters on 2D-3D reconstruction accuracy (Objective 6).

Chapter 7 provides a general discussion of the work done as well as concluding remarks.

Chapter 2

Literature Review

2.1 Preface

This chapter reproduces a research paper, entitled “Review of 2D/3D reconstruction using statistical shape and intensity models and X-ray image synthesis: towards a unified framework”, which was published in 2019 in the "IEEE Reviews in Biomedical Engineering" journal [130]. The paper served as a vehicle with which to structure a literature review in the field of orthopaedic 2D-3D reconstruction. While surveying and collating the existing literature it became apparent that there was an increasing need to to consolidate the differing terminologies used by researchers to describe the same underlying mathematical processes. In doing so, a common conceptual framework and set of terminology was established, as well as a unified mathematical formulation of the problem.

2.2 Abstract

Patient–specific three–dimensional (3D) bone models are useful for a number of clinical applications such as surgery planning, postoperative evaluation as well as implant and prosthesis design. Two–dimensional–to–three–dimensional (2D/3D) reconstruction, also known as model–to–modality or atlas–based 2D/3D registration, provides a means of obtaining a 3D model of a patient’s bone(s) from their 2D radiographs, when 3D imaging modalities are not available. The preferred approach to estimating both shape and density information (that would be present in a patient’s CT data) for 2D/3D reconstruction makes use of digitally reconstructed radiographs and deformable models in an iterative, non–rigid, intensity–based approach. Based on a large number of state–of–the–art 2D/3D bone reconstruction methods, a unified mathematical formulation of the problem is proposed in a common conceptual

framework, using unambiguous terminology. In addition, shortcomings, recent adaptations and persisting challenges are discussed along with insights for future research.

2.3 Introduction

Three–dimensional (3D) reconstructions of patient–specific anatomical structures help medical professionals to better visualize and interact with the volumetric data from 3D imaging modalities such as computed tomography (CT) or magnetic resonance images (MRI) [136]. They have become an especially valuable tool for orthopaedic applications which include the detection of bone–related pathological deformations as well as the quantitative measurement of bone geometry and bone density for surgical planning, implant design, and postoperative evaluations [33, 136]. However, CT and MRI technologies are expensive and have long image acquisition times [181, 28, 193, 78]. Furthermore, hospitals in resource–limited and remote settings often do not have access to these technologies. Those that do, have to consider the artefacts and distortions caused by the presence of metallic implants, as well as the prohibitive costs per scan, for either modality [136]. In order to overcome these challenges, researchers have investigated ways to obtain 3D models of patient–specific anatomical structures from two–dimensional (2D) imaging modalities such as X–ray, dual–energy X–ray (DXA), fluoroscopic and ultrasound images. This approach is known as two–dimensional–to–three–dimensional (2D/3D) reconstruction [101]. Patient–specific models obtained in this way are less accurate with regard to shape and appearance than those obtained from CT or MRI [52]. However, imaging costs and the dose of ionizing radiation to a patient (in the case of CT) is significantly reduced [136, 9, 105]. For example, in the case of preoperative imaging for total hip arthroplasty, the dose is 30% less for conventional 2D X–ray images than for a CT scan [181]. Low–dose imaging systems can further reduce a patient's exposure to ionizing radiation and, in some cases, allow patients to be scanned while they are standing up, enabling the assessment of both their load–bearing posture and musculoskeletal interactions [178]. These advantages make repetitive follow-up imaging and paediatric imaging applications possible [54]. Furthermore, when obtaining 3D patient–specific models from fluoroscopic sequences, in vivo motion tracking is made possible for applications such as 3D knee kinematics studies [69, 9, 142, 163].

Two–dimensional–to–three–dimensional reconstruction is accomplished using a deformable model, which encodes prior knowledge and assumptions about the typical 3D appearance of an anatomical structure [101]. These models can then be manipulated, using a set of parameters, to match the information gathered from a patient's 2D image(s) in a 2D/3D registration [23, 7, 49, 84, 12, 180, 86, 125, 157, 109, 184, 82, 58, 135, 87, 45, 18, 67, 150,

72, 190, 83, 65, 33, 2, 183, 40, 186, 189, 192, 187, 20, 174, 141, 161, 9, 28, 162, 188, 173, 47, 74, 160, 163, 6, 78, 181, 73, 27, 191]. An exemplary specimen (sometimes referred to as a template or atlas) is used to identify a set of characteristic features such as landmarks or contours. The mathematical relationships between these features and their variations are then defined either analytically [23, 7, 84, 86, 125, 109, 184, 58, 87, 45, 183, 40, 141, 74], or statistically [49, 12, 180, 157, 82, 135, 18, 67, 150, 72, 190, 83, 65, 33, 2, 186, 189, 9, 192, 187, 20, 174, 161, 28, 162, 47, 188, 173, 160, 163, 6, 78, 181, 73, 27, 191]. Analytical models (sometimes referred to as free–form models) rely on intuitions regarding the mathematical relationships, such as those of medical experts. Statistical models, on the other hand, are learnt from a set of labelled examples. They are particularly well–suited to the rigid variability present in bony structures and are popular for their ability to represent objects robustly; while such a model is deformed its validity as a representation of the bone–of–interest is preserved (Figure 5.1 illustrates this concept)[95, 136, 186]. A more detailed explanation of statistical models such as these is provided in Section 2.4.1. For a detailed discussion about deformable models, and their use in medical image registration, the reader is referred to [145]. In order to obtain a patient–specific reconstruction, the parameters of the deformable model are tuned according to information inferred from a patient's X–ray images, the target values of which are determined using one of two registration strategies [136].

The first strategy seeks to establish correspondence, and minimize the geometric distance, between features detected on both the deformable model and the patient's X–ray image(s). This is accomplished using a Kriging optimization (also known as a Gaussian process regression) [23, 7, 84, 12, 86, 125, 109, 184, 58, 87, 45, 18, 72, 190, 83, 66, 183, 40, 186, 189, 9, 192, 20, 141, 161, 28, 162, 188, 74]. It is however first necessary to identify these corresponding features. Methods to do so vary from fully manual to fully automatic, but both are time–consuming and prone to error [186, 28, 187]. The corresponding features can either be projected onto the 2D space of the X–ray images or back–projected into the 3D space of the model. Methods of the former category search for the parameters in an iterative scheme which is terminated once the distance value is less than a pre-defined threshold [190, 10, 157, 12, 192, 49, 18]. Methods belonging to the latter category compute the parameters directly, but often do not provide sufficiently accurate reconstructions [74] (Accuracy requirements are discussed in more detail in Section 2.4.1).

The second strategy makes use of simulated X–ray images obtained from the deformable model, and seeks to maximize the score of a similarity measure which performs a comparison of pixel intensity values in the 2D space of the X–ray images [49, 180, 157, 82, 135, 67, 150, 33, 2, 187, 174, 47, 173, 160, 6, 78, 181, 191]. Once again an iterative optimization scheme is

Figure 2.1 A statistical shape model (SSM), illustrating the first (left), second (middle) and third (right) principal components (PCs) of variation. The femur shape with the darkest shade corresponds to the mean configuration of the SSM, and is identical in each of the three depictions. The femur shapes represented in lighter shades correspond to SSM configurations at -3σ and $+3\sigma$. The reader should note that the principal components of an SSM are ordered according to the amount of variation that each accounts for.

adopted, and is concluded once the similarity measure value exceeds a pre–defined threshold (A basic diagram of the algorithm is shown in Figure 2.2). In this case no feature–detection is necessary as the entire 3D object is projected into 2D space. This projection, known as a digitally reconstructed radiograph (DRR), can be as simple a silhouette, thickness–projection or maximum–intensity–projection (MIP) [20, 82, 157], or can model the complexity of the full intensity distribution. Intensity–based approaches are known to be more accurate than feature–based methods, but have a longer computation time [101, 107]. The algorithm suffers from a computational bottleneck resulting from its iterative nature, because an accurate volumetric reconstruction often requires a few thousand DRRs to be rendered (this issue is discussed in more detail in Section 2.4.2) [78]. In addition, methods such as these have a small range of similarity measure values that can be optimized reliably without getting caught in local maxima [107, 101]. Consequently the first instance (or configuration) of the deformable model must be initialized so that its corresponding DRR projection(s) are already somewhat similar to the target X–ray image(s) [106, 101]. This is often accomplished by employing

pose–initialization methods before the registration (optimization and pose–initialization are discussed in greater detail in 2.4.4) [107, 15, 166].

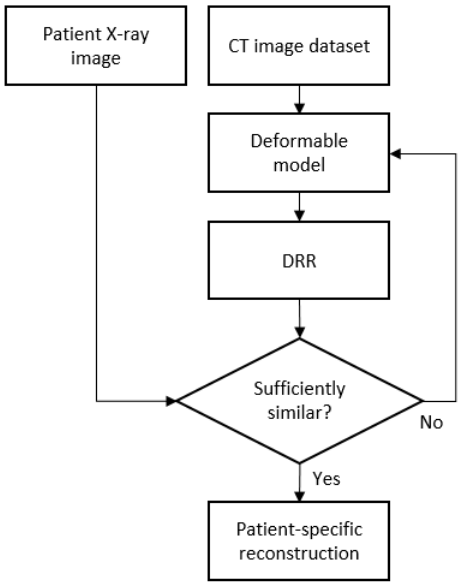


Figure 2.2 Overview of the DRR–based 2D/3D bone reconstruction algorithm. The process is iterative and is concluded once the similarity measure value exceeds a pre–defined threshold.

Two–dimensional–to–three–dimensional bone reconstruction algorithms originally only accounted for shape and therefore only produced surface models [23, 7, 49, 84, 12, 86, 125, 157, 109, 184, 82, 58, 87, 45, 18, 67, 72, 190, 83, 65, 183, 40, 186, 189, 161, 192, 20, 141, 9, 28, 188, 74, 6]. Authors have since sought to include bone density information in their models, which enable volumetric models as well as a number of additional clinical applications [180, 135, 150, 33, 2, 187, 174, 162, 47, 173, 160, 78, 181, 191]. A few clinical application examples include; distinction between cortical and trabecular bone for total hip arthroplasty (cortical regions on the outer hull of a bone absorb X–rays more readily than the trabecular regions which comprise the inner–bone) [47, 181]; estimation of the “local anchorage quality" of osteoporotic bone (sufficient anchorage quality enables prosthesis and implant stability), thereby contributing toward efficient surgical planning for femoral fracture fixations [150]; early identification of individuals at risk of developing osteoarthritis; measuring the progression of osteoarthritis and providing radiographic markers for individuals who require surgery [115, 62, 1, 136]; estimation of the potential risk factors for individuals with osteoporosis (prediction of potential bone fractures and detection of existing vertebral fractures) [131, 57]; aiding in the often problematic task of distinguishing bone deformities from bone fractures [131].

The scope of our review is limited to DRR–based 2D/3D bone reconstruction methods that make use of statistical models and biplanar X–ray images [180, 135, 150, 33, 187, 47, 78, 191]. More specifically, we focus on statistical models of both bone shape and density, which we will refer to as statistical shape and intensity models (SSIMs), and on those trained using CT data since DRR projections of MRI data are unreliable [101]. Statistical shape and intensity models are sometimes referred to as statistical appearance models (SAMs), and are distinguished from statistical shape models (SSMs) which account for shape information alone [187, 136].

Our survey of existing literature found that the terminology used by researchers is ambiguous and inconsistent, and that a unified mathematical formulation of the 2D/3D reconstruction problem has not yet been presented. We remedy this using a common conceptual framework and set of terminology, as well as a unified mathematical formulation of DRR–based 2D/3D reconstruction using SSIMs.

A number of adaptations to this approach have been proposed in an attempt to overcome the limitations of the algorithm. These are discussed according to which of the four main modules of the algorithm they are most relevant to; the SSIM (Section 2.4.1), the DRR rendering method (Section 2.4.2), the similarity measure (Section 2.4.3), the optimization strategy (Section 2.4.4), or the algorithm as a whole (Section 2.4.5).

Finally, we provide a detailed discussion of key issues, persisting challenges and insights for future research (Section III).

2.4 2D/3D bone reconstruction using DRRs

2.4.1 Statistical models: construction, validation, requirements and variations

In order to better understand SSIMs which have been applied to DRR–based 2D/3D reconstruction we first describe the mathematical formulation of SSMs, specifically those that utilize principal component analysis (PCA). The methods used in their construction and validation, as well as the level of accuracy and maximum computation time that is typically required by current clinical applications, are then described.

Construction of SSMs

We closely follow the procedure described by [95] and [136]. The first step in constructing an SSM is to decide how the bone shape is described mathematically. A number of shape features have been proposed including landmark points, dense surface meshes such as Fourier

surfaces and spherical harmonics, medial models such as m–reps and s–reps [122, 123], deformation fields and distance maps; the most commonly used descriptor for bone is landmark points [136]. The simplest means of labelling landmarks is to manually select bone shape features that can be easily identified by an anatomist or radiographer (these usually number in the order of tens) [136]. However, most modern approaches use a dense set of landmarks (which number in the order of thousands). A shape, Γ_n, is first defined as a set of points:

$$\Gamma_n = \{a_p^n \in \mathbb{R}^3, p = 1, ..., P\} \tag{2.1}$$

for the n^{th} shape, Γ_n, $a_p^n = (x_p^n, y_p^n, z_p^n)$ where x_p, y_p and z_p denote the Euclidean coordinates of the p^{th} point. The shape, Γ_n, can then be represented as a vector:

$$\mathbf{s}_n = (x_1^n, y_1^n, z_1^n, ..., x_P^n, y_P^n, z_P^n) \quad \mathbf{s}_n \in \mathbb{R}^{3P} \tag{2.2}$$

Once the same number of points have been used to define all the shapes that comprise the dataset, $n = 1, ..., N$, it is necessary to ensure that the points of each shape are in correspondence with one another. This means any particular point $(a_p^1, ..., a_p^N)$ should describe the same anatomical feature across the dataset. A non-exhaustive list of methods for establishing correspondence among point-sets includes the Iterative Closest Point algorithm or variations thereof, Coherent Point Drift, and Robust Point Matching [14, 114, 35]. Additional information regarding such methods is provided in [81]. Furthermore, since SSMs are good regularizers themselves, they can also be used for landmark–based surface registration, as described by [13].

Next, it is necessary to align the bones so that variations resulting from anything other than shape (such as rotational and translational effects) are eliminated; the most common method for doing this is generalized Procrustes analysis (GPA) [94].

Once aligned, it is possible to model the shape variations using a normal distribution:

$$\mathbf{s} \sim \mathcal{N}(\bar{\mathbf{s}}, S) \tag{2.3}$$

where the mean, $\bar{\mathbf{s}}$, and covariance matrix, S, can be estimated as follows:

$$\bar{\mathbf{s}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{s}_n \tag{2.4}$$

$$S = \frac{1}{N-1} \sum_{n=1}^{N} (\mathbf{s}_n - \bar{\mathbf{s}})(\mathbf{s}_n - \bar{\mathbf{s}})^T \tag{2.5}$$

Finally, principal component analysis (PCA) can be performed to determine common configurations of bone features, represented as an average point distribution and its principal modes of variation (also referred to as eigenmodes or eigenvectors) [136]. The concept of principal components is illustrated in Figure 5.1. This leads to a probabilistic representation for **s**:

$$\mathbf{s} = \bar{\mathbf{s}} + \sum_{m=1}^{M} w_m \sqrt{u_m} \mathbf{e}_m \qquad w_m \sim N(0,1) \tag{2.6}$$

where M is the number of principal components, u_m and $\mathbf{e}_m$ represent the m^{th} eigenvalue and eigenvector respectively, and w_m represents the m^{th} model parameter of **s**. A more detailed mathematical description of this process is provided by [95, 136, 149]. The eigenvectors are intrinsically sorted in descending order so that $u_m > u_{m+1}$ and the variance of **s** is equal to the sum of the individual component variances, since the parameters, $w_1, ..., w_m$, are uncorrelated [95]. Then if u_m decays rapidly, **s** can be accurately approximated using only the first G components:

$$\mathbf{s} \sim \bar{\mathbf{s}} + \sum_{m=1}^{G} w_m \sqrt{u_m} \mathbf{e}_m \tag{2.7}$$

The number of components to use are typically selected by increasing G until the ratio of accumulated variance to total variance, γ, reaches a threshold value:

$$\gamma = \frac{\sum_{m=1}^{G} u_m}{\sum_{m=1}^{M} u_m} \tag{2.8}$$

Commonly accepted values for γ are provided in [63].

Validation of SSMs

In order to assess the quality of an SSM, a number of statistics are typically computed: compactness, specificity and generality [152, 112].

Compactness is a model's ability to faithfully capture shape variance while using a minimum number of principal components [60]. This characteristic measures the cumulative variance that the model can account for, as a function of the number of principal components, M, that are utilized [152].

Specificity measures the ability of the model to generate instances similar to those available within the training set [112]. It is computed as the average distance (such as root–mean–square error (RMSE) or mean–absolute–distance (MAD)) between randomly generated, uniformly distributed, shapes and their nearest match in the training set [152]. The measure is plotted as a function of the number of principal components that are employed

to generate the random shapes. The number of randomly generated shapes are typically large when compared with the number of training examples (often having multiple orders of magnitude), and are proportional to the number of points that comprise the model. An in–depth discussion of distance metrics for evaluating 3D image segmentations is provided in [155].

Generality is the ability of the model to generate instances not explicitly provided by the training set. A leave–one–out strategy is used when a limited number of training examples are available; all but one example is used to train a shape model, which is then fitted to the left–out training example [152]. This process is repeated so that each instance has a turn to be excluded and measured. Ultimately, the average reconstruction error for unseen examples is provided, as a function of the number of principal components used to approximate the left–out training example. Generality is an important characteristic of the model to measure because if the model is over–fitted to the training set, it may have excellent specificity but will hinder its ability to generalize to unseen examples [152].

Variations on SSMs

Recently, PCA–based SSMs have been reformulated as Gaussian process morphable models (GPMMs) by [95]. The advantage of this reformulation is that it does not restrict the covariance function to be the sample covariance matrix (obtained from the training examples), but any valid positive definite covariance function. The covariance function can thus be analytically defined to include prior knowledge and intuitions about a shape, even when training data are not available. Another advantage of this approach is that it is formulated in the continuous domain. This means that dense correspondence can be established for any required resolution, and can therefore be tailored to a specific application. The reader is directed to [95] for an in–depth explanation of GPMMs.

Construction of SSIMs

Digitally reconstructed radiograph–based 2D/3D reconstruction technically requires bone intensity information to be included in the deformable model and, since this is absent from 3D surface models (SSMs), such information is often estimated in a simple manner using projection techniques such as MIP. Efforts have since been made to extend SSMs and accurately encode the intensity information present in example CT datasets. By training the model in a similar fashion to SSMs, they can also account for the average bone intensity and bone intensity variation [38, 84, 86, 58, 87, 7, 109, 125, 65, 45, 72]. The construction of SSIMs involves a few extra steps compared with SSMs, the most important of which

involves establishing correspondence not only on the surface of the shape, but also within the volume [136, 19]. Statistical shape and intensity models can be broadly categorized into two types of representation [19]. One representation models a segmented CT volume (of the bone–of–interest) as a mesh in which mesh–morphing techniques such as those proposed in [56] and [22] are used to establish correspondence. The second representation models the volume as an image using the original CT voxels. In this case correspondence is established using image registration algorithms such as free–form deformation, or the diffeomorphic Demons algorithm [19]. In a comparison of the two types of model representation (mesh and image–based SSIMs), the former was better at reproducing shape information, while the latter was better at reproducing intensity information [19].

Once correspondence has been established, each training instance is fitted to a target reference shape; the mean shape of the SSM is typically used. Then the intensity information is sampled from the mean, and concatenated into an intensity (texture) vector [136]:

$$\mathbf{i}_n = (i_1^n, ..., i_P^n) \quad \mathbf{i}_n \in \mathbb{R}^P \tag{2.9}$$

The intensities are usually normalized to reduce the effects of global intensity variations [149]. Finally, the SSIM can either model the two types of information separately, known as an independent SSIM, or together as a combined SSIM [32, 136].

Independent SSIMs have the advantage of enabling more accurate patient–specific reconstructions, but are more time–consuming to implement [136]. The intensity variation is modelled separately, using PCA, in a similar process to that used to obtain a shape model:

$$\mathbf{i} = \bar{\mathbf{i}} + \sum_{q=1}^{Q} w_q \sqrt{u_q} \mathbf{e}_q \quad w_q \sim N(0,1) \tag{2.10}$$

where Q is the number of principal components of the intensity model, $\mathbf{i}$, u_q and $\mathbf{e}_q$ represent the q^{th} eigenvalue and eigenvector respectively, and w_q represents the q^{th} model parameter of $\mathbf{i}$. An independent SSIM was implemented by [187]; they constructed a separate shape model and intensity model, in a similar fashion to the method provided by [132], using a dataset containing 20 dry, cadaveric femurs. They used the diffeomorphic Demons algorithm [108] to establish dense correspondence between the femur volumes in an automated fashion before completing two distinct PCAs for shape and intensity.

Combined SSIMs are easier to implement, have fewer parameters to optimize, and require less memory than independent SSIMs, ultimately reducing the time taken to search for patient–specific model parameters [136]. They model the correlations between bone intensity and shape information which will affect the quality of the reconstructed model,

however, the relationship between these subspaces is not well–studied and requires further investigation [33]. The set of shape and intensity model parameters are combined into a vector as follows [136]:

$$\mathbf{w}_{\text{combined}} = \begin{pmatrix} \mathbf{W}_m \mathbf{w}_m \\ \mathbf{w}_q \end{pmatrix} = \begin{pmatrix} \mathbf{W}_m \mathbf{E}_m^T (\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{E}_q^T (\mathbf{i} - \bar{\mathbf{i}}) \end{pmatrix} \tag{2.11}$$

where $\mathbf{E}_m^T$ and $\mathbf{E}_q^T$ are the transposed eigenvector matrices of shape and intensity, respectively. A diagonal matrix of shape parameters, $\mathbf{W}_m$, is used to reconcile the shape and intensity model parameters, since they have different magnitudes [136]:

$$\mathbf{W}_m = \beta \mathbf{I} \tag{2.12}$$

where $\mathbf{I}$ is a unit matrix and β is the ratio of total intensity variation to the total shape variation [136]:

$$\beta = \frac{\sum_{m=1}^{M} u_m}{\sum_{q=1}^{Q} u_q} \tag{2.13}$$

A more compact model is then obtained by discarding the correlation between the shape and intensity model parameters, by applying a third PCA on $\mathbf{w}_{\text{combined}}$ [136].

An alternative approach to constructing a combined SSIM is to perform a single PCA on a volume vector:

$$\mathbf{v}_n = (x_1^n, y_1^n, z_1^n, i_1^n, ..., x_P^n, y_P^n, z_P^n, i_P^n) \quad \mathbf{v}_n \in \mathbb{R}^{4P} \tag{2.14}$$

similarly to the process described in Section 2.4.1, where the n^{th} volume, $\mathbf{v}_n$, is represented by a discrete set of P landmark points, x_p^n, y_p^n and z_p^n denote a point's Euclidean coordinates, and i_p^n denotes its corresponding intensity value.

Validation of SSIMs

The method and means with which the quality of the newly included intensity information is validated is not discussed in great detail in the literature. This is most likely due to the fact that the level of fidelity that is required is application–specific. The authors of [19] and [182] made use of dice coefficients to measure the segmentation accuracy of different bone regions, such as cortical and trabecular bone. These provided a percentage of volume overlap between manually segmented ground–truth volumes and their model reconstructions. An in–depth review of metrics that have been used for evaluating 3D medical image segmentations, such as volume overlap, is provided by [155]. These, however, only evaluate shape information. In order to gauge the accuracy of the reconstructed intensity information, voxel–wise com-

parisons have been proposed [188, 174, 66]. In fact any intensity–based similarity metrics, such as those discussed in Section 2.4.3, can be extended to 3D and used. Futhermore, SSIM analogs for model quality metrics (such as generality, specificity and compactness) have not been encountered in the literature.

Variations on SSIMs

A combined SSIM known as the “InShape" model was proposed by [51], which was trained with a dataset of 15 femur specimens. They made use of Euclidean distance maps to represent the surface of the model and an intensity model similar to the one proposed by [38] to represent the intensity information. These were then combined into a spatial–intensity distribution using a level–set segmentation. The resulting SSIM, therefore, essentially stores texture–based features [51].

The authors of [180] followed the original method provided by [38] for constructing an SSIM, but opted for a novel combined representation of shape and intensity information. A hierarchical tetrahedral mesh was used to describe bone shape. While this type of data structure is more involved, it provides a high degree of flexibility and is superior when adapting to local shape structures [179]. Once established, each tetrahedron in the tetrahedral mesh was assigned an analytical function in the form of a Bernstein polynomial using a barycentric coordinate system. Bernstein polynomials provide a more efficient means of representing intensity information compared with storing an intensity value for each voxel. The intensity functions are continuous, and in explicit form, which make them easier to deform or integrate and thus ideal for non–rigid registration as well as efficient DRR rendering [179]. The greater the number of voxels encapsulated by each tetrahedron the more memory-efficient the storage of the density model, the shorter the rendering time of a DRR [179]. However, when the number of tetrahedra used approaches the number of voxels in a CT image there is an improvement in the average intensity difference and standard deviation of intensity difference when compared to a ground-truth DRR. The advantages of representing the model in this way are therefore most apparent when a coarser sampling is used. In addition, the barycentric coordinate system is symmetric, and normalized, which causes the intensity functions to be shape invariant despite the combined nature of the SSIM [179]. This characteristic is advantageous when performing computations on, and deforming, the SSIM. This SSIM paradigm has since become a popular method of incorporating CT intensity values into SSMs; it is used by [135], [47], and [78].

The authors of [135] extended the method proposed by [180] by altering the way in which the tetrahedral mesh is constructed so that it is more conducive to their novel DRR rendering approach. This is discussed in more detail in Section 2.2. They also only made

use of the voxel intensity values of a single patient while the method in [179] performed statistical analysis on the intensity information of all the patients.

Clinical performance requirements for statistical models of bone

There is little information in the literature regarding the clinical performance requirements of 2D/3D registration using SSIMs. One example does provide some guidance with regard to shape information; after an analysis of common orthopaedic procedures the following specifications for the accuracy and robustness of a bone shape model were proposed by [91], in order to be useful for surgical guidance:

- Root–mean–squared registration error ranges of 1 *mm* – 1.5 *mm* (2 *mm* – 3 *mm* in the worst case).
- The registration is successful in its first attempt at least 95% of the time.
- The registration takes no more than 1 minute.
- Simple and minimal preoperative and intraoperative user interaction.

To the best of our knowledge, specifications for the accuracy and robustness of the estimation of intensity information using 2D/3D registration are not available in the literature and appear to be application–specific. For example, [173] note that the trabecular region is predominantly affected by osteoporosis, which is the primary cause of compression fractures. The accuracy of this region would therefore be prioritized when developing SSIMs for clinical applications aimed at osteoporosis.

2.4.2 Rendering of DRRs

Digitally reconstructed radiograph projection is a rendering technique that simulates the X–ray imaging process in order to produce a synthetic X–ray image from a CT volume or volumetric model (An example is shown in Figure 2.3). The most commonly used method to do so is “ray–casting" [110]. It is derived from the Beer–Lambert Law (Equation 2.15), which describes the attenuation of X–rays through matter. In this approach the value of each pixel of the DRR image is computed by evaluating each voxel of a CT volume (or instance from a SSIM) which is encountered along the path of a single ray between the current pixel and a center–of–projection (COP) (see Figure 2.4). When a voxel is evaluated, its corresponding linear attenuation coefficient (LAC), as well as the distance the ray has “travelled", is considered. A line integral along the path of a single ray is evaluated where I is the X–ray signal intensity, L is the path from the X–ray source (COP) to a detector pixel,

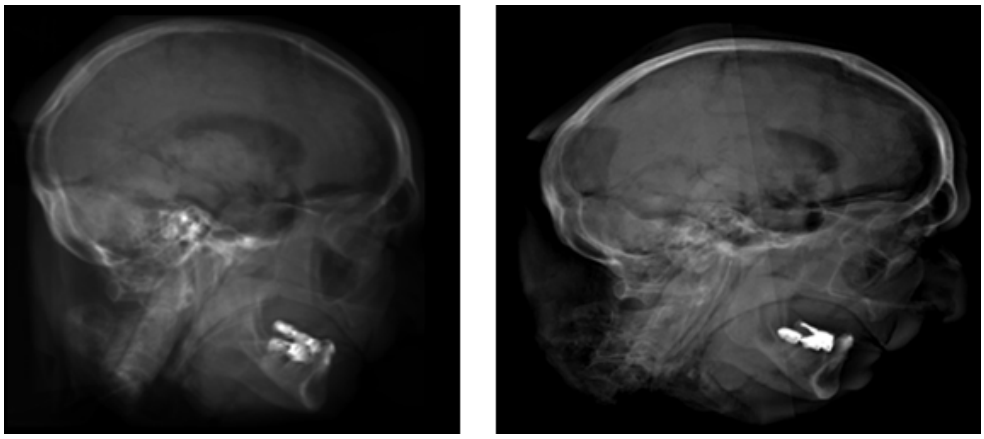


Figure 2.3 Example of a DRR (left) obtained using ray–casting through a CT volume of a patient, contrasted against an actual X–ray image of the same patient (for the Lateral view).

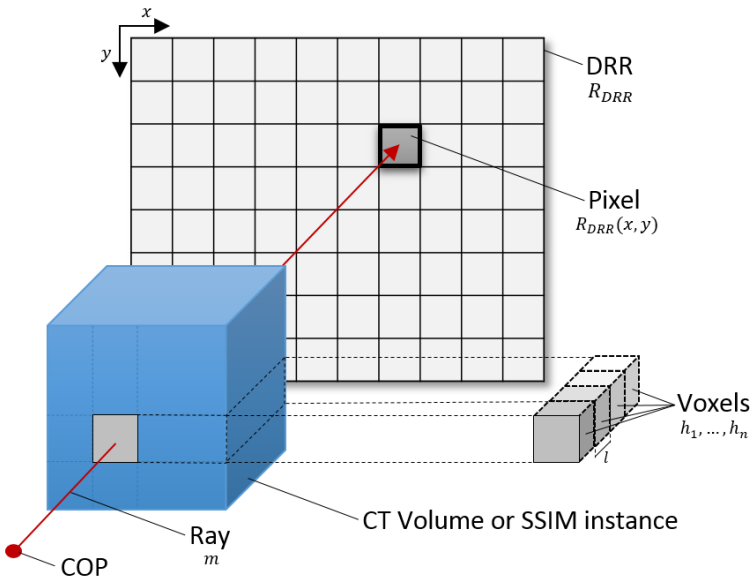


Figure 2.4 An illustration of the ray–casting method. Each DRR pixel intensity value $R_{\text{DRR}}(x,y)$, is computed by evaluating the CT voxels encountered along a ray m, between a COP and the current pixel.

$\mu(E,r)$ is the energy and position–dependent LAC of the material that the X–ray passes through and $I_0(E)$ is the incident X–ray intensity and energy spectrum [70]:

$$I = \int I_0(E) \times \exp(-\int_L \mu(E,r)dl)\, dE \tag{2.15}$$

This formulation is further simplified by modelling the X–ray source as monoenergetic; where more complex physical phenomena such as “scatter contamination", “beam hardening", and “veiling glare" are disregarded; and then discretized along the ray path to produce a voxelized

representation [70, 43, 147]:

$$I = I_0 \exp(-\sum_n \mu_n l_{n,m}) \tag{2.16}$$

where n and m designate a specific voxel and ray, respectively; l is the intersection length and μ is the LAC. The corresponding LAC for a voxel's Hounsfield unit is obtained using Equation 3.13. This value is contingent on the relative LAC of water, for the particular energy spectrum of the incident rays [147].

$$h_n = \frac{(\mu_n - \mu_w) \times 10^3}{\mu_w} \tag{2.17}$$

where h_n is a CT Hounsfield unit, μ_w is the LAC of water for a specific CT energy, and μ_n is the LAC for the current voxel. The COP is either modelled as being close in a perspective projection scheme, or infinitely far away in an orthographic projection scheme where the rays are parallel to one another (See Figure 2.5). The most commonly used scheme is perspective projection, which is more challenging since it involves more complex projection geometry and because it is possible for adjacent voxels to be traversed simultaneously. A number of sampling methods (also known as voxel traversal, ray traversal, or compositing methods) have been proposed to address this issue, often implementing some type of voxel interpolation. These are however beyond the scope of this paper, but are discussed in detail in [171] and [64]. The DRR rendering method can be viewed as a transform, $\mathscr{P} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$. A general formulation of the ray-casting projection through an SSIM, regardless of the sampling or projection strategy that is used, can then be described as follows [98]:

$$\mathscr{P}(\mathbf{v}(\mathbf{w})) = R_{\text{DRR}} \tag{2.18}$$

$$R_{DRR}(x,y) = I_0 \exp(-\int_{L_{x,y}} \mu_l \, dl) \tag{2.19}$$

where $\mathbf{v}(\mathbf{w})$ is an SSIM instance obtained with a specific set of model parameters $\mathbf{w}$, and $R_{\text{DRR}}(x,y)$ is a function that returns an accumulated intensity value along ray-path $L_{x,y}$ from the X–ray source (COP) to a detector pixel at the 2D coordinates, (x,y). For projection geometry-specific formulations of this equation the reader is directed to [98].

Variations on DRR rendering

DRR–based approaches are the most frequently reported type of 2D/3D registration methods in literature [101]. They are known to be more accurate than feature–based methods at the cost of having a longer computation time [106, 101]. The iterative nature of the algorithm causes a computational bottleneck to emerge as accurate volumetric reconstruction often

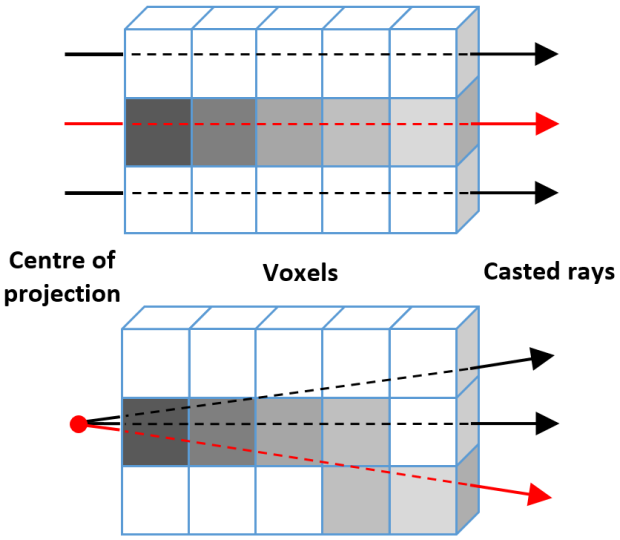


Figure 2.5 An illustration of two DRR ray projection schemes. Orthographic (or orthogonal) projection models the COP as being infinitely far away (above) as opposed to perspective projection which models the COP as a point–source (below). Interpolation is typically used in order to overcome the simultaneous traversal of adjacent voxels when sampling along perspective–projected rays. The increasing greyscale values of the voxels indicate the order in which the voxels have been traversed.

requires a few thousand DRRs to be rendered [78, 107]. Efforts have since been made to reduce the time taken to obtain a DRR projection. These can be roughly categorized as hardware and software–based methods, but are not mutually exclusive distinctions [101].

Hardware–based approaches exploit parallel processing through the use of commercial graphics processing units (GPUs) to speed up the ray–casting algorithm, and consequently the registration process, in some cases by up to 25 times [101, 85, 156, 146].

Software–based approaches apply concepts from the computer graphics community to provide more efficient rendering algorithms. These take the form of image–order and object–order techniques. Image–order techniques involve backward projection and are typically performed pixel–by–pixel. These typically include adaptations to the ray–casting method (which itself is an image–order technique) either by reducing the number of rays, or number of samples that are considered as is the case with Monte Carlo volume rendering [89]. Object–order techniques involve forward projection and are typically performed voxel–by–voxel. These include alternatives to the standard ray–casting approach such as shear–warp factorization, wobbled splatting, Fourier volume rendering and cylindrical harmonics, among others [71, 80, 16, 100, 168]. Most object–order methods are geared towards rigid registration for purposes such as preinterventional patient alignment (image–guided registration therapy) [43]. Here it is assumed that the volume is static, while the camera pose/COP is searched

for. It therefore makes sense that a popular strategy for object–order methods is to spend computational resources encoding the volume in such a way that, once initialized, can obtain a DRR from different COPs with ease. Non–rigid reconstruction, on the other hand, seeks to find a patient's specific volume by iteratively deforming an appearance model while the COP remains stationary. Image–order techniques are more amenable to this strategy.

Some methods, such as those proposed by [135], [47], and [78], have been specifically applied to the reconstruction paradigm, and utilize a combination of hardware and software–based approaches. The SSIM proposed by [180] (see Section 2.4.1) was accompanied by a complimentary DRR rendering algorithm that adheres to the non–linear intensity distribution of its tetrahedra [47]. The algorithm proposed by [135] improved upon the DRR rendering method [179]. They noted that a X–ray attenuation value does not need to be accumulated according to the order of visibility of tetrahedra along a line integral. Thus each tetrahedron can be processed independently using a projected tetrahedra approach (replacing the numerical evaluation rays intersecting the SSIM), making the algorithm both a closed–form solution and ideal for GPU implementation. This characteristic also compliments the method in [179] which uses barycentric coordinates, allowing for the simultaneous classification of the projected outline of the tetrahedra. The approach provided by [47] further improved that of [135], by focusing effort on the DRR rendering step of the reconstruction algorithm. The authors of [47] provided a way to combine the deformation and tetrahedral mesh projection steps thus enabling a full GPU implementation. They employed cell–based ray–casting which reuses the closed–form method for ray integral computation, but implemented a novel tetrahedron thickness calculation instead of the projected tetrahedra approach. This permits the advantages of a barycentric coordinate system while circumventing the need to classify the projected outline of the tetrahedra [78, 47]. A summary of notable DRR rendering methods is provided in Table 2.1 (note that $\mathcal{O}(n^2)$ represents quadratic time, $\mathcal{O}(n^3)$ represents cubic time and $\mathcal{O}(n^2\log[n])$ represents quadratic–logarithmic time).

Finally, the accuracy of the registration can be further improved through the generation of more realistic DRRs, which will inherently be more similar to a patient's X–ray image(s) [51, 150]. The ray–casting method only models the attenuation of primary photons while disregarding secondary effects such as beam hardening. Techniques to account for these effects were proposed by [147], which produce more realistic DRR renderings but also increases the already taxing computational complexity of the DRR rendering module.

Table 2.1 Notable DRR rendering methods which include software, hardware, and subspace/segmentation–based approaches

Author(s)	Name of technique	Voxel/ tetrahedra–based	Time complexity (if available)	GPU–based acceleration	Comments
[143, 68]	Ray–casting	Voxel	$\mathcal{O}(n^3)$	None	- Used as a benchmark. - Computationally inefficient.
[135]	Projected tetrahedra	Tetrahedra	Not available	Partial	- Suitable for mesh–based 2D/3D reconstruction.
[133]	Light fields	Voxel	$\mathcal{O}(n^2)$	No	- Reduced time complexity. - The 4D space must be densely sampled to provide satisfactory images.
[100, 117]	Fourier volume rendering	Voxel	$\mathcal{O}(n^2\log[n])$	Full	- Large memory demands. - Independence of image quality on the sampling step size. - Perspective projection not supported.
[80]	Shear–warp factorization	Voxel	$\mathcal{O}(n^3)$	None	- Well suited rigid 2D/3D registration. - Reduced computations when recomputing projections.
[16]	Splatting	Voxel	$\mathcal{O}(n^3)$	None	- Very accurate, but difficult to implement. - Aliasing artefacts.
[168]	Cylindrical harmonics	Voxel	Not available	None	- Fast. - Reduced computations when recomputing projections.
[47, 78]	Cell–based ray–casting with novel tetrahedral thickness calculation	Tetrahedra	Not available	Full	- Suitable for mesh–based 2D/3D reconstruction.

2.4.3 Similarity measures

One of the primary factors influencing the accuracy of intensity–based 2D/3D reconstruction algorithms is the similarity measure, the choice of which can produce drastically different registration results [186, 169]. The DRR(s) and X–ray image(s), R_{DRR} and R_{XRAY}, are compared to one another with regard to some statistic or set of features and a numerical value is produced which quantifies the similarity of information between them. In some cases researchers refer to a cost, loss, error, merit, criterion, energy, or objective function, which the optimization strategy seeks to minimize (or maximize).

In the context of deformable image–based 2D/3D reconstruction the similarity measure can be framed as a Q–dimensional function where Q represents the number of parameters of the SSIM. The value of each location of a Q–dimensional parametric search space then corresponds to the similarity measure value for a specific set of model parameters, $\mathbf{w} = \{w_1, ..., w_Q\}$ (an instance of the SSIM) [144, 101].

The similarity measure values for each X–ray image view/COP can be consolidated as follows:

$$\text{sim}_{\text{tot}}(\mathbf{w}) = \sum_{c=1}^{C} \text{sim}(\mathscr{P}_c(\mathbf{v}(\mathbf{w})), R_{XRAY,c}) \tag{2.20}$$

where $\mathscr{P}_c$ represents the DRR rendering method for a COP c, and C is the total number of patient X–ray images.

Variations on similarity measures

The mutual information (MI) similarity measure is commonly used by 2D/3D reconstruction algorithms [169]. It compares the probability distribution of the pixel intensity values (histograms) in a DRR and X–ray image. While it is robust to noisy and occluded X–ray images, it only considers intensities, thus disregarding any spatial information [135]. A number of MI variations have since been proposed which incorporate spatial information. The asymmetric multi–feature MI similarity measure uses additional spatial features in the form of intensity gradients and is especially effective when reconstructing from a small number of patient X–ray images [101]. The performance of two novel variations of the MI similarity measure, distance coefficient MI and distance weighted MI, were compared by [76]. They found that these measures outperformed conventional MI without having a significant impact on computation time. The MI similarity measure was also adapted to include spatial information by [185], by incorporating the Kullback–Leibler bound into a Markov random field model. Their experiments show that their measure, which they have

named the Maximization of Mutual Information, not only outperforms the MI similarity measure, but is also robust to occlusions present in X–ray images.

The researchers in [135] discuss the similarity measure-related complications that arise when the statistical model is trained using a dry-bone dataset. Images of live subjects often contain other organs, as well as soft tissue, which adds noise and reduces the contrast of the images. Despite the merits of MI they found that it did not perform well and therefore included a soft-tissue model in their SSIM in order to ameliorate these effects. Alternatively, the researchers in [79] proposed the use of normalized cross correlation (NCC) as an initial similarity measure and then automatically switched to using the variance weighted sum of localized normalized correlation (VLNC). The NCC similarity measure is not very accurate, but reduces the influence of regions containing foreign objects. The VLNC focuses on high-variance regions, assuming that these contain relevant information. In doing so they improve the accuracy of the final registration.

An extensive study of six similarity measures was completed in the context of intensity–based 3D–2D registration by [121]. These included cross correlation, entropy, mutual information, gradient correlation, pattern intensity and gradient difference. By comparing these against a gold standard (landmark–based registration), the measures were ranked with regard to how accurate and robust they were. The study concluded that the pattern intensity and gradient difference similarity measures were accurate and robust, even when interventional instruments, thin–line structures and soft tissue were present [186].

The authors of [101] have suggested that the similarity measure utilized by the optimization strategy be specifically adapted to the understanding of the image formation process as well as the relationship between the intensities of the DRR and X–ray images. One such similarity measure was proposed by [111], which was tailor-made to the statistics of CT and X–ray image acquisition. In this approach Poisson and Gaussian probability distributions were used to model the intensity values of the two modalities (since the photon noise which is present in authentic X–ray images is Poisson-like in nature). Ultimately, new similarity measures were calculated from the assumed distributions, using maximum likelihood estimation, which were especially robust to image noise [111, 101]. Another such similarity measure was presented by [42], who established the feasibility of constructing a similarity measure using coefficients from an orthogonal set of base functions, and by decomposing X–ray and DRR images into orthogonal Zernike moments [42, 101]. The advantages of doing so include better robustness to histogram differences, invariance to in-plane rotation, and control over the level of detail that is considered. These bespoke similarity measures, however, have yet to be applied specifically to the 2D/3D reconstruction paradigm; their

performance in this regard, as well as in comparison to other similarity measures, is therefore unclear.

2.4.4 Optimization strategies

The optimization strategy for a deformable image-based 2D/3D registration algorithm is iterative, and describes how a patient-specific bone reconstruction is searched for. This is accomplished by providing a procedure for how the model parameters for the next iteration are chosen based on the differences between the patient's X–ray image(s) and the DRR projection(s) for the current iteration. As previously mentioned, these differences are measured using one or more similarity measures, which the optimization strategy seeks to maximize:

$$\mathbf{w}_{\text{patient}} =_{\mathbf{w}} \text{sim}_{\text{tot}}(\mathbf{w}) + R(\mathbf{w}) \tag{2.21}$$

where $\mathbf{w}_{\text{patient}}$ is the final set of optimal model parameters selected to represent the patient's bone volume. The regularization term, $R(\mathbf{w})$, accounts for the nature and constraints of the transform, $\mathscr{P}_c$ [145]. These include prior knowledge regarding the bone's shape and intensity, ensuring smooth deformations and dampening the effects of outliers (for an interesting discussion regarding how the regularization term is related to the covariance matrix, S, the reader is referred to [151]). Figure 2.6 provides a useful overview of the DRR-based 2D/3D reconstruction algorithm, including important mathematical formulae.

The similarity measure, when represented as a function of registration error, should ideally be a convex, monotonic function, having a sharp minimum (indicating the true registered position or best possible similarity value) and successively decreasing values the further away from the minimum they are [101, 144]. This ensures good registration results when implementing iterative, local optimization strategies such as Powell's method, the downhill–simplex method, gradient–descent, [150, 135, 47, 101].

However, as is the case with intensity–based 2D/3D registration, the registration error function is highly non–convex, making the optimization strategy susceptible to getting caught in local minima [145, 107]. The similarity measure therefore has a small capture–range, which is the subset of values around the global minimum which do form a convex, monotonic function (see Figure 2.7) [144].

Variations on the optimization strategy

In order to widen the capture range, and increase chances of convergence, researchers have implemented global optimization methods and heuristic search strategies such as simulated

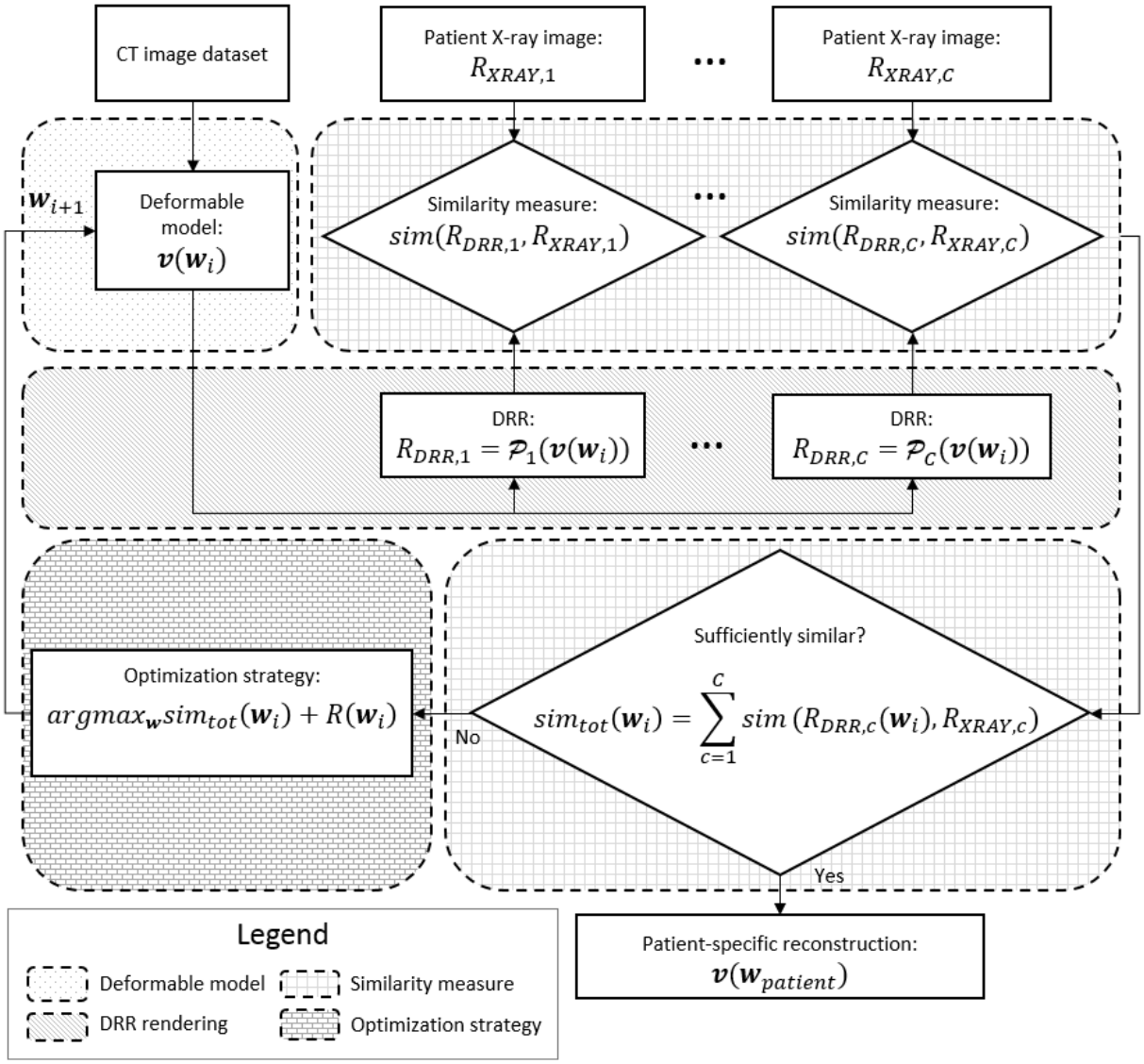


Figure 2.6 A detailed overview of the DRR–based 2D/3D bone reconstruction algorithm (including a mathematical formulation). The modules which comprise it include the deformable model, of which an SSIM $\mathbf{v}(\mathbf{w}_i)$ is one type, DRR rendering R_{DRR}, similarity measure sim_{tot}, and optimization strategy. The process is iterative, where $\mathbf{w}_i$ is the set of model parameters for iteration i, and which is concluded once the value of sim_{tot} exceeds a pre–defined threshold.

annealing and Monte Carlo random sampling [99, 41, 101]. While these increase the probability of finding the global maximum, the authors of [101] note that they are problem–specific. We are not aware of any DRR–based 2D/3D reconstruction methods that make use of these.

A simpler and more well established approach is to use a hierarchical multi–resolution and/or multi–scale optimization strategy, adopting a coarse–to–fine approach to the parameter

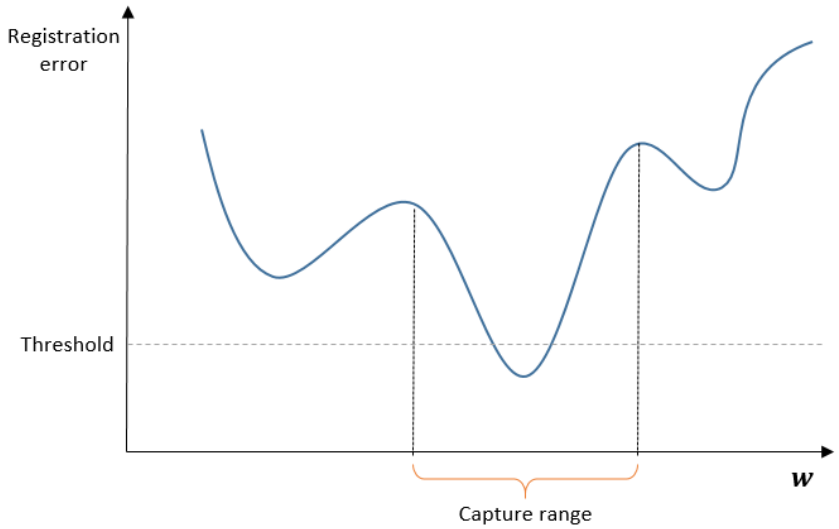


Figure 2.7 A plot of intensity–based 2D/3D registration error as a function of the model parameter values. The range of values around the global minimum which form a convex, monotonic function is known as the capture–range.

search [101, 153, 180]. They first downsample the images to reduce their complexity (and avoid local maxima) and then successively introduce finer detail in stages. The complexity which is gradually increased refers either to that of the data, warp or model [153]. An extensive survey of multi–resolution approaches is provided by[88]. The authors of [180], [33] and [47] implemented similar multi–resolution optimization strategies. They downsampled the resolution of the X–ray images and reduced the number of model parameters which, during later iterations, were both increased to ensure an accurate reconstruction of the pelvis bone. These optimization strategies not only increase the capture range of the algorithm (by avoiding local maxima), but also improve its computation time. The authors of [101] noted that downsampling the image may suppress some image features while making other non–corresponding features more similar. However, they also noted that these approaches are especially suitable for 2D/3D reconstruction.

Linear optimization strategies such as the downhill simplex method, which was implemented and applied to the femur and pelvis bones by [135], are sometimes preferred to non–linear strategies as they are often computationally cheaper [20, 47]. However, from what is evident in intensity–based 2D/3D bone reconstruction literature, a number of non–linear alternative optimization strategies have been proposed which provide more accurate model reconstructions [103]. Furthermore, linear strategies are limited in that they have an inability to cope with large deformations [145]. Levenberg–Marquardt optimization is a non–linear strategy that is reported by [78], [187]; both applied it to the femur bone because it provided accurate reconstructions while still outperforming gradient–descent–based approaches with

regard to computation time [78]. Levenberg–Marquardt optimization was also implemented by [191], who applied it to the pelvis.

Another way to avoid local maxima is to implement a pose initialization method before proceeding with the optimization strategy [107, 164]. Such methods aim to initialize the first SSIM instance in such a way that its initial DRR projection(s) are relatively close to that of the target X–ray images. This is typically accomplished with a manual landmarking process, which is time–consuming. An accurate initialization method requiring only three manually labeled landmarks has been developed by [113], which uses epipolar geometry to assist in localizing points between two X–ray perspectives. However, methods such as these are still prone to human error [28, 29]. Fully automatic methods have since been explored, but face many challenges; noise, poor image contrast and the superposition of bone structures at different depths make it difficult to distinguish the bone–of–interest [29]. Random forest regression was implemented by [181] for fully automated contour extraction and pose initialization. This method provided accurate results, but was only tested using simulated patient X–ray images. They also did not comment on the method's influence on the computation time of the overall algorithm.

Finally, supervised learning–based similarity measures have been demonstrated by researchers to widen the capture–range of intensity–based 2D/3D registration algorithms, but are far more computationally expensive [107]. We are also not aware of any DRR–based 2D/3D reconstruction methods that make use of these.

A summary of the optimization strategies discussed in this section, which have been specifically applied to DRR–based 2D/3D bone reconstruction, is provided in Table 2.2.

2.4.5 Investigations on the algorithm as a whole

The use of two X–ray images enables a significant improvement in the accuracy of the reconstructed model than if only a single X–ray image is used. However, according to [180], the use of additional X–ray images has diminished returns in terms of accuracy. This is arguably not worth the extra computation time or additional radiation exposure to the patient. It was demonstrated in [180] that the best registration accuracy was obtained when the angle between the two perspectives is orthogonal. This makes sense since the information provided by the two images in this case is the least correlated. However, it does not necessarily follow that this is always the case. A good example is the pelvis bone, which has a symmetric structure; an AP and LAT X–ray image combination will have a large superposition of structure - especially in regions containing joints. While orthogonality is important, the selection of the best imaging direction is dependent on the anatomical structure-of-interest.

Table 2.2 A summary and comparison of optimization strategies which have been used for DRR–based 2D/3D bone reconstruction

Author(s)	Nature of SSIM	DRR rendering method	Optimization strategy	Similarity measure	Type of bone	Number of X–ray images	Average reconstruction error (mm)	Type of optimization
[180]	Combined	Not clear (probably ray–casting)	Novel multi–resolution non–linear optimization	(Normalized) Mutual information	Pelvis	2 or more	2.0	Non–linear
[135]	Combined	Projected tetrahedra	Downhill simplex algorithm	(Normalized) Mutual information	Femur and pelvis	2 or more	2.0	Non–linear
[187]	Independent	Ray–casting	Levenberg–Marquardt	Maximization of mutual information	Femur	2 or more	1.5	Non–linear
[47]	Combined	Cell–based ray–casting with novel tetrahedral thickness calculation	Gradient descent with X–ray sub–sampling	(Normalized) Mutual information	Pelvis	1	2.0	Linear
[78]	Combined	Cell–based ray–casting with novel tetrahedral thickness calculation	Levenberg–Marquardt	(Normalized) Mutual information	Femur	2	1.18	Nonlinear
[150]	Combined	Ray–casting	Evolutionary	Mean reciprocal square difference normalized mutual information	Femur	2 or more	4.24	Unclear
[191]	Combined	Ray–casting	Levenberg–Marquardt	Maximization of mutual information	Femur	2 or more	1.2	Non–linear

The 2D/3D reconstruction of joint structures using an articulated statistical shape and intensity model (ASSIM) has been investigated by [46]. They reconstructed pelvic and femoral bones simultaneously in order to infer two patient-specific surgically relevant parameters from a single X–ray image. They extended their previous SSIM, presented in [47], to include both the pelvic and femoral bones, but constrained the 3D rotation of the proximal femur such that its center of rotation corresponds to the center of a sphere made to fit the acetabulum. Their method produced results close to the CT gold standard, but became less accurate when faced with outliers where some regions pertinent to the inference were not visible in the AP X–ray image. They proposed increasing the number of training examples in order to overcome this limitation [46].

The authors of [150] investigated the performance of a number of similarity measure–optimization strategy combinations for appearance–model–based 2D/3D registration of the human femur [136, 51, 150]. These included Powell optimization, 1+1 evolutionary optimization, mean reciprocal square difference metric, Matte's mutual information metric and the normalized mutual information metric [126, 128, 102, 124]. Using the InShape SSIM, the following parameters were optimized: the six degrees of freedom of the rigid 3D transformation (three rotational and three translational) as well as the modes of shape and intensity variation. They found that the mean reciprocal square difference metric in combination with 1+1 evolutionary optimization, as well as normalized mutual information measure combined with the 1+1 evolutionary optimization strategy, resulted in the best performance.

2.5 Discussion

During our literature review we noted a number of challenges and insights regarding DRR–based 2D/3D bone reconstruction, which may serve as future avenues of research. These are discussed according to the algorithm module to which they are most relevant.

2.5.1 Statistical shape and intensity models

While SSIMs are robust to artefacts and noise, they require a large amount of training data in order to obtain a model capable of expressing all possible target shapes. The generalization of the classical SSM approach, known as GPMMs, and introduced by [95], makes it possible to intuitively and analytically define covariance functions, which enable the construction of expressive shape priors even when only a few example shapes are available (essentially providing a framework for combining template and statistically based models, and allowing

for dense point correspondence at any resolution). Their approach could be extended to include SSIMs, ultimately providing more accurate bone shape and intensity estimations.

Furthermore, it is unclear whether independent or combined SSIMs are superior for the purposes of non–rigid 2D/3D bone reconstruction; the relationship between the shape and intensity subspaces is not well studied and requires further investigation.

2.5.2 Digitally reconstructed radiographs

The original patient X–ray image can be regarded as the gold–standard, equivalent to a maximum similarity measure value. However, this is not realistically possible because of the presence of noise. Also, a DRR rendering algorithm's failure to account for secondary imaging effects such as veiling glare and beam–hardening further prevents the DRR from attaining a gold standard value [147]. We therefore define the "ideal" threshold, which takes this compromise into account. A further concession in accuracy is made when we consider that DRRs will actually be obtained from an SSIM instance and not the patient CT volume, since it is not available. Thus another value, the "best–case" threshold, is defined, which accounts for the consequent loss of information resulting from the PCA approximation. Finally, the "sufficient" threshold observes the fact that the best–case threshold may be unattainable since the model parameter search is limited to the 2D domain, as well as the performance of the chosen optimization strategy. It may also be the case that the reconstruction accuracy necessary for a particular application is attained prior to achieving the best–case threshold. An illustration of these concepts is provided in Figure 2.8.

Some researchers have established the feasibility of using partial DRRs to guide the model parameter search. However, the decision regarding which regions to select is made on intuition. Should there be access to a corresponding CT and X–ray image dataset for a number of patients, a potentially insightful study can be completed. Firstly, an SSIM can be built using the (hand–segmented) patient CT volumes. The ideal similarity threshold value (for a patient) can be computed using a DRR obtained from their CT volume, and their X–ray image. The best–case similarity value can be obtained in a similar fashion, using a near–perfect model fit (searched for in the 3D domain using the patient's CT volume). Then, by comparing the original X–ray image to the DRR obtained from the CT volume as well as the DRR obtained from the SSIM, the regions and/or features which should be focused on can be determined empirically.

While a number of DRR rendering methods have been specifically proposed for application to 2D/3D registration, only the ray–casting and tetrahedra–based projection methods have been applied to 2D/3D bone reconstruction. The performance of different DRR volume rendering methods in this context have not yet been compared. Object–order methods such

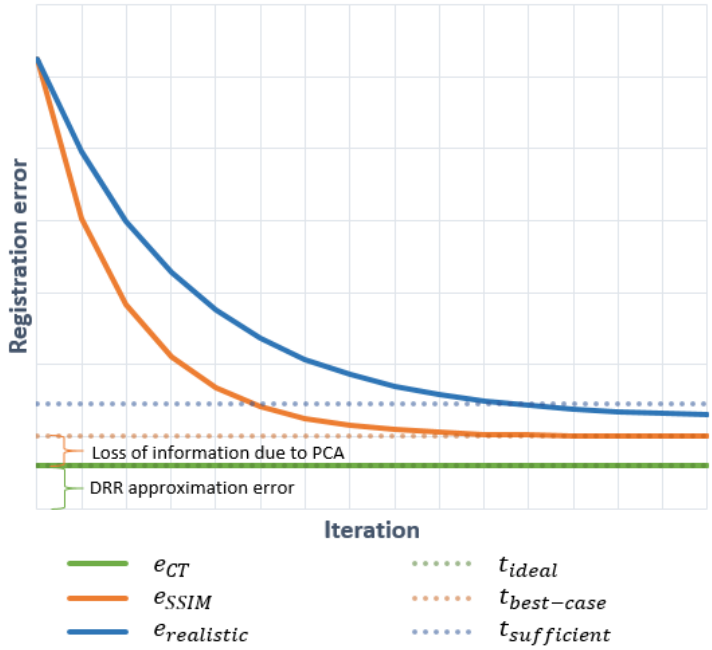


Figure 2.8 Iterative 2D/3D reconstruction optimizations, depicted as decreasing registration error (increasing similarity). The best possible registration accuracy, varies under different circumstances: the ideal threshold corresponds to a DRR obtained directly from the patient CT volume ($e_{CT} = 1 - \text{sim}(R_{DRR,CT}, R_{XRAY})$); the best–case threshold corresponds to a DRR obtained from an SSIM instance, the model parameters of which were searched for using the CT volume ($e_{SSIM} = 1 - \text{sim}(R_{DRR,SSIM}, R_{XRAY})$); the sufficient threshold also corresponds to a DRR obtained from an SSIM instance, but accounts for the fact that the best–case threshold may not be reachable when the model parameters are searched for using only the patient's X–ray image(s).

as cylindrical harmonics, shear–warp factorization and frequency domain–based rendering spend computational resources encoding the volume in such a way that obtaining a DRR from a large number of COPs is made much simpler. This allows DRR projections to be iteratively re–rendered at a reduced computational cost until an optimal COP is found. While this is beneficial for 2D/3D registration (where the assumption is that the object does not change), this is not the case for 2D/3D reconstruction where the COP(s) are fixed and the object is iteratively deformed. However, these methods are of use for the initial pose estimation; it may serve a 2D/3D reconstruction algorithm well to include a two–part DRR rendering module where an image–order technique takes over from an object–order technique once the pose has been established. If additional accuracy is required, the methods provided by [147], which account for secondary physics effects, can be employed to further improve the registration. However, it is important to note that the DRR itself is not the goal of the 2D/3D reconstruction algorithm; it is only a means to determine a model reconstruction (optimizing

a set of model parameters). Any additional DRR rendering accuracy, other than what would enable an accurate reconstruction, is unnecessary, and it would be beneficial to rather trade this accuracy for a faster computation time (See Figure 2.8).

As is shown in Table 2.1, many rendering methods have a $\mathcal{O}(n^3)$ time complexity. These methods should be avoided (at least once the pose of the volume/X–ray source has been established).

2.5.3 Similarity measure

In the majority of cases, the mutual information similarity measure, or some variation thereof, has been used for image–based 2D/3D reconstruction (and is also popular for rigid 2D/3D registration). Newer similarity measures, such as those proposed by [186], [111] and [42] have been specifically adapted to the context of X–ray images, but have not yet been applied to the context of reconstruction. These may prove to be superior to conventional similarity measures.

2.5.4 The algorithm as a whole

The 2D/3D reconstruction of an anatomical structure from a single X–ray image is a valuable area of research because clinical routines that require only one are far more common and cost less. However, since depth and scale are correlated, the precise estimation of either the size of the anatomical structure, or its depth, becomes more challenging when constrained to a single view [182]. The researchers in [83] first demonstrated such an algorithm using a feature-based approach to reconstruct a shape model. Their model was comprised of average templates of both 2D thickness images and height maps that were projected from 3D CT training images. The single X–ray image informed how the templates were aligned and warped using carefully selected landmarks, and then combined to obtain an estimated 3D shape. The researchers in [162] were the first to reconstruct both shape and intensity information from a single X–ray image. They extended [83]'s method by also warping the volumetric BMD, and normalized it, to match that of the input X–ray image. Furthermore, having used a calibration phantom, they were able to estimate actual BMD values. In [47] the researchers reconstructed both shape and intensity information, but made use of an SSIM and no calibration phantom. However, the focus of their research was on the computation time and accuracy of their proposed DRR renderer. In [46] they reported the estimation accuracy of surgically relevant pelvic measurements, but did not explicitly report on the intensity reconstruction accuracy - it was only used as a proxy to improve the pelvic shape estimation.

The aforementioned methods, however, all utilized an artificial X–ray image (DRR) to represent the patient data. The researchers in [186] were the first to demonstrate such an algorithm using a real X–ray image, but only reconstructed a shape model. They first proposed a surface–based iterative scaled rigid registration using a standard AP X–ray image, and then a surface–based iterative affine registration using 2D lateral fluoroscopy [186, 189]. They exploited two parameters, namely the image scale (in mm/pixel) and the distance from the COP to the detector (the imaging plane), which can be retrieved from a DICOM image (provided that the anatomical structure was imaged using a standard clinical procedure). The researchers in [175, 66, 174] and [160] have reported success when using a single image of actual patients, and also reported on the accuracy of the reconstructed shape and intensity information, but only when using DXA. Dual–energy X–ray images provides two advantages over conventional X–ray images. Firstly, DXA uses orthographic projection geometry which eliminates the need for camera calibration [182]. Spatial distortions, such as depth–dependent object magnification effects, which complicate shape estimation are thus avoided [175]. Secondly, DXA is the standard for BMD estimation and the use of a DXA image over a conventional X–ray image, along with a calibration phantom, allows for accurate BMD estimation [120]. However, DXA requires a specialized machine, to which access may be limited. It is not part of a typical clinical routine and has a limited field–of–view. Therefore, we believe that the 2D/3D reconstruction of both shape and intensity information of an anatomical structure from a single conventional X–ray image is still an unsolved problem, and worthy of future research efforts.

There is currently not enough data to objectively compare the performance of different DRR–based 2D/3D bone reconstruction algorithms (in order to determine how suitable they are) [136]. Researchers employ different datasets and are often unclear about the evaluation methodology that they use, nor does a standard methodology exist, to validate and benchmark these algorithms [101]. Furthermore, researchers either focus on different bones or use different types of modules (SSIMs, DRR rendering techniques, similarity measures and optimization strategies) within their algorithm. It is clear that the interactions between these modules are complex and greatly affect the overall accuracy of the model reconstructions, as well as the time spent computing them. It would therefore be worthwhile to formulate a standardized means of comparing and benchmarking 2D/3D bone reconstruction algorithms in order to aid future research efforts. To the best of our knowledge, only [150] have made an effort to compare different 2D/3D reconstruction algorithms; they implemented and compared the performance of different similarity measure–optimization strategy combinations. An ideal evaluation methodology should include a complete open–source algorithm, which would allow the different modules to be interchanged, and measured (this may not always

be possible since algorithm modules such as those proposed by [180, 135] and [47] are interwoven), and would include a standard dataset (for at least one type of bone).

Of the reconstruction algorithms that we have investigated, only [135] and [46] have shown that their algorithm can be applied to more than one type of bone (femur and pelvis), and still yield a sufficiently accurate reconstruction. This is worth considering since unique complexities and challenges exist for different types of bone, such as bone structure superposition when dealing with scapulae, or intricate bone substructures as is the case with the skull. Therefore, the success of an algorithm when applied to one particular type of bone does not necessarily guarantee its success when applied to another.

2.6 Conclusion

We have presented a review of the literature pertaining to DRR–based 2D/3D bone reconstruction from X–ray images using statistical shape and intensity models. This was completed using unambiguous terminology, and a unified mathematical formulation of the problem, in a common conceptual framework. We have also provided a discussion of the shortcomings, recent adaptations and persisting challenges of this approach along with insights for future research.

Ultimately, 2D/3D bone reconstruction methods have the potential to enable clinical procedures that normally require access to CT and MRI machines and would otherwise be impossible to perform given only a patient's X–ray images. This is especially pertinent to resource–limited settings where access to these 3D imaging technologies may not be possible. The option of a relatively cheap and more easily accessible compromise will have a beneficial impact. In addition, by minimizing a patient's exposure to ionizing radiation (compared with CT) their implicit risk of cancer can be reduced. Imaging artefacts that are caused by the presence of metal implants can also be avoided. The reconstruction of 3D patient–specific bone models from X–ray images is therefore an important and powerful technique in medical imaging, and is worthy of future research efforts.

Chapter 3

A volume rendering pipeline for simulating X-ray images

3.1 Introduction

In the previous chapter, an extensive literature review in the field of 2D-3D reconstruction was presented using a common conceptual framework and unified mathematical formulation of the problem. X-ray simulation, which is one of the modules presented within the framework, will be the focus of this chapter. X-ray image simulation is the process of generating an artificial X-ray image, also referred to as a digitally reconstructed radiograph (DRR), and is usually derived from a computed tomography (CT) voxel volume. As discussed in the Literature Review, there are advantages in adopting a DRR-based approach to 2D-3D reconstruction. However, a number of challenges which have attracted much less attention in the literature than other aspects of the reconstruction algorithm, are still encountered. Parameters which affect the geometry of the 2D projection, such as the distance from the X-ray source to the photon detector, vary from one X-ray machine to another. Moreover, some X-ray machines have unusual ray geometry, which produce significantly different projections than the conventional "point-source" setup. The field of computer graphics offers a useful framework for addressing these often-overlooked challenges. However, employing a volume rendering pipeline in this context necessitates that it be adapted to the unique characteristics of X-ray machines.

This chapter will start by discussing the 2D-3D reconstruction-specific design requirements for the DRR renderer. Next, the "volume rendering framework" will be introduced and modified to focus on the steps comprising the "volume ray-casting" method. Each step will then be revisited while describing the adaptations necessary to meet the design requirements.

Next, details regarding the software implementation of the adapted volume renderer together with a number of visual examples will be presented, followed by a discussion of the work done.

3.2 Design requirements

Before a standard volume renderer was adapted for the purpose of simulating X-ray images, and for the purpose of 2D-3D reconstruction, the necessary design requirements were first analysed, as follows:

- The accuracy of geometric projections from 3D to a 2D plane is of utmost importance as this has a direct influence on the possible reconstruction accuracy. In order to ensure a high degree of accuracy, the imaging setup of unique X-ray machines must be closely adhered to.
- Commonly used projection geometries (perspective, parallel, fan-beam) and aggregation methods (DRR, silhouette, maximum-intensity-projection, Hounsfield Unit-thresholding) should be provided, and easily configured.
- Machine-specific parameters, including source-to-object (S2O) distance, source-to-detector (S2D) distance, pixel-spacing, image-width, image-height, contrast-enhancement window and contrast-enhancement level, should also be easily configured.
- Since the aforementioned parameters are readily accessible in the standard "DICOM" format used by X-ray machine manufacturers, they should also be automatically extracted and used to calibrate the DRR renderer.
- The X-ray image (pixel data) should be automatically loaded into the renderer, which should provide the ability to switch the rendering mode between the DRR itself, and a difference image of the DRR and the target X-ray image.
- Real-time interaction with large CT volumes is a requirement for visual exploration, and for performing manual 2D-3D registration. Therefore, a parallelised implementation of the renderer should be used in order to leverage hardware-based acceleration (GPU) and increase the rendering frame-rate above 30 frames-per-second.
- The renderer should provide the ability to specify its sampling rate, the reduction of which makes real-time interaction possible even when working with very large volumes.
- The DRR renderer should support the commonly used "Nifty" format for CT volumes.

3.3 Volume rendering basics

Volume rendering concerns the visualisation of 3D scalar fields on a 2D image plane [170]. In order to depict all important information in a single 2D projection, data is rendered in a semi-transparent fashion using some kind of optical model for light transport. The commonly-used emission-absorption model [134] computes the amount of radiative energy $I(t)$, and its derivative with respect to length t, using the following differential equation:

$$\frac{dI(t)}{dt} = g(t) - \tau(t)I(t) \tag{3.1}$$

where g(t) is the emission of light from the participating medium (the volumetric data to be rendered), and the "extinction coefficient" $\tau(t)$, which defines the rate at which light is attenuated by the medium. The extinction coefficient, together with radiance I(t), forms the "absorption term". The radiance which exits the volume at point $t = D$ and arrives at the detector is determined by integrating Equation 3.1 from the point $t = t_0$ at which the incident radiance I_0 enters to point $t = D$:

$$I(D) = I_0 \exp(-\int_{t_0}^{D} \tau(t')dt') + \int_{t_0}^{D} g(t) \exp(-\int_{t}^{D} \tau(t')dt')dt. \tag{3.2}$$

A simple approach [170] to evaluating the integral is using an equidistant sampling scheme, equivalent to a Riemann-sum approximation of the volume rendering integral:

$$I(D) \approx I_0 \prod_{i=1}^{n} e^{-\tau(i\Delta p + t_0)\Delta p} + \sum_{i=1}^{n} g(i\Delta p + t_0)\Delta p \prod_{j=i+1}^{n} e^{-\tau(i\Delta p + t_0)\Delta p} \tag{3.3}$$

over n partitions of length $\Delta p = (D - t_0)/n$. The source term then becomes

$$g_i = g(i\Delta p + t_0)\Delta p, \tag{3.4}$$

and the extinction coefficient becomes

$$t_i = e^{-\tau(i\Delta p + t_0)\Delta p}. \tag{3.5}$$

The absorption-emission model does not account for scatter-radiation. More advanced optical models have been explored, which incorporate scatter-radiation, and which produce more realistic renderings [5, 172], but are also more computationally intensive, and thus impede the ability of a renderer to perform in real-time. In order to evaluate the aforementioned volume rendering equation (emission-absorption model), the process is conceptualised as a

sequence of subsequent steps referred to as the volume rendering pipeline: data traversal, interpolation, gradient classification, shading, and compositing [170]. While the field of computer graphics describes a number of techniques for completing these steps[1], the configuration of steps known as "volume ray-casting" is best suited to the task of X-ray simulation. Conceptually, however, these steps do not map in a one-to-one fashion from the broader process of volume rendering to the more specific process of volume ray-casting. Moreover, certain aspects of the process are either irrelevant or require modification for the purposes of X-ray simulation. Before providing an overview of previous work, the conceptual framework was first modified in order to better contextualise concepts relevant to X-ray simulation.

First, the approach to data traversal was easier to conceptualise as two separate steps: casting and sampling, where X-ray-specific considerations are essential for the former but not for the latter. In contrast, since the interpolation step did not involve any non-trivial operations, it was unnecessary for it to be considered separately, but rather as part of the sampling step. Finally, gradient classification is used when modelling surface-based light reflections. It is what distinguishes other volume rendering methods such as "ray-tracing" from ray-casting [170] and was therefore excluded. Figure 3.1 provides an illustration of the revised volume rendering pipeline, which comprises the following steps: casting, sampling, shading, and compositing. What follows is a more detailed description of each step.

3.4 Adapting a standard volume ray-caster for simulating X-ray images

The previously described volume rendering paradigm must be adapted in order to adhere to the unique requirements of X-ray imaging and can thus be made to more accurately simulate X-ray images. In this section, each of the steps comprising the revised volume rendering pipeline are described in detail along with any novel adaptions that were required for X-ray simulation. These adaptions include: Machine-specific ray geometry, stochastic sampling, Hounsfield Unit conversion as shading, and aggregation based on an absorption-only model (Beer-Lambert law). Finally, while hardware-based acceleration is not itself a novelty, the unique parallelisation scheme that was employed is described.

[1] These steps can also be categorised as image-order and object-order methods, as described in Chapter 2.4.2.

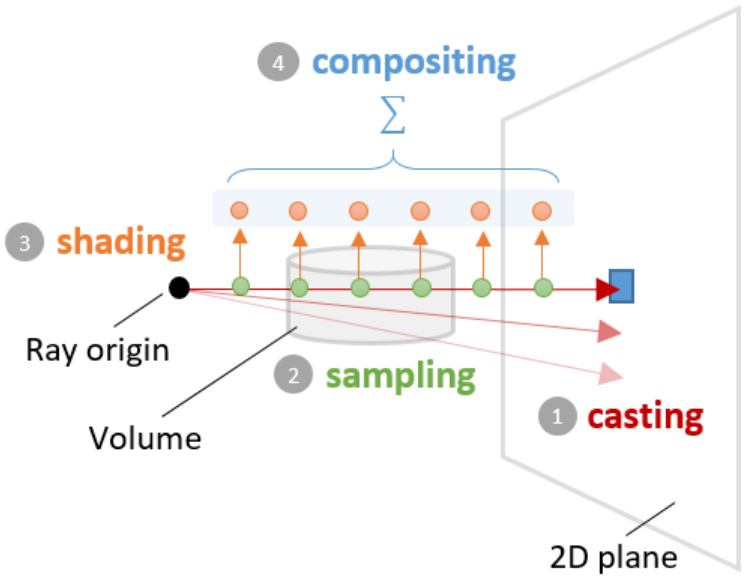


Figure 3.1 A simplified overview of the steps comprising the "volume ray-casting" rendering pipeline. The first step, "casting", defines the ray geometry. The second step, "sampling", defines how samples are selected along the defined ray paths. The third step, "shading" involves the mapping of the sampled values using a look-up-table. The fourth and final step, "compositing" describes how the mapped values are aggregated to obtained a final pixel intensity value.

3.4.1 Ray-casting using specialised geometry

During the process of volume ray-casting, a single ray is "cast" for each pixel on a predefined 2D grid. Each ray starts at the camera and traverses through the volumetric data to be rendered in steps while collecting colour samples along the way using Equation 3.20.

Thus, the 3D sample-point coordinates p_i for the i^{th} step can be computed using the vector form of the straight-line equation:

$$p_i = o + i\hat{r} \tag{3.6}$$

where o is the ray-origin and $\hat{r}$ is a normalised ray-direction vector. The simplicity of this approach allows it to be easily parallelised as the subsequent steps in the pipeline are identical for each ray that is modelled.

It is next necessary to define the set of ray-origin(s) and ray-directions for all the pixels of the rendered image according to some projection scheme such as perspective (rays diverging from a single point), or orthographic projection (rays parallel to one another). Furthermore, the volumetric data has to be placed in the coordinate system of the defined rays. There is

a number of considerations to be made when doing so. First, there are typically multiple objects in the scene to render. Furthermore, for many applications the ability to reorient objects (for example, objects moving within the scene) as well as the camera (moving around the scene from a first-person viewpoint) is a requirement. In a typical volume rendering pipeline, the computer graphics concept of the model-view-projection (MVP) matrix is used to efficiently reconcile the separate coordinate systems of the objects to be rendered, the scene (the shared global 3D Cartesian coordinate system also referred to as the "world"), and the 2D plane corresponding to the output pixels which are rendered (also referred to as the "camera"). The reference frame of the voxels comprising each object is first transformed to a common one - referred to as "world space" - using a corresponding model matrix for each of the objects. When adjusting the pose of an object, identical operations are performed on all the voxels that comprise it. However, if multiple pose adjustments are performed (as is the case in an iterative registration) rounding errors may accumulate, due to the floating-point arithmetic[2] used by computers, and may result in undesired volume deformations. A matrix is therefore used as an efficient way to keep track of object-pose adjustments; the original voxel volume coordinates are preserved, while a record of the resultant translation and rotation deltas[3] is stored in a matrix. A model matrix M therefore comprises an object's translation transform T, rotation transform R, and scale transform S:

$$\begin{aligned} M &= T \cdot R \cdot S \\ p_{world} &= M \cdot p_{model} \end{aligned} \tag{3.7}$$

Multiplying a point-coordinate vector p by this model matrix transforms it into world space. The transformed coordinates are subsequently transformed again to an assumed viewpoint in "camera space" - using a view matrix, which is simply the inverse of the camera's model matrix[4]:

$$\begin{aligned} V &= C^{-1} \\ p_{camera} &= V \cdot M \cdot p \end{aligned} \tag{3.8}$$

[2]Since computers have a finite storage capacity, they use floating point numbers instead of real numbers. A floating-point representation (essentially scientific notation) makes it possible to represent and perform computations on a wide range of magnitudes. However, errors may be introduced each time a floating-point operation is performed, since the amount of memory set aside for the result may be insufficient.

[3]Euler angles do not provide a mathematically complete description of the object orientation, since certain coordinate configurations can cause the degrees-of-freedom to fall below 3. Quaternions are often used instead because they provide a coordinate system with 4 degrees of freedom. In addition, they have other useful mathematical properties; they require fewer operations - which is especially desirable for reducing computation time - and also make it easier to subdivide larger rotations into smaller ones - enabling smoother rotations when used in an interactive renderer.

[4]Thus, strictly speaking, the scene moves around the camera, rather than vice versa.

Finally, a projection onto a 2D plane residing in "homogeneous space" is performed (where the camera/detector-pixels are located) using a projection matrix. The projection matrix alone specifies the type of projection (orthogonal, perspective, or fan-beam) and also checks which voxel coordinates are visible according to the detector (camera) parameters, thus avoiding unnecessary ray-casting operations. Taken together, this series of operations can be combined into a single matrix - referred to as the MVP matrix (see Figure 3.2):

$$p_{homogeneous} = P \cdot V \cdot M \cdot p \tag{3.9}$$

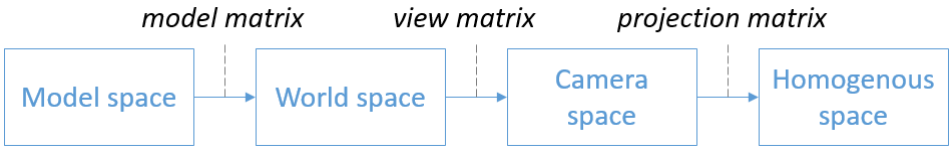


Figure 3.2 An illustration of how model, view, and projection matrices are used to transform points or voxels between model, world, camera, and homogeneous spaces.

For the purposes of the DRR renderer, the camera (X-ray detector) remains stationery. The camera-space can therefore be considered the same as the world-space, and thus makes the view matrix unnecessary. Furthermore, by defining the projected rays in this same space, rather than separately in a projection matrix, only a single model matrix has to be tracked for each new frame that the DRR renderer outputs. Unnecessary computations can instead be avoided using the ray-box intersection technique, which disregards rays that do not intersect the bounding-box of the volume [75]. The desired projection geometry can be easily switched (simply by changing the definition of the ray-origin) and computed in an online fashion, rather than having to keep track of different projection matrices.

$$\bar{o} = \begin{cases} (0, 0, -d_{S2O}) & \text{if perspective} \\ (p_{detector,x}, p_{detector,y}, -d_{S2O}) & \text{if parallel} \\ (0, p_{detector,y}, -d_{S2O}) & \text{if fan-beam} \end{cases} \tag{3.10}$$

Taking this approach also resolves a key difference between the conventional ray geometry used in ray-casting and the kind required to properly model an X-ray machine, which is that incoming rays are convergent for the former, and divergent for the latter, as shown in Figure 3.4. Thus, for classic ray-casting, the further away an object is positioned from the location of the detector, the more the ray geometry will approximate an orthographic (parallel) projection, while the inverse is true for an X-ray machine. Finally, the specialised case of "fan-beam" projection should also be considered. Low-dose X-ray machines, such as

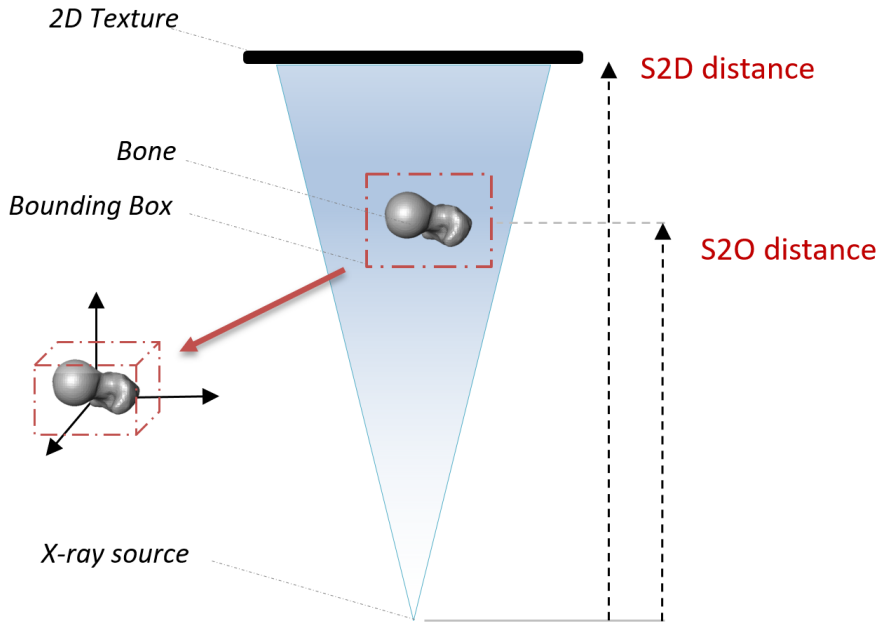


Figure 3.3 Source-to-object distance (d_{S2O}), source-to-detector distance (d_{S2D}).

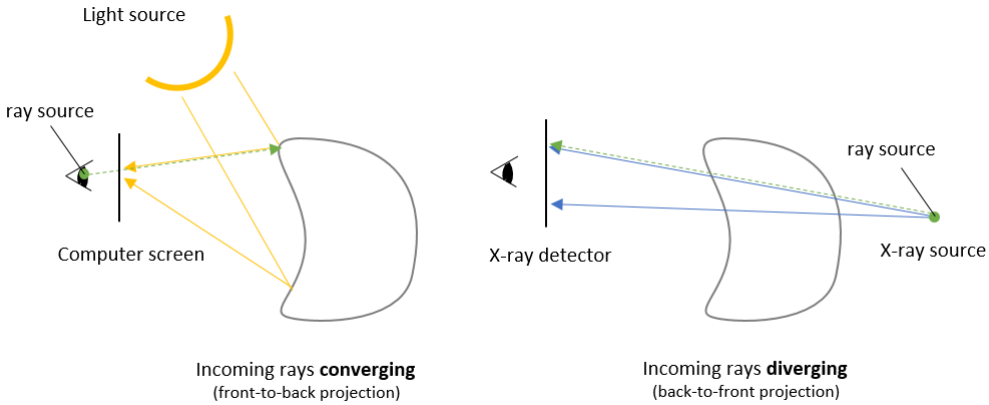


Figure 3.4 Incoming rays converge on the computer screen for a classic ray-casting model, whereas X-rays diverge from a point-source towards the X-ray detector.

the Lodox Statscan [93] and the EOS X-ray imaging system [48] employ fan-beam geometry to reduce the X-ray exposure area, thus reducing the dose of ionising radiation a patient receives. First, a collimator[5] is used to allow only a narrow strip of X-rays to exit the X-ray source, and second, using a gantry[6] to fire the X-rays along a parallel scan direction.

[5] A collimator is used to focus a beam of particles or waves which - in the case of X-ray imaging - reduces the cross-section of the projection on the image plane.

[6] In a linear slot-scanning X-ray machine, the gantry contains electronics such as the X-ray tube and detector in a housing that moves along a bridge-like frame in a fixed path during image acquisition.

Fan-beam ray geometry can therefore be modelled using rays that are parallel in one axis direction (corresponding to the scan-direction), and perspective in the other (see Figure 3.5).

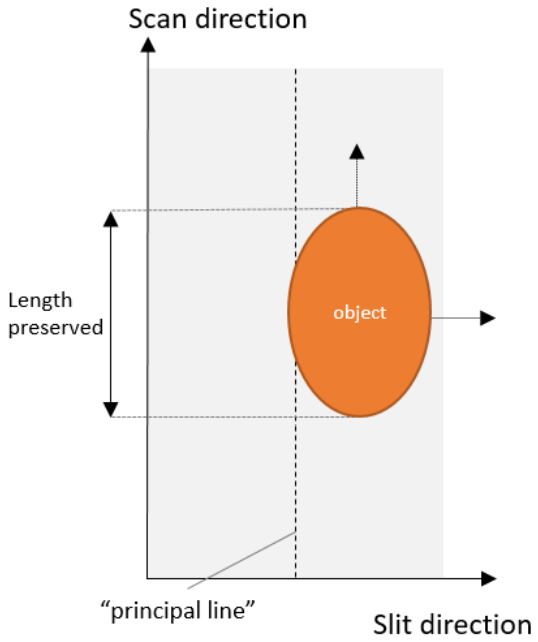


Figure 3.5 Direction-dependent distortion resulting from fan-beam ray geometry (a top-down view is shown). Object-length is preserved along the y-axis (scan-direction), while for the x-axis (slot-direction), features are distorted in proportion to their distance from the "principal-plane" (located along the x-axis).

3.4.2 Stochastic sampling

Next, the volumetric data is sampled at positions along a given ray, for which a variety of ray-traversal methods have been proposed (see [171] and [64]). A simple approach is to aggregate each voxel encountered along the ray, where the set of encountered voxels can be efficiently determined using the Bresenham line algorithm [21]. However, this strategy introduces bias because, in some cases, perspective geometry may allow more than one adjacent voxel to be encountered simultaneously (see Figure 2.5 in Chapter 2). It is therefore preferable to take equidistant samples along the ray. However, sample positions do not always lie perfectly on voxel-grid locations, and simply using nearest-neighbouring voxels introduces small biases which can accumulate and distort DRR quality. Furthermore, aliasing can occur in regions where the number of casted rays (pixels) is greater than the available voxel resolution; when there are large differences between neighbouring voxels, the result

appears as a step-like pattern - commonly referred to as "jaggies". Voxel-values are therefore typically interpolated[7] in order to overcome these pitfalls.

While interpolation between neighbouring voxels can suppress the appearance of jaggies, it does not prevent another type of aliasing, referred to as Moiré patterns. Since ray-casting is a point-sampling process, it is assumed that casted rays are infinitely thin, despite the fact that pixels have a finite width. To overcome this incongruency, the accumulated intensity of a single ray, which is incident at the centre of each pixel, is considered to be the intensity of the entire pixel [37]. However, because pixels occur at regular intervals, frequency-based aliasing can occur, the effect of which is observed as Moiré patterns (see Figure 3.6). One way to avoid these patterns is with stochastic sampling where every point is assigned a finite probability of being hit [36]. In doing so, Moiré patterns are substituted for noise, which the human eye is less sensitive to (see Figure 3.6).

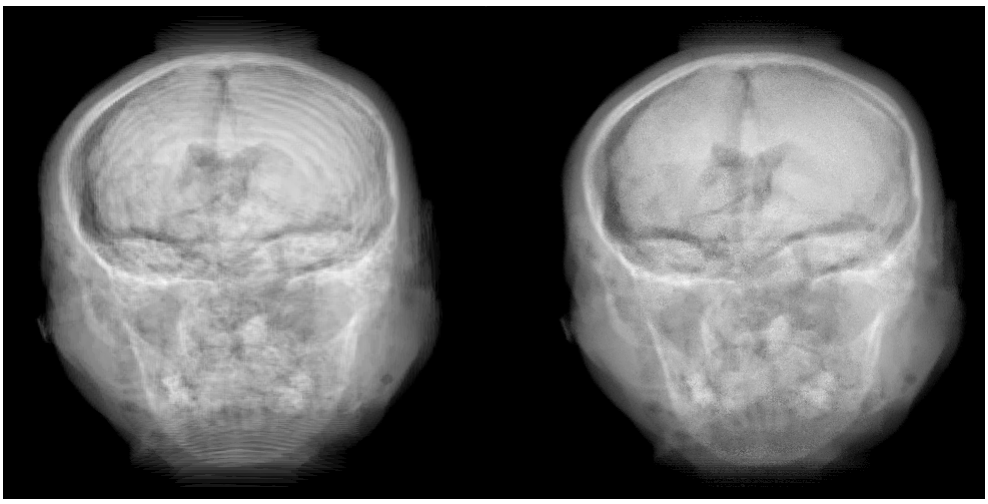


Figure 3.6 Two DRRs are shown, generated using identical parameters, except for sampling strategy. Regular sampling (shown on the left) causes Moire patterns, which can be substituted for noise resulting by using stochastic sampling instead (shown on the right).

Standard trilinear interpolation was used for the adapted volume renderer, in order to overcome jaggies. However, instead of the conventional stochastic sampling strategy - which assigns a finite probability of collision to every point - a slightly different approach was followed. In doing so, the goal of mitigating Moiré patterns is still achieved but with increased computational efficiency, as a random perturbation is made for each ray rather than for each sample point:

$$o' = o + d\hat{r} \tag{3.11}$$

[7] Bilinear and trilinear interpolation is easily performed on the GPU by setting a flag, and is sufficient for the requirements of volume ray-casting.

where o is the ray-origin and $\hat{r}$ is a normalised direction vector. The perturbation d is made to be much smaller (three orders of magnitude) than the ray-traversal step-size.

3.4.3 Shading for Hounsfield Unit conversion

Shading is conventionally used to assign a material colour, represented by four channels, corresponding to red, blue, green, and alpha, (RGBA) to a sample-point dependent on its corresponding, or desired, optical properties. A transfer function, t, maps a voxel value v at point coordinate p, to a new value:

$$g(p) = t(v(p)). \tag{3.12}$$

For example, the source term can be augmented to include a local illumination model, which provides information about one or many light sources within a scene (see Equation 3.20). The extinction coefficient can be augmented such that it assigns increasing transparency to materials the further away a voxel is from the back of the volume (in relation to the camera). Thus, more information can be visualised simultaneously despite the true material properties of a volume. Transfer functions have been studied for a wide variety of applications; for an extensive survey, the reader is referred to [92].

Conventionally, shading is used to render material colours based on optical properties such as ambient light. However, for X-ray imaging, rendered pixels do not represent material colours, but instead are assigned an intensity based on material density. Computed tomography volumes comprise voxels - each with a corresponding Hounsfield radiodensity value, h_n. A CT Hounsfield value is an integer ranging between -1000 and 3000, in Hounsfield Units (HU). For X-ray simulation, each Hounsfield value has to be converted to a Linear Attenuation Coefficient (LAC), μ_n (in cm^{-1}), providing a normalised measure of how absorptive a voxel's corresponding material is [8]:

$$h_n = \frac{(\mu_n - \mu_w) \times 10^3}{\mu_w} \tag{3.13}$$

Computed tomography machines are typically calibrated so that $\mu_{water} = 0.25$, causing the Hounsfield value of water to correspond to 0, and the Hounsfield value of air to correspond to -1000 [8].

For the adapted volume renderer, the conversion was performed as an adaption of the shading step (see Equation 3.16), using a simple 1D transfer function to output a LAC value, rather than a RGBA tuple:

$$t(h_i) = \mu_w(h_i \times 10^3 + 1) \tag{3.14}$$

The transfer function can also act as a thresholding function, so that only certain types of tissue are rendered:

$$t(h_i) = \begin{cases} \mu_w(h_i \times 10^3 + 1) & \text{if } h_i \geq h_t \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

where h_t is the threshold Hounsfield value. Such a strategy is effective when performing a registration in stages, or when registering a CT-volume that has soft tissues present. Alternatively, transfer functions can be used to perform contrast enhancement within specific value ranges (see Figure 3.7).

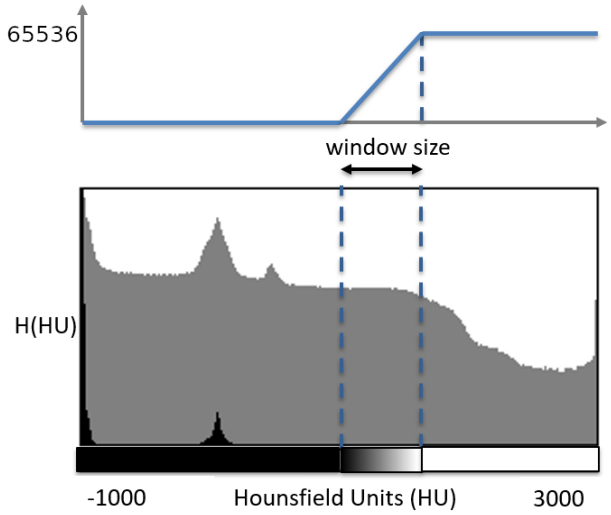


Figure 3.7 The window/level look-up-table is a linear scaling function for contrast enhancement. The original Hounsfield values of a CT volume are shown below (standard histogram in grey, logarithmic histogram in black) while the new values are shown above (as 16-bit grey-levels). Values below the window are mapped to black while values above the window are mapped to white. Values within the window are scaled uniformly.

3.4.4 Compositing using the Beer-Lambert law

Next, the process known as compositing is used to iteratively aggregate the shaded samples and thus compute a pixel's colour and alpha[8].

There are two approaches to compositing: back-to-front and front-to-back compositing. While the ordering does not affect the appearance of the final image, back-to-front compositing - where the ray is traversed from the back of the volume to the camera - is preferred when using graphics hardware because it avoids the additional overhead of keeping track

[8] In the computer graphics community, opacity, the complement of transparency, is often referred to as alpha.

of accumulated alpha values. Front-to-back compositing is preferred for software-based rendering because additional processing can be avoided once the accumulated transparency falls below a pre-defined threshold, thus making it faster to compute [139, 170].

The back-to-front iteration equation is as follows:

$$C_{destination} \leftarrow t_{source} C_{destination} + C_{source} \tag{3.16}$$

where C_{source} and transparency t_{source} are assigned a value by a transfer function which maps the values based on the optical properties at the current location on the ray. The output colour for the corresponding camera pixel, $C_{destination}$, is accumulated through successive updates as the ray is traversed.

Conventionally, the discretized volume rendering integral is computed iteratively through the process of compositing (see Equation 3.20). When modelling X-rays it is assumed that the participating medium[9] is made up of cold, perfectly black particles, which do not emit or scatter any light, but instead absorb all the light (X-rays) it intercepts [104]. The absorption-only model is therefore defined as

$$\frac{dI(t)}{dt} = -\tau(t)I(t). \tag{3.17}$$

The radiance which exits the volume at point $t = D$ and arrives at the detector is determined by integrating Equation 3.17 from the point $t = t_0$ at which the incident radiance I_0 enters to point $t = D$:

$$I(D) = I_0 \exp(-\int_{t_0}^{D} \tau(t')dt') = I_0 T(t) \tag{3.18}$$

where $T(t)$ is defined as transparency [104]. The Beer-Lambert law, which describes the attenuation of light through matter, is then used to define the transparency in terms of linear attenuation coefficients:

$$T(t) = \exp(-\int_{t_0}^{D} \tau(t')dt') = \exp(-\int_{t_0}^{D} \mu(t)dt) \tag{3.19}$$

For ray-casting, an equidistant sampling scheme is used to evaluate the integral, which is equivalent to a Riemann-sum approximation of the volume rendering integral, over n segments of length $l = (D - t_0)/n$:

$$I(D) \approx I_0 \exp(-\sum_{i=1}^{n} \mu l) \tag{3.20}$$

[9] the volumetric data to be rendered - located between the energy source and the detector.

This simplified evaluation is independent of the order of ray-traversal, and therefore avoids the additional complexity which would otherwise be necessary [104]. Using the Beer-Lambert law to aggregate the sampled LACs encountered along the ray is similar to taking an average. However, other types of aggregation can also be used, such as a maximum-intensity-projection or silhouette (see Section 3.6).

Once the sampled values along each ray have been composited and outputted to the pixels comprising the modelled detector, the result is a DRR. In radiology, it is typical to perform a post-processing step for contrast enhancement, which is incorporated into modern X-ray viewing software as standard, and for which values are reserved in the DICOM specification. The DRR contrast enhancement method that was used for the adapted volume renderer is identical to the one described in Figure 3.7 except that the look-up-table is employed for the pixels in DRR rather than for CT voxels.

3.4.5 Continuous rendering for visual inspection

One of the primary advantages of the volume rendering framework is that it allows pose and projection parameters to be modified "on-the-fly" in an interactive fashion with ease. This ability is useful not only for gaining an intuition for X-ray imaging, but also when manually performing a 2D-3D registration through visual inspection. In order to provide an interactive mode for the renderer, the aforementioned algorithm is placed in a loop, iterating at a frequency of > 20 Hz in order to provide a smooth experience and not distract the user [11, 24, 25]. For each iteration, the keyboard and mouse is queried for input (if any is entered) and the model matrix is modified accordingly before another kernel pass is made. Other hyper-parameters such as the contrast-enhancement level and window can also be dynamically modified to further aid visual inspection. The loop continues indefinitely, until an exit operation is triggered.

3.5 Implementation details

The step-by-step procedure for simulating an X-ray image is described in this section. To illustrate the procedure, the scenario where a single X-ray image, and single CT volume are used to perform a manual 2D-3D alignment, is considered.

It is assumed that the CT data is provided in the commonly used "Nifty" format, while the X-ray data is provided in the clinical-standard DICOM format. Raw 2D X-ray pixel data as well as relevant X-ray machine parameters are automatically extracted from the DICOM file. When compared to CPUs, graphics processing units (GPUs) have far more arithmetic

logic units and therefore excel at executing parallel computations. The ray-casting algorithm lends itself to a GPU implementation, because it is easily parallelised; the evaluations of each ray integral are independent of one another and therefore a GPU thread can be assigned to each pixel. Even an unsophisticated parallelisation scheme provides a significant reduction in computation time when compared to a CPU-only approach. Specialised GPU memory objects are prepared (see Step 1 of Figure 3.8) which store the CT volume, target X-ray image, output 2D image, model matrix, voxel-volume dimensions (required in order to determine ray projection geometry correctly), and transfer function, as well as hyperparameters, which include the source-to-object distance (d_{S2O}), source-to-detector distance (d_{S2D}), sampling-rate, LAC of water μ_{water}, incident intensity I_0, and projection-type. These memory objects are then passed to the GPU, which executes an identical kernel program for each thread in parallel, before being collated and passed back to the CPU. The number of threads is set according to the number of pixels in the image. The GPU workflow/pipeline is shown in Step 2 of Figure 3.8.

The identical set of computations executed by the kernel program for each thread, are as follows. The pixel index assigned to the current GPU thread is retrieved and used to determine the detector pixel coordinates $\bar{p}_{det}$. The detector pixel coordinates, together with the X-ray source coordinates, $\bar{o}$, are then used to compute the ray-direction vector, $\hat{r}$. The X-ray source coordinates are determined depending on the X-ray projection geometry specified by the corresponding input hyper-parameter; perspective geometry: $\bar{o} = (0, 0, -d_{S2O})$, parallel geometry: $\bar{o} = (p_{det,x}, p_{det,y}, -d_{S2O})$, fan-beam geometry: $\bar{o} = (0, p_{det,y}, -d_{S2O})$. Furthermore, X-ray source coordinates are offset by a small random amount along the ray-direction vector (stochastic sampling) to avoid Moiré patterns. The 3D grid of scalar values corresponding to the CT data is mapped into the coordinate system of the X-ray machine (to which the X-ray source and detector pixel were referenced) using the model-matrix. The intersection of the ray-direction vector with the bounding box of the CT volume is tested to determined whether or not to proceed with the ray-traversal; if no intersection is determined, the pixel value corresponding to the DRR is assigned an intensity of zero. Next, the step-size is determined such that the number of samples is proportional to the voxel-spacing of the volume, as defined by the sampling rate multiplier. The 3D coordinate of the sample is then determined using Equation 3.6, The corresponding HU value(s) is determined and interpolated (if necessary) on the voxel-grid of volume. Equation 3.15 is then used to determine the corresponding LAC for the HU value, which is multiplied by the step-size. The result is added to the running sum. The sum, together with the I_0 hyperparameter, is then used to compute the DRR pixel intensity using Equation 3.20. The contrast-enhanced pixel value is subsequently computed, using the contrast enhancement window and level hyperparameters (previously extracted

from the DICOM file). If a target X-ray image was provided, the target pixel intensity value for the 2D coordinate corresponding to the current pixel index is determined. The DRR pixel intensity value is subtracted from it and marked as the kernel output for the current GPU thread. Finally, the output pixel value for each GPU thread is obtained and collated into a 2D GPU memory object before being transferred back to CPU memory, as shown in Step 3 of Figure 3.8. Alternatively, if the renderer is used interactively - as is the case for the task of manual 2D-3D alignment - the 2D GPU memory object is successively updated and rendered, while accepting input from the mouse and keyboard.

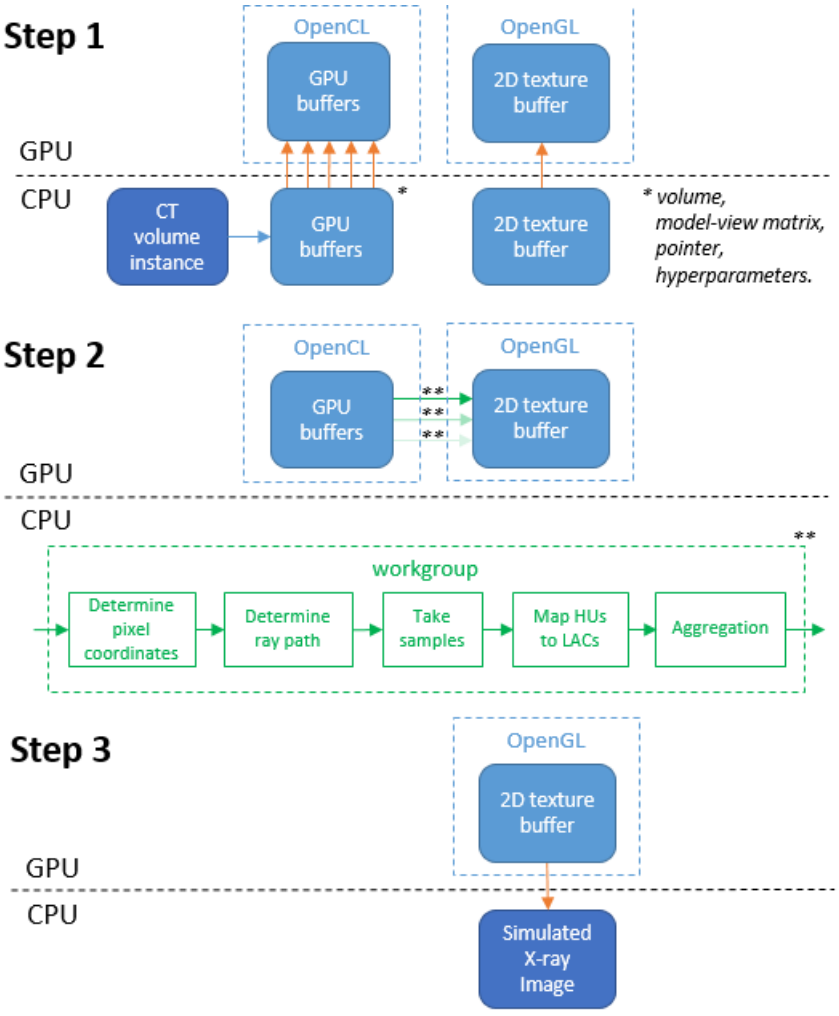


Figure 3.8 The hardware-accelerated X-ray simulation process that was implemented to perform hardware accelerated X-ray simulation comprises three steps. First, special buffers are passed to a GPU kernel program. Second, the kernel executes an identical set of computations for each parallel stream (workgroup), before finally collating the results and passing them back to the CPU. The number of streams used are decided beforehand, which, in the case of the ray-casting is set to be equal to the number of pixels in the image.

3.5.1 Application Programming Interface

Ultimately, the DRR renderer takes the form of an application programming interface (API), which provides a set of objects and methods for rendering simulated X-ray images. Table 3.1 provides a list of important specifiable parameters and methods offered by the API.

3.5.2 Software frameworks

The core of the renderer was programmed with Scala, version 2.11 [119]. For GPU-related operations, OpenGL version 1.1 [140], and OpenCL version 1.2 [77] were used. A small Java-based toolbox was programmed to perform rotation, translation and scaling operations using quaternions, as well as for converting between the quaternion and standard matrix representations.

Keyboard and mouse input was handled using the Lightweight Java Gaming Library (LWJGL) framework, version 2.9 [90], and also provided wrappers for OpenCL and OpenGL (originally programmed in C). Thus, an “interactive mode" for the renderer was made possible, where a user is able to manually adjust object-pose, and toggle various settings such as sampling rate and ray geometry.

The Amira suite [158], version 5.4.3, was used to load computed tomography (CT) slices, in the DICOM format, obtain a voxel grid, resample it to a 1x1x1 voxel spacing using a Lanczos filter[10] [44] (if necessary), and save it as a "Nifty" data structure. The Scalismo framework[11][159], version 0.17, was then used to load the CT nifty volume for further use, using the Scala programming language.

The ImageJ software toolbox, version 1.54b [137], was used to load the "target" X-ray image, and extract the relevant X-ray machine parameters, from the standard DICOM formatted file. The image was converted to a four-channel "PixelImage" format of the Scalismo framework.

3.6 Qualitative Results

In this section, visual examples of output from the adapted volume renderer are provided in order to demonstrate its functionality. First, the renderer is capable of accommodating three

[10]The Lanczos filter is known to yield competitive results for both magnification and minification resampling operations. However, while avoiding blurring effects, the Lanczos filter is more prone to produce excessive "ringing" effects.

[11]The Scalismo framework is a software library specialised in statistical shape modeling and model-based image analysis. The framework is programmed in Scala and is developed and maintained by the Graphics and Vision Research Group (GRAVIS) at the University of Basel.

Table 3.1 Parameters provided by the DRR Renderer API.

Parameter	Data-type	Description
volume	3D array of 16-bit integer values	The CT volume.
targetImage (optional)	2D array of 32-bit Floating-point values	The target X-ray image.
intensity	32-bit Floating-point	Incident energy of the modelled X-ray, I_0.
acWater	32-bit Floating-point	Linear attenuation coefficient of water, μ_{water}.
samplingRate	16-bit integer	Rate at which samples are taken along each ray (as a multiple of the voxel-length).
distanceS2O	32-bit Floating-point	Distance from the source (ray-origin) to the object (CT data) to be imaged.
distanceS2D	32-bit Floating-point	Distance from the source (ray-origin) to the detector.
roll	32-bit Floating-point	Side-to-side tilt (in degrees) on the x-axis.
pitch	32-bit Floating-point	Backward-forward tilt (in degrees) on the y-axis.
yaw	32-bit Floating-point	Left-right turn (in degrees) on the z-axis.
translationX	32-bit Floating-point	Left-right movement (in degrees) along the x-axis.
translationY	32-bit Floating-point	Up-down movement (in degrees) along the y-axis.
scale	32-bit Floating-point	Scaling of the simulated X-ray image (default value normalises the scaling to the X or Y dimension, depending on which is larger).
detectorWidth	16-bit integer	Width of the simulated X-ray image in pixels.
detectorHeight	16-bit integer	Height of the simulated X-ray image in pixels.
projectionType	String flag: "PERSPECTIVE", "PARALLEL", "FANBEAM"	Perspective, Parallel, or Fan-beam.
hounsfieldUnitThreshold	16-bit integer	The threshold (in HU) for which voxels encountered along the ray have a smaller value, are excluded.
diffImage	Boolean	Whether to show the difference image.
interactive	Boolean	Whether to turn engage "interactive mode" of the renderer.
renderMode	String flag: "DRR", "MIP", "SIL"	Aggregation method to use: standard, maximum-intensity-projection or silhouette.
offscreenRendering	Boolean	Whether to turn engage the "offscreen-renderering mode" of the renderer (overrides the interactive mode).
contrastEnhancement	Boolean	Whether to use window-level contrast enhancement.
contrastEnhancementLevel	16-bit integer	Pixel intensity value (16-bit) for contrast enhancement "level".
contrastEnhancementWindow	16-bit integer	Pixel intensity value (16-bit) for for contrast enhancement "window".

different projection geometries used by X-ray machines, namely perspective, orthographic (parallel), and fan-beam, as shown in Figure 3.9.

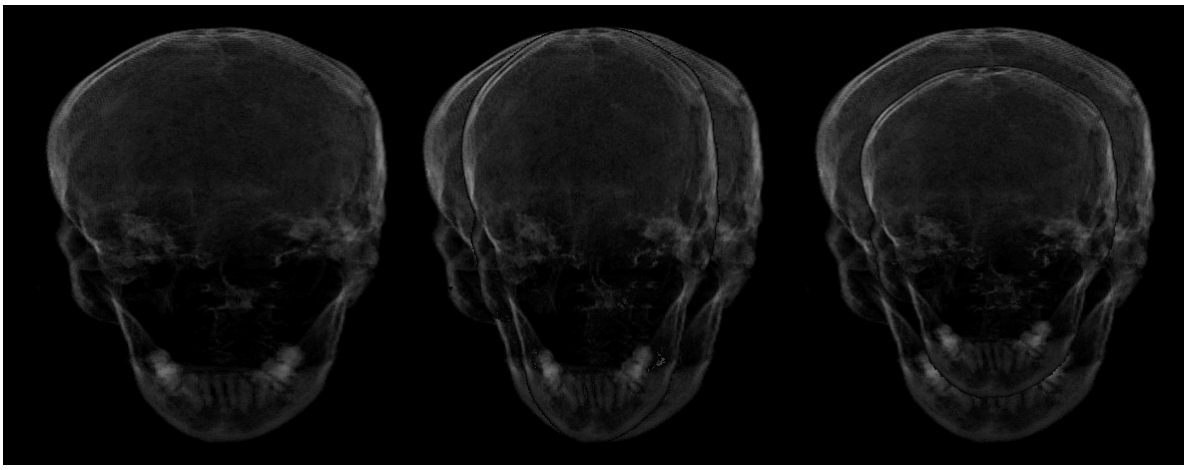


Figure 3.9 Three pairs of simulated X-ray images (DRRs) of the human skull corresponding to three different projection geometries, orthographic/parallel (left), fan-beam (middle), perspective (right). Each pair contains a projection overlaid onto another with a larger S2O distance in order to emphasise the characteristic way in which each image is warped. Parallel projection does not change as a function of S2O distance, while perspective projection approximates parallel projection as the S2O increases. Fan-beam projection provides parallel projection in the scan-direction (y-axis), and perspective projection in the slot-direction (x-axis).

The renderer readily models different X-ray machines using information accompanying the raw X-ray pixel data in the DICOM format. Figure 3.10 shows DRRs obtained while modelling the Lodox Statscan X-ray system. The system employs fan-beam geometry. This type of geometry was introduced to reduce the amount of ionising radiation a patient receives, and is rarely encountered outside the field of X-ray imaging. Furthermore, the imaged object is a human skull, which is particularly complex, having a large number of cavities and interconnected macro- and micro-structures. While, from a global perspective, there seems to be a significant agreement between the real and simulated X-ray images shown in Figure 3.10, at a local level, many discrepancies are apparent. Aside from those resulting from errors in the manual registration performed to align the projections, and artefacts resulting from the presence of metallic dental implants, discrepancies are predominantly due to the unstable pose of the jaw-bone and soft tissue, which can shift during the imaging protocol. Nonetheless, Figure 3.10 demonstrates that the DRR renderer can produce visually similar simulated X-ray images when dealing with bone structures as complex as the skull[12].

[12]Extensive testing on simulated X-ray images of the skull is presented in [17].

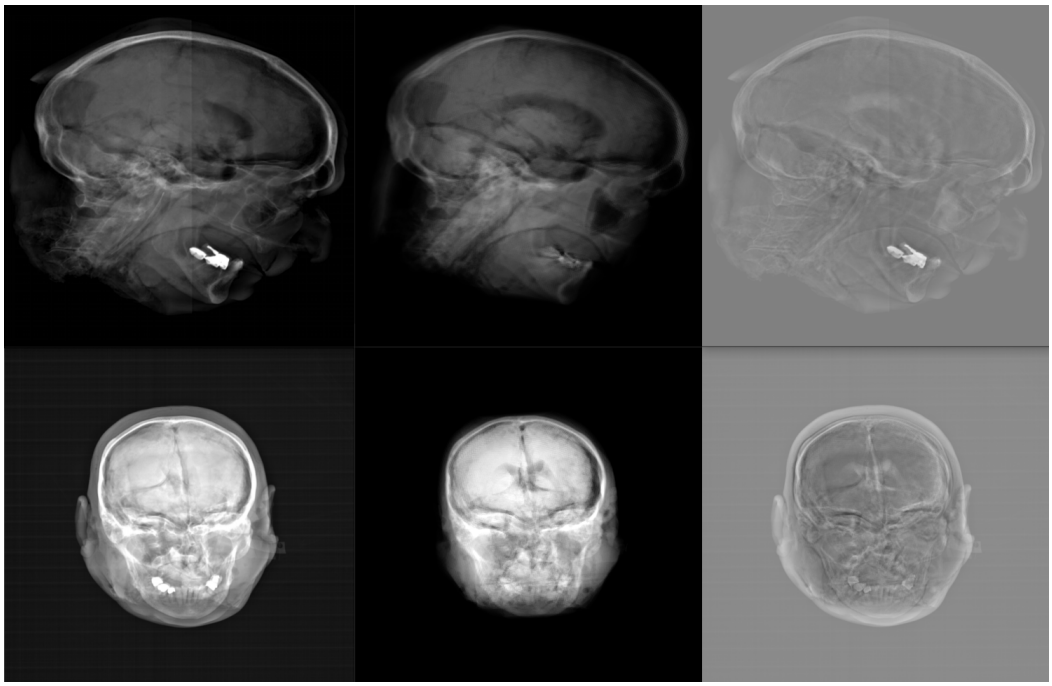


Figure 3.10 Digitally reconstructed radiographs obtained while modelling the Lodox Statscan X-ray system for lateral view (LAT), shown on the top row, and anterior-posterior (AP) view, shown on the bottom row. The original X-ray images are shown in the left column while the simulated X-ray images shown in the middle column. Finally, for a better comparison, difference images are provided in the right column. In order to better visualise the differences, any given pixel intensity value of the difference image, i_d, was computed as $0.5(i_r - i_s + 1)$ where i_r and i_s are the corresponding pixel intensity values of the real X-ray image and simulated X-ray image, respectively.

Manual 2D-3D registration through visual inspection is made possible with the interactive mode of the renderer. Figure 3.11 shows the result of such a registration, demonstrating a high degree of accuracy with which the renderer is able to approximate real X-ray machines.

A useful feature of the DRR renderer is its ability to utilise a transfer function. A common use-case is for the transfer function to act as a thresholding function so that only certain types of tissue are rendered, as is shown in Figure 3.12. Such a strategy is effective when performing a registration in stages, or when registering a CT-volume that has soft tissue present (see Figure 3.13).

The DRR renderer is also capable of switching between different compositing functions. Figure 3.15 provides example renderings of the femur bone while using four example compositing functions: standard DRR, maximum-intensity projection, silhouette, and a

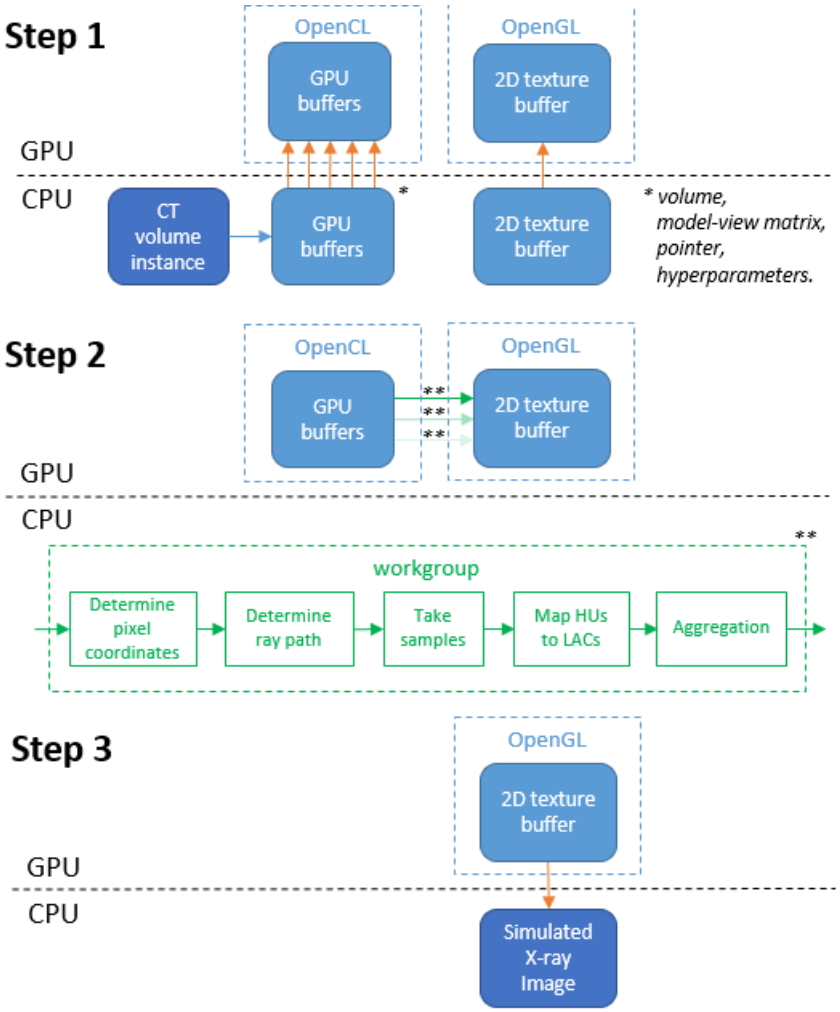


Figure 3.11 The result of a manual 2D-3D registration performed with the help of the "interactive mode" of the DRR renderer. Once registered, it is evident that the X-ray image (outputted by an actual X-ray machine) was approximated with a high degree of accuracy.

custom function based on statistical moments. Compared to the case of the standard DRR compositing function, this functionality allows for additional valuable insights to be gained while interactively exploring a CT volume.

3.7 Discussion

In this chapter, a standard "ray-casting" volume rendering pipeline was adapted for the simulation of X-ray images, and for use in a 2D-3D reconstruction algorithm. Taking a volume rendering approach provided an intuitive framework for understanding and modelling the unique imaging characteristics of X-ray machines. Furthermore, additional computations

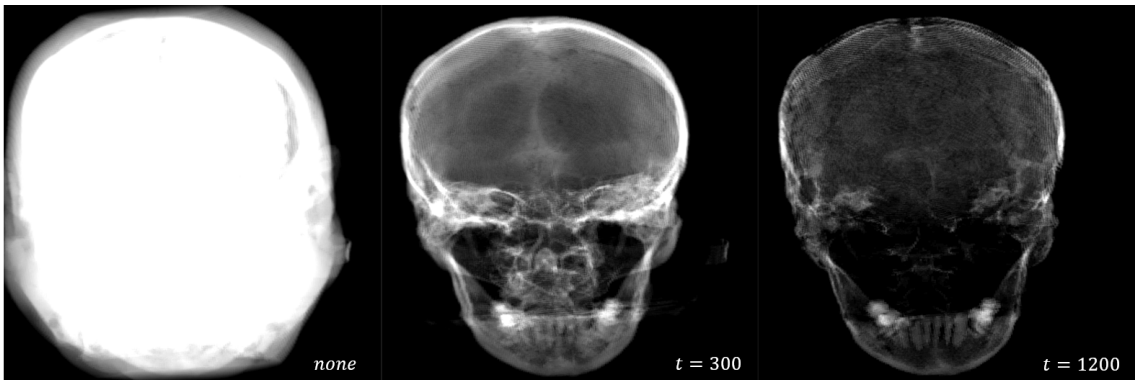


Figure 3.12 Simulated X-ray images of a human head, with HU values thresholded at the following values: none (left), 300, (middle), 1200 (right).

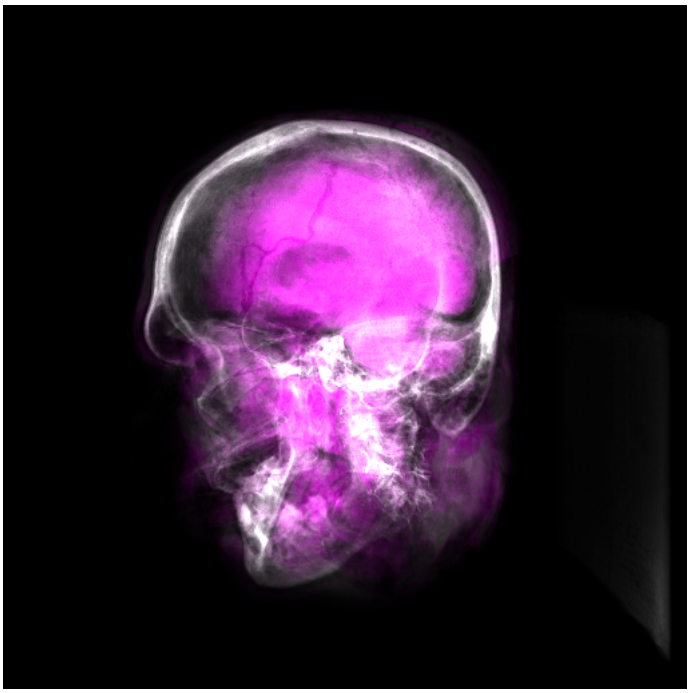


Figure 3.13 Volume rendering of a CT volume of a human skull with regions corresponding to soft-tissue highlighted in purple using a transfer function.

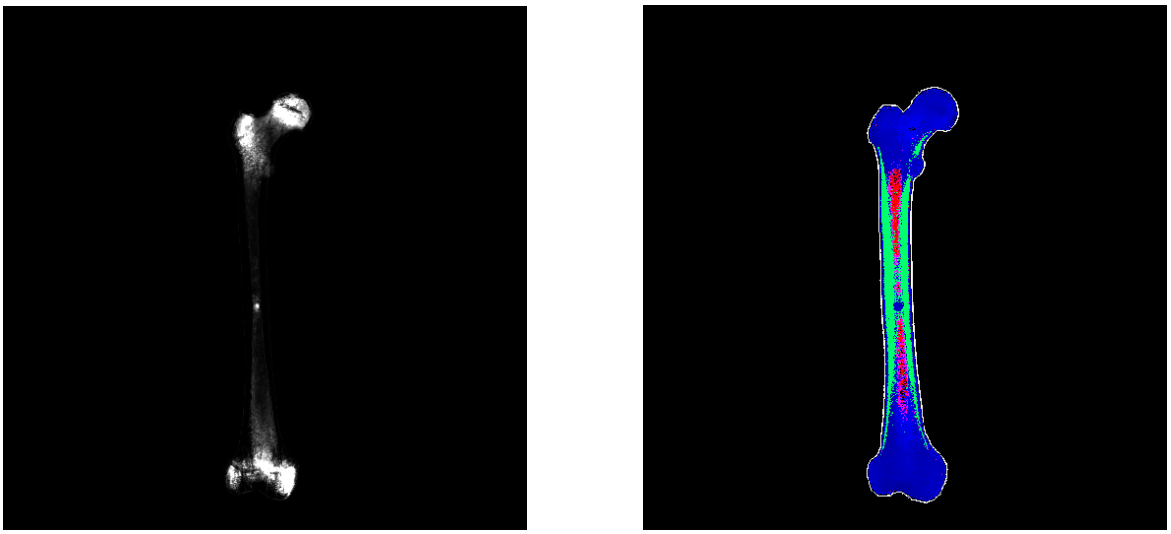
required for X-ray simulation such as the conversion of HUs into LACs were easily incorporated into the pipeline. The ray-casting method for volume rendering was used, which casts one ray for each pixel of the output image. Its relative simplicity is appealing because of the ease with which it is parallelised. It is also amenable to useful techniques such as ray-box intersection, which excludes the rendering of any voxels which do not lie within the viewport of the render - providing additional gains in frames-per-second. The disadvantage of the one-ray-per-pixel assumption is that the appearance of jaggies is more likely. To mitigate this

(a) MIP (b) Silhouette (c) Custom

Figure 3.14 The DRR renderer provides a selection of aggregation types, which include standard DRR (left), a maximum-intensity projection (MIP) (centre), and silhouette (right).

side-effect, each sample point was assigned a value aggregated from its nearest voxel values using trilinear interpolation. Regular sampling intervals produce Moiré patterns, which were counteracted by introducing small random offsets to the ray-origin of each ray along its corresponding ray direction. In doing so, the Moiré patterns were substituted for noise, which the human eye is less sensitive to, and which is better suited to clinical applications where X-ray images are still mostly inspected visually. Hounsfield conversion was performed quickly and efficiently using the GPU kernel; only voxels which lie within the viewport of the renderer are converted, instead of converting all the voxels beforehand. A simple absorption-only variation of the volume rendering equation was used to composite the converted samples encountered along a given ray. The benefit of doing this is better rendering performance with regard to frames-per-second since the equation is less computationally complex. However, more complex secondary imaging effects (such as scatter-radiation, beam-hardening and veiling-glare) were not included, which, in principle, make for an increased disparity between a real X-ray image and its simulation. Such phenomena mostly influence intensity information rather than geometric information. Also, for the case of fan-beam X-ray machines, veiling glare effects are already significantly reduced in the scan-direction by the collimator.

The iterative nature of DRR-based 2D-3D reconstruction, and the presence of many local optima, necessitate that thousands of X-ray images need to be simulated to ensure the parameter space is sufficiently explored. Researchers have explored ways to mitigate the computational bottleneck caused by the DRR rendering process by parallelising the ray-casting algorithm in order to take advantage of hardware-based acceleration. This was also done for the DRR renderer, which provided a significant reduction in the time it takes to obtain a DRR (a quantitative evaluation of the performance of the renderer is described in



(a) LAC distribution skewness (b) Multi-class segmentation

segmented according to the skewness of the distribution of intensity values along each ray.

Figure 3.15 The DRR renderer provides the ability to define a custom aggregation function, allowing specific regions of interest to be identified. For example, shown on the left are bone regions intersected by rays whose distribution of LAC values have a relatively greater skewness (third statistical moment). Furthermore, by employing multiple thresholds based on a selection of statistical measures, it was possible to segment multiple regions (shown on the right), which loosely correspond to three types of bone: marrow (red), cortical (blue), and trabecular (green).

Chapter 4). Real-time interaction was made possible and is helpful for gaining an intuition for bone-specific nuances as well as the manner in which machine-specific parameters such as S2O distance or different rendering modes like maximum-intensity-projection affect bone appearance.

When provided with a standard DICOM X-ray image, the DRR renderer automatically configures itself to the specific parameters of the X-ray machine used to obtain the X-ray image. Thus, together with its interactive functionality, the renderer makes for a more agreeable user-experience. However, the DRR renderer only accepts volumes stored in the Nifty format. Computed tomography machines output CT slices in the DICOM format, which first need to be collated into the Nifty format in order to be used by the renderer. The user is therefore still required to do so manually using third-party software and the renderer could thus be improved by automating this process. Furthermore, the renderer also requires volumes to be in isotropic 1-mm volume spacing in order to correctly calibrate the X-ray projection geometry. Volumes that are provided with a different volume-spacing

are automatically resampled to meet this requirement, which may result in a loss of high frequency information.

3.8 Conclusion

This chapter presented a volume rendering approach to X-ray simulation, and the development of a DRR renderer, which was designed accordingly. Such an approach allows machine-specific X-ray projection geometry to be accurately and efficiently modelled, while a significant reduction in computation time is made possible through a hardware-accelerated parallelisation of the rendering algorithm. The chapter to follow presents an experimental investigation to provide insight into what extent machine-specific parameters influence DRR fidelity.

Chapter 4

Evaluating the influence of machine-specific parameters on DRR fidelity

4.1 Introduction

The standard approach to simulating X-ray images includes implicit X-ray machine-specific calibration settings. These settings affect both the resulting geometry and intensity profile of simulated X-ray images, and will thus also impact the accuracy of estimated volumes when used in a 2D-3D reconstruction algorithm. However, despite the fact that X-ray machines vary significantly with regard to their imaging setup, the importance of correct projection calibration is given little attention in literature. In the previous chapter, a DRR renderer was developed using a standard volume rendering framework, adapted to the unique requirements of X-ray imaging. The renderer can be calibrated to machine-specific parameters. While the ultimate aim of the project was to better understand the influence of machine-specific parameters on 2D-3D reconstruction accuracy, it was first necessary to validate the renderer and investigate the influence of the machine-specific parameters on DRR fidelity.

What follows is a detailed description of a set of three experiments that were performed. The first two focused on the source-to-object distance (d_{S2O}) and incident energy (I_0) machine-specific parameters while the third focused on the time it takes to render a simulated X-ray image. A detailed description of the materials and methods used are provided, as well as a discussion of the results of each experiment. A general discussion of the insights gained concludes the chapter.

4.2 Materials

A calibration "phantom" (Figure 4.1), of known physical dimensions (mm) and density (HA/cm^3), manufactured by QRM GmbH [127], was imaged using both CT and fan-beam X-ray modalities. Thus, using the CT volume as an input, the simulated X-ray images produced by the renderer could be compared to ground-truth X-ray images.

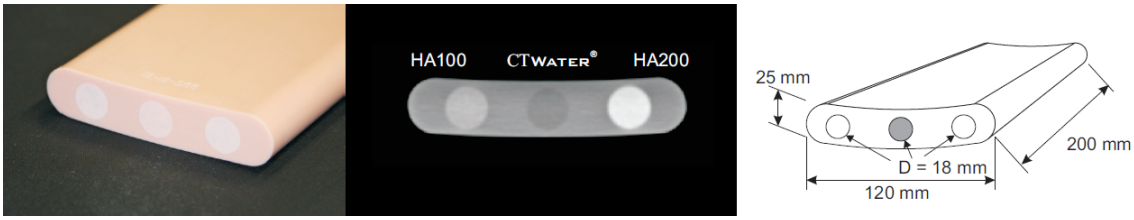


Figure 4.1 A photograph (left), and CT image (middle), and labelled schematics (right), taken of the QRM-BDC bone-density calibration phantom. "HA100", "CTWater", and "HA200", correspond to proprietary composite materials used to simulate specific bone mineral densities: 100 HA/cm^3, 0 HA/cm^3, and 200 HA/cm^3, respectively [127].

The dataset itself comprised a 3D CT image and 2 corresponding 2D X-ray images, one for the anterior-posterior orientation (947 × 1896 pixels, $d_{S2O} = 987$ mm, $d_{S2D} = 1300$ mm), and one for the lateral orientation (947 × 1764 pixels, $d_{S2O} = 918$ mm, $d_{S2D} = 1300$ mm) of the object. Each X-ray image was saved in "DICOM" format by the X-ray machine. The DICOM format is typically used by manufacturers to store digital X-ray images. The 2D X-ray images were imaged using an EOS low-dose X-ray imaging system [48] at IMT Atlantique, INSERM in Brest, France. The EOS system employs a vertical, biplanar scanning procedure resulting in stereo-radiographic images (see Chapter 3.4.1).

A CT image was acquired of the phantom at IMT Atlantique, INSERM in Brest, France. The original CT image comprises 649 slices, each comprising 768 × 768 pixels. Each CT image has a uniform trans-axial pixel size of 0.260417 mm, and slice-spacing of 0.8 mm, thus resulting in a voxel size of 0.260417 × 0.260417 × 0.8 mm. The phantom was originally placed in a water bath to simulate the presence of soft tissues. For the purposes of calibrating the DRR renderer, however, the phantom region in each CT slice was hand-segmented in order to remove this confounding factor. A CT volume was subsequently extracted and cropped to only include the bounding box of the phantom. Regions which previously corresponded to water were assigned a HU value of -1000, which corresponds to air. Finally, the CT volume was resampled to a uniform voxel size of 1 mm using a Lanczos filter[1], resulting in a volume

[1]The Lanczos filter is known to yield competitive results for both magnification and minification resampling operations. However, while avoiding blurring effects, the Lanczos filter is more prone to produce excessive "ringing" effects.

with dimensions of 140 × 37 × 209 mm (or pixels). All image processing was performed using the Amira suite [158], version 5.4.3, a specialised software platform for 3D processing, analysis and visualisation.

The DICOM format, contains important parameters which affect the resulting 2D geometry of an imaged object, including source-to-object distance (d_{S2O}), and source-to-detector distance (d_{S2D}). When provided with a DICOM-formatted X-ray image, the DRR renderer reads these parameters and configures itself accordingly.

4.3 Experiment 1: Assessing the accuracy of 2D geometric information

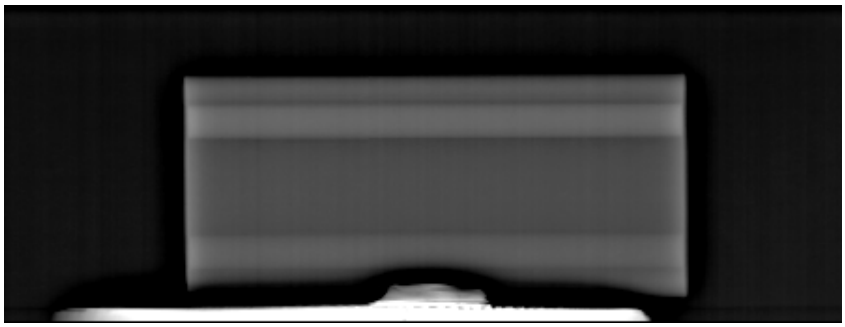
An investigation was conducted to determine how accurately the 2D geometry of a calibrated shape (the phantom shown in Figure 4.2)) could be approximated when provided with the correct machine-specific parameters, and to what extent deviations from the correct value for (d_{S2O}) and (d_{S2D}) affect this accuracy. While the DRR renderer can accommodate various kinds of projection geometry, only fan-beam geometry was investigated, since it is the only geometry for which matching data (a CT volume and X-ray image pair for the same object) were available. Nonetheless, fan-beam geometry is arguably the most challenging type of projection as it induces anisotropic distortions, and is unique in this regard. It furthermore encapsulates the more well-known projections, employing perspective projection in the slot-direction and parallel projection in the scan-direction (see section 3.4.1). This difference is, in fact, what brings about the anisotropy. It was therefore assumed that the validation of the renderer for this type of geometry will also be valid for the more commonly known ones.

Methods

First, the volume renderer was automatically calibrated using the machine-specific parameters of the X-ray machine, which were extracted from the DICOM file. The ground-truth X-ray image was segmented into two classes corresponding to the pixels belonging to the phantom, and the remaining pixels (see Figure 4.2). Next, the correct pose of the phantom was determined manually, using the interactive mode of renderer and while using the "silhouette" mode[2] of the renderer (see Section 3.4.5).

Silhouette projections were also obtained for a set of d_{S2O} offset magnitudes ranging from -500 cm to d_{S2D} cm in increments of 10 cm. The accuracy of a given calibration offset

[2] A simplified DRR which has binary intensity values.



(a) Ground-truth X-ray image

(b) Segmented ground-truth X-ray image.

Figure 4.2 Target X-ray image for the test phantom.

was evaluated as the degree to which its corresponding projection overlaps with the ground-truth X-ray image, which was measured using 2D Dice coefficients [155]. It was therefore assumed that a correctly calibrated renderer will produce the largest overlap, and that this overlap steadily decreases in proportion to the magnitude of calibration offset. Furthermore, despite a correct calibration, discrepancies between the projection and ground-truth were to be expected due to small errors arising from the processes of volume segmentation and manual registration. Finally, the previous step was repeated for d_{S2D}, for a set of offset magnitudes ranging from d_{S2O} cm to 500 cm in increments of 10 cm.

4.3.1 Results and discussion

Two sets of 2D Dice coefficient measurements corresponding to d_{S2O} and d_{S2D} were plotted as a function of offset (cm), as shown in Figure 4.3. As expected, the simulated X-ray images were most similar to the target X-ray image ($DC_{2D} \sim 0.98$) when machine-specific parameters were used. The overlap between the Dice coefficient values decrease linearly in proportion to the magnitude of the applied offset. The rate at which the Dice coefficient values decrease is greater for the d_{S2O} parameter (where a d_{S2O} offset of $10cm$ leads to a Dice coefficient error between $0.05 - 0.065$) than for the d_{S2D} parameter (where a d_{S2D} offset of

$10cm$ correspond to a Dice coefficient error between $0.04 - 0.054$). This is because the d_{S2O} is associated with distortions proportional to the distance a given ray is from the principal line; d_{S2D} is associated primarily with an isotropic magnification effect for the case of perspective projection, but is still partly the case for fan-beam projection, in the slot direction. The rate at which the 2D dice coefficient decreases is greater when the distance is reduced rather than increased, and is the case for both the S2O distance, and the S2D distance. This is because both perspective and fan-beam projections approximate a parallel-beam projection as an object's distance from the X-ray source increases. The distortions which occur are amplified more greatly as the distance decreases. For the d_{S2O} parameter, the constant Dice coefficient values correspond to the point at which the object is magnified beyond the field of view of the detector. A similar series of values would eventually be seen for the d_{S2D}, as the projection approximated an orthographic projection.

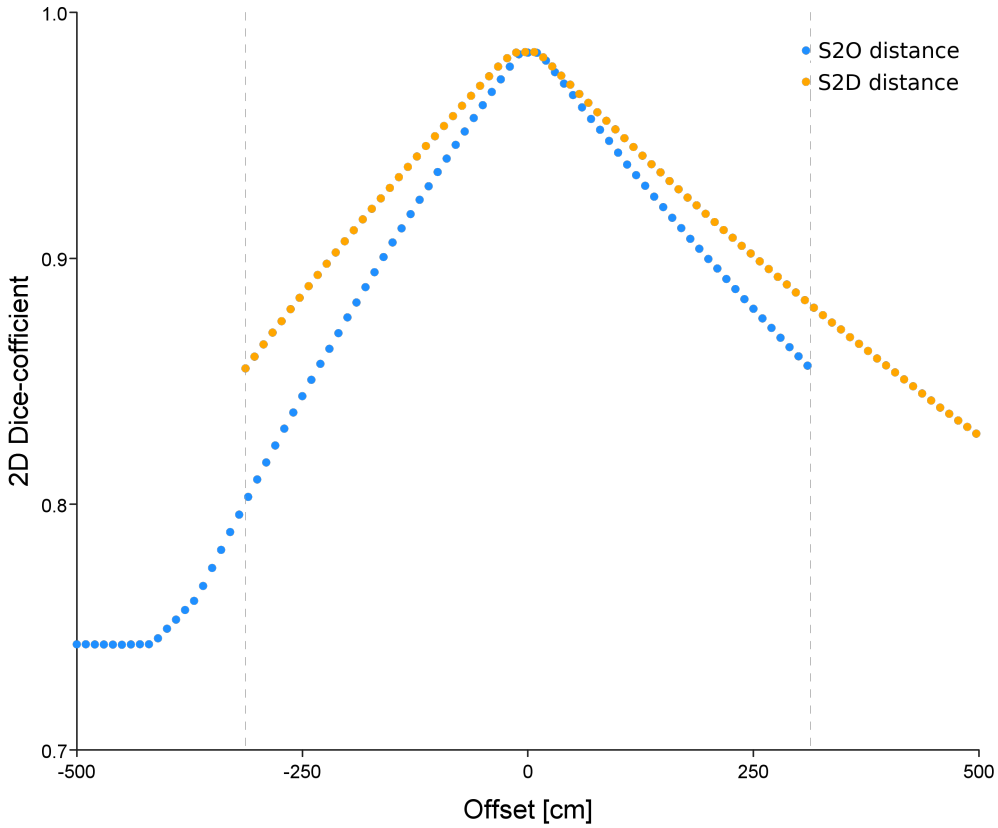


Figure 4.3 The 2D overlap (measured by the 2D Dice coefficient) between the ground-truth and a simulated silhouette projection, with d_{S2O} offset values ranging from -500 cm to d_{S2D} cm in increments of 10 cm, and d_{S2D} offset values ranging from d_{S2O} cm to 500 cm, in increments of 10 cm.

The discrepancy between the hand-registered DRR and ground-truth was small (Figure 4.4). Such small discrepancies are to be expected and can be attributed to a combination of segmentation error, registration error, and human error. Segmentation errors occurred both during the segmentation of the ground-truth X-ray image, and the phantom CT image. Portions of the phantom along the bottom-edge of the X-ray image were not clearly imaged. For the X-ray image, a concentration of artefacts are present around the bottom of the image, causing reduced visibility of the phantom. While the material properties of the platform used to hold the phantom in place are unknown, it is possible that they caused an increase in the amount of scatter radiation in this region, thus making the successful traversal of photons to the detector pixels corresponding to this region less likely. For the CT, segmentation errors are primarily due to the partial volume effect, resulting from the limited resolution of the CT image, and causing pixels to be mis-classified as background. Registration errors occur because more than one 3D appearance can account for the same 2D projection. Furthermore, since the registration was performed manually, human error is also a factor.

(a) Difference image when S2O distance offset is equal to 497*mm*.

(b) Difference image when S2O distance offset is equal to 0*mm*.

Figure 4.4 Difference images for S2O distance offset values. In order to better visualise the differences, any given pixel intensity value of the difference image, i_d, was computed as $0.5(i_r - i_s + 1)$ where i_r and i_s are the corresponding pixel intensity values of the real X-ray image and simulated X-ray image, respectively.

4.4 Experiment 2: Calibration and accuracy of simulated pixel intensities

An empirical calibration of the renderer was conducted with respect to its incident-energy parameter, I_0, which affects the resultant pixel intensities that comprise the simulated X-ray images. This was accomplished by tuning the parameter until the intensity difference between the simulated X-ray image and the real X-ray image was minimised. Furthermore, the extent to which deviations from the empirically determined correct value for I_0 affect the fidelity of DRR intensity information was investigated. The pixel intensity information that is outputted by the DRR renderer is affected by two of the tunable parameters of the DRR renderer, namely, the LAC of water, μ_{water}, and the incident energy, I_0, which is attributed to each modelled X-ray (see Chapter 3.4.4). Under the assumption that the CT machine used to image the phantom was correctly calibrated, μ_{water} was assigned a value of 0.25. This experiment was designed to calibrate the value of I_0, and in doing so, determine an upper bound of the accuracy for rendered pixel intensities. Despite a thorough machine-specific calibration of the volume renderer, deviations from the ground-truth X-ray images are to be expected because phenomena such as secondary imaging effects and photon noise have not been modelled (see Equation 3.19).

4.4.1 Methods

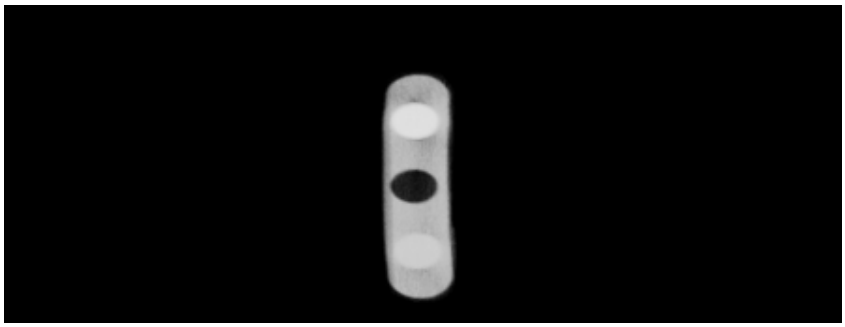
As was done for the previous experiment, the volume renderer was first calibrated according to the machine-specific parameters of the X-ray machine. Next, the correct phantom pose was determined using the interactive mode of renderer (see Section 3.4.5). A set of simulated X-ray images were then generated with I_0 values ranging from $6.5 - 8.5$ in increments of 0.5, which was assumed sufficient for the investigation. For each of the simulated images as well as the ground-truth X-ray image (Figure 4.5), a set of 25 pixel intensities were sampled from two regions, corresponding to density values of $100 HA/cm^3$ and $200 HA/cm^3$ (see Figure 4.1). Since these regions had uniform density and no other objects - or regions corresponding to other densities - occluded the path from the X-ray source to the detector, it was assumed sufficient to estimate the average pixel intensity using 25 samples. Finally, the average pixels intensities were plotted for the two regions, for each I_0 value.

4.4.2 Results and discussion

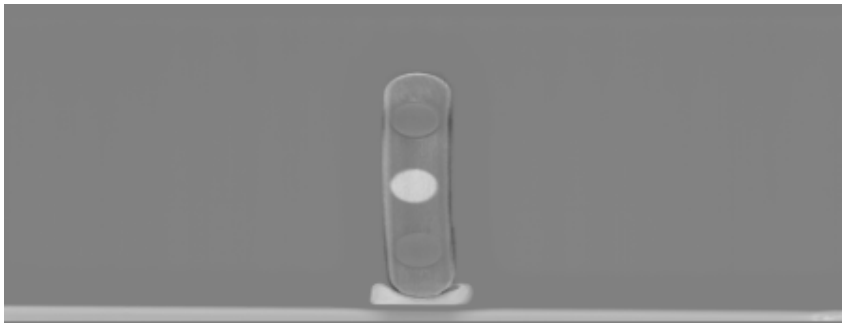
The hand-registered phantom pose is shown in Figure 4.6. For a given I_0 value, the difference between the pixel intensity values for the two density regions has been emphasised to better

Figure 4.5 Target X-ray image for the test phantom.

depict the slope. An incident energy (I_0) offset of 0.5 resulted in an error of 3350 16-bit grey levels (~ 0.051 normalised pixel grey-levels). The pixel intensity error scales linearly with the I_0 offset, and at the same rate (or slope) regardless of the bone density region from which the intensities were sampled.



(a) Digitally Reconstructed Radiograph



(b) Difference Image

Figure 4.6 Resulting DRR after calibrating the DRR renderer ($I_0 = 7.5$). Note that the discrepancy present in the middle circular region is due to it having not been included in the segmentation and did not affect the experimental results. The difference image was computed as described in Figure 4.4.

The DRR renderer models the Beer-Lambert law such that I_0 is not influenced by other variables. Therefore, since I_0 acts as a simple coefficient, the error in intensity per I_0 is

identical for each of the two sets of I_0 values (Figure 4.7). Of course, this would not be the case if secondary imaging effects such as veiling glare, photon noise, and beam hardening were modelled (see Section 2.4.2). By taking a number of pixel intensity samples, secondary imaging effects such as veiling glare, scatter radiation, and photon noise, are mitigated. The effects of beam-hardening, however, are unaffected, and this partly explains why the ground-truth intensity values from the two density regions have smaller errors (emphasised by its slope) than those from the renderer.

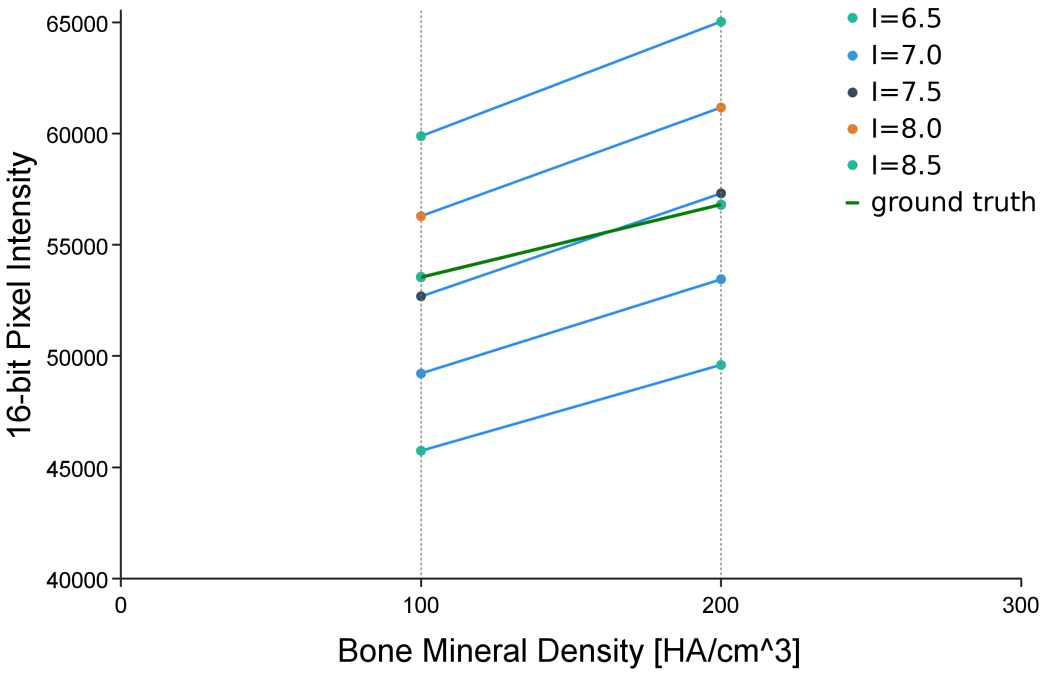


Figure 4.7 The median pixel intensity (16-bit) as sampled on ground-truth images (green), and simulated X-ray images (other colors), for two regions corresponding to a bone mineral densities of $100 HA/cm^3$ and $200 HA/cm^3$, respectively.

4.5 Experiment 3: Computation time vs. volume size

An investigation was conducted to determine the extent to which the desired number of 2D image pixels for the simulated X-ray image, the number of voxels comprising the 3D volume, and the number of samples taken along each casted ray, influenced the time it took to render a simulated X-ray image. Measurements were taken both when using, and not using, hardware-based acceleration. Doing so helped quantify the extent to which rendering time can be reduced when dedicated graphics hardware is available. Hardware-based acceleration has been reported to reduce DRR rendering time by a factor of up to 25 [101]. It was

therefore expected that similar significant performance gains would be measured during the experiment. Since the set of operations executed for each ray are identical, the renderer was designed such that the desired size of the simulated X-ray image corresponds to the number of parallel threads that the GPU executes (see section 3.5). Graphics processing units excel in their ability to execute parallel threads (in the order of tens of thousands), compared to CPUs (in the order of thousands - even with hyper-threading). Most of the reduction in the computation time is achieved through these parallel threads. The number of voxels comprising the 3D volume is limited by the amount of the Random Access Memory (RAM) available to the GPU or CPU. Since GPUs typically have far less RAM at their disposal than a CPUs, a trade-off occurs when employing hardware acceleration. Furthermore, this limitation is compounded when a deformable model is also loaded into GPU memory (as will be demonstrated in Chapter 5). However, GPU technology continues to develop at a significant rate. Thus, such considerations may become less of a concern - at least when modelling a single volume. Finally, the number of samples taken along each casted ray presents a trade-off between computation time and DRR fidelity. While the lower-bound is dictated by the Nyquist theorem [118], some scenarios - such as multi-scale registration - may warrant a smaller sampling rate, where the high-frequency content of the image may be of less importance, and a reduced computation time is favoured.

4.5.1 Methods

The experimental procedure comprised three sub-procedures used to gather frame-rate data corresponding to variations in the 2D image size, 3D volume size, and sampling rate multiplier, respectively. For all three sub-procedures, the rate at which the volume renderer can generate DRRs was measured in frames per second (FPS). Such measurements are subject to noise resulting from background processes executed by the operating system. Therefore, a total of 60 FPS measurements were recorded for each data-point over a period of 60 seconds, and the average was taken. The first set of measurements were recorded for each 2D image size in a sequence of values, where the number of pixels in each subsequent image scales by n^2. A 500^3 volume and a sampling rate multiplier of 1 were used throughout. A second set of measurements were recorded for each 3D volume size in sequence of values, where the number of voxels in each subsequent volume scales by n^3. A 100^2 image and a sampling rate multiplier of 1 was used throughout. A final set of measurements were recorded for each sampling size multiplier in a series, where the magnitude of each multiplier scales by n^3. A 500^3 volume and a 100^2 image were used throughout. The used ranges were selected to encapsulate typical volume sizes and image sizes encountered in reconstruction applications reported in the literature, from smaller bones like the scapula [154] to larger

bones such as the femur [30]. The results were computed using an Intel Xeon 3.5 GHz CPU with a 16-GB RAM, and a NVIDIA GTX 970 GPU with a 8-GB RAM.

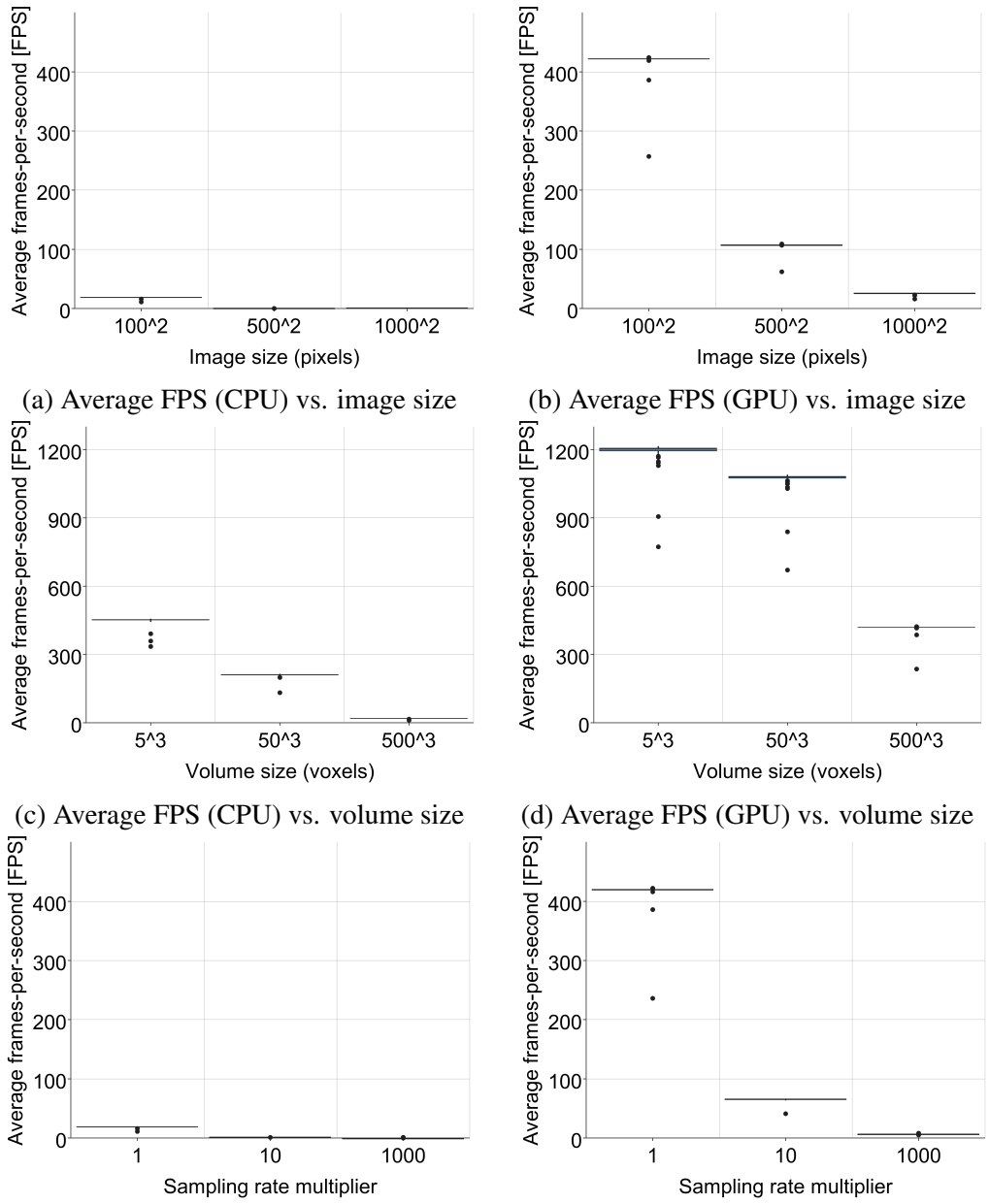
4.5.2 Results and discussion

Figures 4.8a and 4.8b show the average FPS as a function of the pixel size of the simulated X-ray image. Here it can be seen that the time it takes to obtain a DRR scales exponentially as the number of pixels increases logarithmically in the two spatial dimensions. The number of pixels correspond to the number of parallel threads assigned to the GPU. If the number of desired parallel threads exceeds that capacity of the GPU used; which can typically accommodate tens of thousands parallel threads; the GPU will split the computations into separate batches, resulting in an additional computation time overhead. The desired size of the simulated X-ray image therefore has a significant impact on level of performance that can be expected from the renderer.

Figure 4.8c and 4.8d show the average FPS as a function of volume size. Here it can be seen that the computation time associated with the number of voxels that comprise the volume did not scale linearly. Since the ray-casting algorithm is agnostic to the 3D volume it receives, volume size should only affect the amount of RAM used by the GPU. Nonetheless, computation time was indeed influenced, since the greater the "spread" of the volume in relation to the 2D detector plane, the more rays that intersect with the bounding box of the volume and therefore be marked for traversal by the ray-box-intersection algorithm. Therefore, past a certain the volume size threshold, the only additional computation time cost incurred will be by the process of copying the volume data into the GPU memory. Volume size is therefore a less important factor to consider with regard to FPS-based performance compared to the desired size of the simulated X-ray images. Instead, the volume is important when determining the maximum volume size that could be accommodated by the memory capacity of the employed CPU or GPU hardware.

Figure 4.8e and 4.8f show the average FPS as a function of the number of samples taken along each ray. Similar to the number of pixels comprising the 2D image, computation time scaled exponentially as the number of samples taken along each ray was increased logarithmically. Significant performance gains can therefore be realised by reducing the number of samples taken along each simulated X-ray. For example, such a strategy might prove especially useful in multi-stage registration, where the number of samples taken along each ray is increased as the algorithm progresses and gradually takes into account finer details.

Generally, the hardware-accelerated implementation of the renderer enabled a reduction in computation time by a factor of as much as 14, while factors of up to 25 are reported in



(a) Average FPS (CPU) vs. image size

(b) Average FPS (GPU) vs. image size

(c) Average FPS (CPU) vs. volume size

(d) Average FPS (GPU) vs. volume size

(e) Average FPS (CPU) vs. sampling rate

(f) Average FPS (GPU) vs. sampling rate

Figure 4.8 Results of computation time experiments. Average frames-per-second (FPS) results on CPUs are shown on the left, while those on GPUs are shown on the right. The first row shows FPS as a function of the size of the 2D detector (number of pixels). The second row shows FPS as a function of the CT volume size (number of voxels). The third row show FPS as a function of the sampling rate (multiple of the standard sampling rate).

literature [101]. It should however be noted that the CPU ray-casting method used for the comparisons in this section is not optimised, but rather an "out-of-the-box" OpenCL method for compiling GPU kernels so that they may be run on the CPU. The methods reported in the

literature also include object-order methods, which - while faster - are less compatible with deformable CT volumes. The renderer developed and presented in this chapter uses image-order rendering (volume ray-casting), which allowed for further reductions in computation time when the model deformation step was also parallelised, and hardware-accelerated (detailed in Chapter 5).

4.6 Discussion

In this chapter, three experiments were performed to validate the adapted renderer with respect to geometric information, intensity information, and computation time. For the first two experiments, two of the configurable machine-specific parameters of the adapted renderer were focused on: the distance from the X-ray source to the object to be imaged, and modelled incident energy of the X-ray source, respectively. Sets of simulated X-ray images were rendered as each correctly calibrated parameter was offset by an incrementing magnitude. The resulting rendering errors were then measured through comparison to a ground-truth X-ray image - with known calibration parameters - of an object of known geometry and density.

The first experiment demonstrated that the renderer is capable of accurately modeling machine-specific ray geometry. Furthermore, it was observed that every 1 cm with which the S2O distance was offset resulted in an additional 2D Dice overlap error of 0.4%.

The second experiment detailed the empirical calibration of the renderer with respect to its incident-energy parameter, I_0, which affects the resultant pixel intensities that comprise the simulated X-ray images. When $I_0 = 7.5$, the difference in pixel intensities between the simulated X-ray image and ground-truth X-ray image was minimised. Furthermore, a normalised pixel intensity error of 0.12 was incurred for every 1-point offset of I_0. The results from the first two experiments indicate that an incorrect calibration of the DRR renderer can cause significant errors in DRR fidelity with respect to both geometric and intensity information.

The final experiment investigated how three factors influenced the time it took to render a simulated X-ray image - both with and without hardware-based acceleration. These factors included: the desired size of the simulated X-ray image, the size of the volume to be rendered, and the sampling rate multiplier which determined the number of samples taken along each simulated X-ray. Doing so made it possible to quantify the extent to which rendering time can be reduced when dedicated graphics hardware is available. Generally, a FPS performance increase by a factor of as much as 14 was observed. The desired image-size and sampling rate multiplier were shown to be most important to consider when concerned with FPS

performance, while volume-size was shown to be important when determining the memory requirement of the renderer.

4.7 Conclusion

This chapter presented an experimental investigation which demonstrated that machine-specific calibration parameters have a notable impact on DRR fidelity. However, X-ray simulation only makes up one component of the 2D-3D reconstruction algorithm, and is only a means to obtaining accurate reconstructions. In the chapters to follow, the development and validation of such a 2D-3D reconstruction algorithm is described, followed by an experimental investigation of the influence of machine-specific calibration parameters on reconstruction accuracy.

Chapter 5

Digitally Reconstructed Radiograph-based 2D-3D Reconstruction

5.1 Introduction

The state-of-the-art approach to orthopaedic 2D-3D reconstruction is to iteratively warp a deformable 3D model of the bone-of-interest, project a simulated X-ray image from it, and compare the result to a target X-ray image using a similarity measure. In the previous chapters, an X-ray simulation method was developed (Chapter 3) and validated (Chapter 4). From the validation, machine-specific calibration parameters were found to have a notable impact on DRR fidelity. However, simulated X-ray images are only a means for reconstructing 3D bone appearance, and, therefore, the influence of machine-specific parameters on 2D-3D reconstruction accuracy requires investigation. To do so, a novel 2D-3D reconstruction algorithm was developed. The algorithm comprises a voxel-based deformable femur model, and a Bayesian inferencing framework, implemented on the GPU. This chapter describes the development of both these components, which includes the application of their respective theoretical frameworks, and technical implementations. Thereafter, final insights and conclusions are discussed, concluding the chapter.

5.2 Modelling femur computed tomography with voxel-based Gaussian Process Morphable Models

The statistical shape model (SSM) was originally proposed by [39] as an efficient way to model shape variations using linear combinations of point distributions, subject to the assumption that it is possible to model such variations using a normal distribution. The SSM, which is essentially a type of point distribution model (PDM), has since been reformulated as a Gaussian process, and is referred to as a Gaussian process morphable model (GPMM)[1] [95]. The GPMM formulation offers a number of advantages, some of which are useful in the context of 2D-3D reconstruction. First, GPMMs are well-suited to probabilistic frameworks (such as the one presented in Section 5.3), where a measure of uncertainty is able to make the parameter-search more robust. Second, geometry is preserved when vector operations are performed, making interpolation possible. The ability to interpolate allows GPMMs to be resampled onto other domains, such as those of a different size or spacing. This type of flexibility is useful where complex parameter search strategies such as multi-resolution optimisation are required. Gaussian process morphable models permit marginalisation of the model, thus allowing select regions-of-interest to be focused on. Third, GPMMs provide more flexibility than SSMs; not only can the model's covariance function be estimated from example data but, with the help of kernel functions such as splines or radial basis functions, can also be defined analytically, where insufficient or no training data are available. Furthermore, the covariance function can be easily combined with other covariance functions, augmented (for example, such that it has multi-scale or anisotropic properties), or made to incorporate additional priors such as landmark constraints. What follows is a mathematical description of the GPMM, which closely follows the one presented by [95]. For reasons outlined in Section 5.2.2, a voxel-based approach to GPMMs was required. Therefore, the mathematical description was adapted to be applied to CT voxel-volumes. Thereafter, the procedure used to build the GPMM used for the 2D-3D reconstruction is described and validated using both artificial and real data, as provided in Chapter 6.

5.2.1 Previous work

The standard SSM-based formulation is provided in Chapter 2. What follows is a reformulation of these equations, using the approach proposed by [95]. Equation 2.1 is first reformulated to model deformation vectors, rather than point distributions, with which the

[1] Unlike PDMs, GPMMs define deformations in the continuous domain, but can also at any time be discretized to a form that is mathematically equivalent to a PDM, enabling the use of classical shape modeling algorithms.

vertices that comprise the reference shape are deformed:

$$\hat{\Gamma}_n = \{\bar{a}_p + \hat{\mathbf{d}}(a_p^n) \mid a_p \in \mathbb{R}^3\} \tag{5.1}$$

where $\bar{a}_p$ is the mean Euclidean position of the p^{th} point, $\bar{a}_p = \Sigma a_p / N$, and $\hat{\mathbf{d}}$ is a deformation vector from $\bar{a}_p$ to point a_p^n.

Equation 2.6 is then given a probabilistic representation, which can be computed using the Karhunen-Loéve (KL) expansion[2]:

$$\mathbf{s} \sim \bar{\mathbf{s}} + \sum_{q=1}^{Q} w_q \sqrt{\lambda_q} \boldsymbol{\phi}_q \qquad w_q \sim N(0,1) \tag{5.2}$$

where Q is the number of principal components used, λ_q and $\boldsymbol{\phi}_q$ represent the q^{th} eigenvalue and eigenvector, respectively, and w_q represents the q^{th} model parameter.

While analytically defined covariance matrices can encode expert insights, they might still express deformations that are not necessarily anatomically meaningful. Instead, the covariance matrix can be estimated using samples (a set of deformation fields obtained by fitting the reference to each of the shapes in a training set). However, estimating the covariance matrix of the Gaussian process is computationally expensive. The size of the covariance matrix increases exponentially with inclusion of more training examples, and the greater the resolution (that is, number of vertices or voxels used). For this reason, it is common to use a low-rank approximation of the covariance matrix, which provides a significant decrease in training time [95][3].

5.2.2 Modelling voxel-volumes using GPMMs

The aforementioned GPMM-based approach has been previously used to model surfaces as triangle meshes [95], and volumes as tetrahedral meshes [50]. While the tetrahedra-based approach provides an efficient representation of a bone volume (see Chapter 2.4.1), taking a voxel-based approach may have some advantages. Despite requiring more memory, and thus being more computationally expensive to compute, it can be hypothesised that segmentation errors such as the partial volume effect [55], and the “shrinking problem"[165, 59] may be avoided by a voxel-based approach. Since the structure-of-interest does not need to be

[2]The Karhunen-Loéve expansion allows stochastic processes to be represented as a a linear combination of orthogonal functions. It is closely related to principal component analysis, but also applicable to Gaussian processes.

[3]More detail on the trade-off between computation time and resolution with respect to the covariance matrix is provided in Chapter 2.4.1.

delineated beforehand, surrounding structures such as soft tissues can be included in the model, and this could mitigate the shrinking problem without having to explicitly define a background model. Furthermore, as researchers have begun to explore articulated models [34, 148], and multi-object models [26, 31, 50], a voxel-based approach lends itself more freely to the process of modelling more than one anatomical structure, and especially towards modelling entire CT volumes. Finally, when implemented together with a transfer function, which allows the model to be conditioned on specific structures (see Chapter 3), a voxel-based approach does not require one to commit to a structure-of-interest beforehand, but instead could be decided according to the clinical task at hand, or adapted in an online fashion during the parameter search.

Thus, a voxel-based approach was developed to model femur CT volumes for the developed 2D-3D reconstruction algorithm. This is similar to the procedure used for vertices, except that the fundamental feature of representation is a cuboid, not a mesh vertex, and that the domain on which the structure is defined is a regular grid - not an unstructured one. Thus the entire volume, $\mathbf{v}$, can be modelled, and not merely a segmentation from it:

$$\mathbf{v} \sim \bar{\mathbf{v}} + \sum_{m=1}^{Q} w_q \sqrt{\lambda_q} \boldsymbol{\phi}_q \qquad w_q \sim N(0,1) \tag{5.3}$$

where Q is the number of principal components used, λ_q and $\boldsymbol{\phi}_q$ represent the q^{th} eigenvalue (standard deviation) and eigenvector (deformation vector), respectively, and w_q represents the q^{th} model coefficient (model parameter). Each cuboid (voxel) comprising the volume is assigned a uniform intensity value, which, in the case of CT, corresponds to a HU value.

To train a voxel-based GPMM of the femur, a slight variation on the procedure described in Chapter 2.4.1 was followed. These steps are illustrated in Figure 5.1 and include: (1) binarising the training examples, (2) registering the training volumes, and (3) building the GPMM.

Step 2 itself is comprised of the following steps (illustrated in Figure 5.3): (2.1) alignment of the training examples to the reference volume; (2.2) building a free-form deformation model (FFDM) from the reference volume; and finally (2.3) using the FFDM to establish correspondence between the reference volume and the rest of the training examples. An in-depth explanation follows.

The training set comprised a binarised CT volume, which served as the reference, and a set of binarised 3D surface meshes (a rationale for binarisation is provided further down). First, each volume was annotated with six landmarks (see Figure 5.2), and subsequently aligned to the reference volume using generalised Procrustes analysis (GPA). Next, the reference volume was used to create a preliminary Gaussian process (FFDM); a number

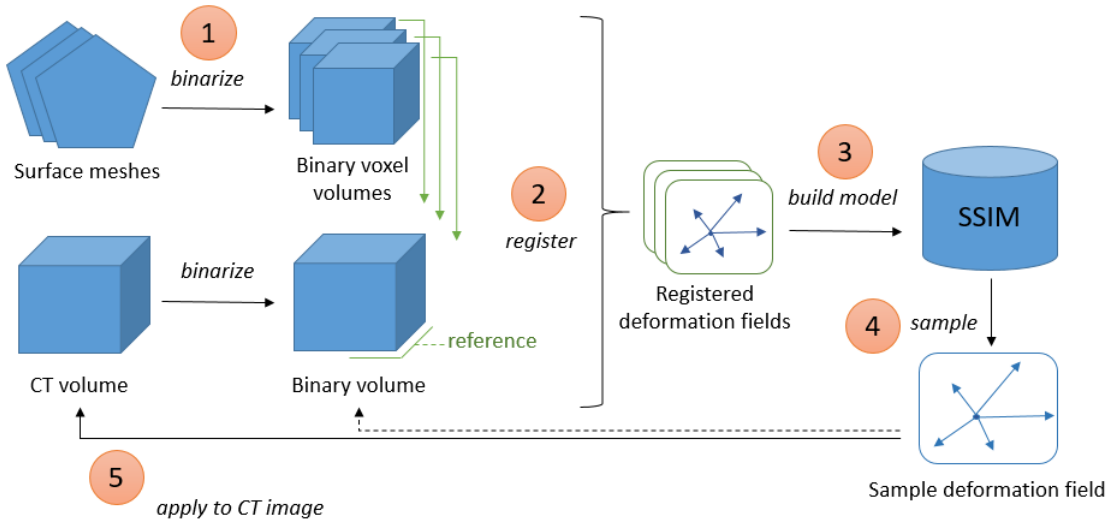


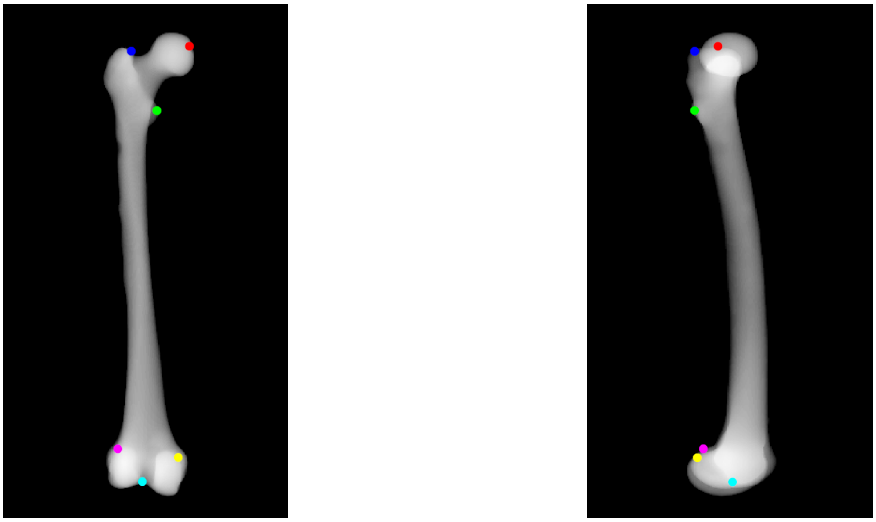
Figure 5.1 Procedure for building a voxel-based deformation model. The key step is that any sampled deformation field is not applied to the binary reference volume, but rather to the original CT volume from which the reference was derived.

of random voxels was sampled from the reference image using a Gaussian kernel, which was then used to build the FFDM. The Gaussian kernel parameters can be determined empirically according to the desired level and type of flexibility; the amplitude determines the magnitude to which the variability of a given voxel will affect its neighbours while the sigma parameter determines the size of the radius encompassing surrounding voxels, which will be considered as neighbours. The more flexible the kernel is made, the better it can fit to the training examples. However, if the kernel is too flexible, the model will be susceptible to over-fitting (details regarding hyper-parameter selection are provided in Section 5.2.3). Next, a non-rigid image-based 3D-3D registration strategy[4] was used to establish inter-volume correspondence, by iteratively deforming the FFDM to fit each training example, resulting in a set of deformation-vector-fields (DVFs)[5]. Finally, the GPMM was obtained by performing a KL expansion (Equation 5.2) on the DVF set. When sampling from the model (Step 4), the sample deformation field was not applied to the binary reference volume, but rather to the original CT volume from which the reference was derived (Step 5). Doing so provides model instances with an intensity distribution that is not binary, but rather one that is more representative of actual CT. The use of binary training volumes made the 3D-3D registration step much simpler, by avoiding the complex local minima present in a full-intensity CT. However, doing so severely limits the flexibility of the deformable model, since CT intensity

[4] An image-based 3D-3D registration uses the similarity of voxel intensity values to guide the registration instead of the distance between mesh vertices, as is typical of a mesh-based registration methods.

[5] For a detailed mathematical description of this process, the reader is referred to [95].

values are not explicitly modelled. Regardless, intensity information was available to guide the 2D-3D reconstruction algorithm, albeit, informed exclusively by values comprising the reference CT volume, and was assumed sufficient for the research objective at hand.



(a) Anterior-posterior view (b) Lateral view

Figure 5.2 Six landmarks were annotated on each of the volumes in the training set. Each landmark has a unique colour so it can be identified in the either of the two viewing angles shown, 0 degrees (left) and 90 degrees (right).

5.2.3 Implementation details

Datasets

A total of 19 3D surface meshes were obtained with permission from the SICAS Medical Repository in Switzerland[6]. All 19 meshes were converted to binarised CT volumes to be used for model building.

A previously imaged adult femur bone CT volume was obtained with ethical approval from the Faculty of Health Sciences Human Research Ethics Committee (HREC) of the University of Cape Town (reference HREC 258/2016). The femur bone was obtained from a dissected leg, with the surrounding tissues removed, collected from a cadaveric specimen at the Division of Clinical Anatomy and Biological Anthropology, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, South Africa. Having been anonymised, information regarding age-range and sex were unknown.

[6]SICAS:http://www.si-cas.com/

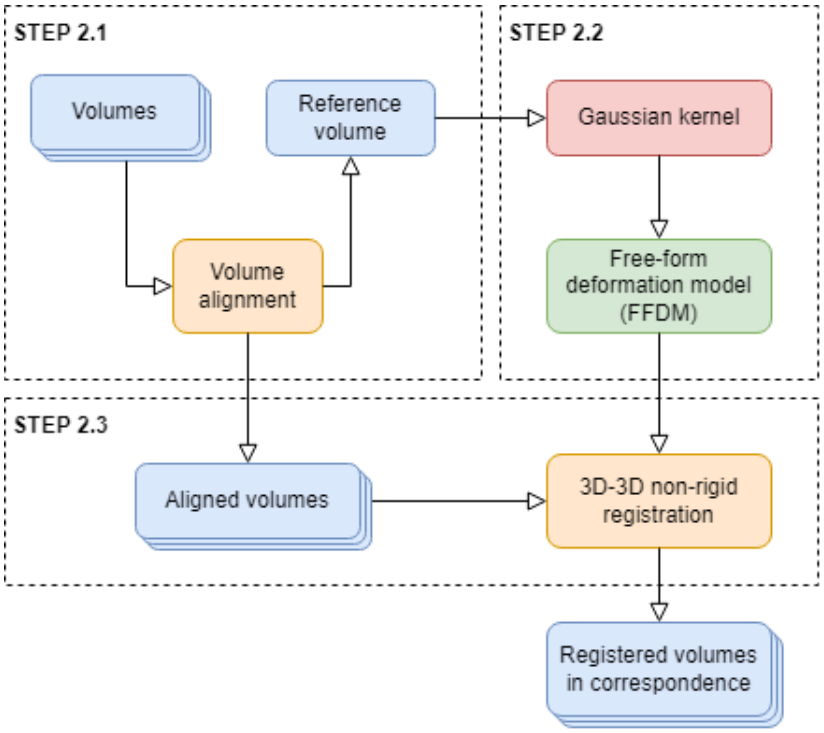


Figure 5.3 Procedure for registering the training volumes used to build a voxel-based GPMM. Step 2.1 is the alignment of the training examples to the reference volume. Step 2.2 is the building a free-form deformation model (FFDM) from the reference volume. Step 2.3 is the use of the FFDM to establish correspondence between the reference volume and the rest of the training examples. Volumes are shown in blue, algorithms in orange, the FFDM in green, and Gaussian kernel in red.

The CT image was imaged at IMT Atlantique, INSERM in Brest, France, comprising 668 slices, each comprising 768 × 768 pixels. The CT image has a uniform trans-axial pixel size of 0.260417 mm, and slice-spacing of 0.8 mm, thus resulting in a voxel size of 0.260417 × 0.260417 × 0.8 mm. The imaged femur bone was originally placed in a water bath to simulate the presence of soft tissues surrounding the bone. Therefore, since the research presented in this dissertation only considered the simplified setting of dry-bone specimens, the femur region in each CT slice was hand-segmented. A CT volume was extracted and cropped to only include the bounding box of the femur. Regions which previously corresponded to water were assigned a HU value of -1000, which corresponds to air. Finally, the CT volume was subsequently resampled to a uniform voxel size of 1 mm using a Lanczos filter, resulting in a volume with dimensions of 200 × 200 × 535 mm. All image processing was performed using the Amira suite [158], version 5.4.3. For model-building, the surface meshes were registered to the CT volume.

Model-building details and hyper-parameter selection

Before training the model with femur volumes, the procedure was first tested in a simpler context, using an existing "lollipop model" [53]. The lollipop model is analytically defined, and has only two modes of variation: the angle of the "shaft", and the elongation of the "head" (see Figure 5.4). In order to validate the voxel-based approach, 100 randomly generated lollipops were generated and subsequently used to train a voxel-based GPMM. Then, the first two principal components of the model were inspected, with the expectation that they will be similar to that of the original analytically defined model. The procedure used to build the lollipop GPMM was the same as the one mentioned before, except for the landmark-based alignment, since the lollipop instances are already aligned. All model building was performed using the Scalismo framework[7], version 0.17 [159]. A Gaussian kernel with amplitude of 20, and sigma of 20 was used to build the FFDM[8]. To establish correspondence, a mean squared error (MSE) metric was used together with a stochastic gradient descent optimisation strategy, with a step-size of 0.25 and a total of 100 iterations per image. To reduce computation time, the MSE was computed on a randomly selected subset of voxels (sample size $= 100$ voxels) instead of each voxel within the volume. In addition, the input volumes to the optimisation strategy were pre-processed; their respective voxel-values were normalised $\{0;1\}$, and subjected to a distance transform, as well as a Gaussian filter ($\sigma = 1.5$), in order to avoid local minima during the registration. The lollipop instances, and the GPMM, were defined on a regular voxel grid with dimensions $70 \times 50 \times 120$. The two principal components of the voxel-based lollipop GPMM are shown in Figure 5.4. Note that the purpose of building this model was not to validate the quality of the instances themselves, but rather to validate that the general variation captured by each principal component corresponds to that of the original mesh model. This was indeed the case; the first principal component shows a variation in the shaft of the lollipop, while the second principal component shows a variation in the elongation of its head. Next, the femur model was trained using 19 binary training volumes. The subsequently mentioned hyper-parameters were determined empirically by iteratively inspecting the result and adjusting the set of parameters accordingly. A Gaussian kernel with an amplitude of 30, and standard deviation (σ) of 50 was used to build the FFDM. To establish correspondence, as was the case for the lollipop GPMM, a MSE metric was used together with a stochastic gradient-descent optimisation strategy, with a step-size of 0.05 and a total of 350 iterations per image. Furthermore, the MSE was computed using

[7]The Scalismo framework is a software library specialised in statistical shape modeling and model-based image analysis. The framework is programmed in Scala and is developed and maintained by the Graphics and Vision Research Group (GRAVIS) at the University of Basel.

[8]For insight into the influence of these parameters on the resulting FFDM, the reader is referred to [95].

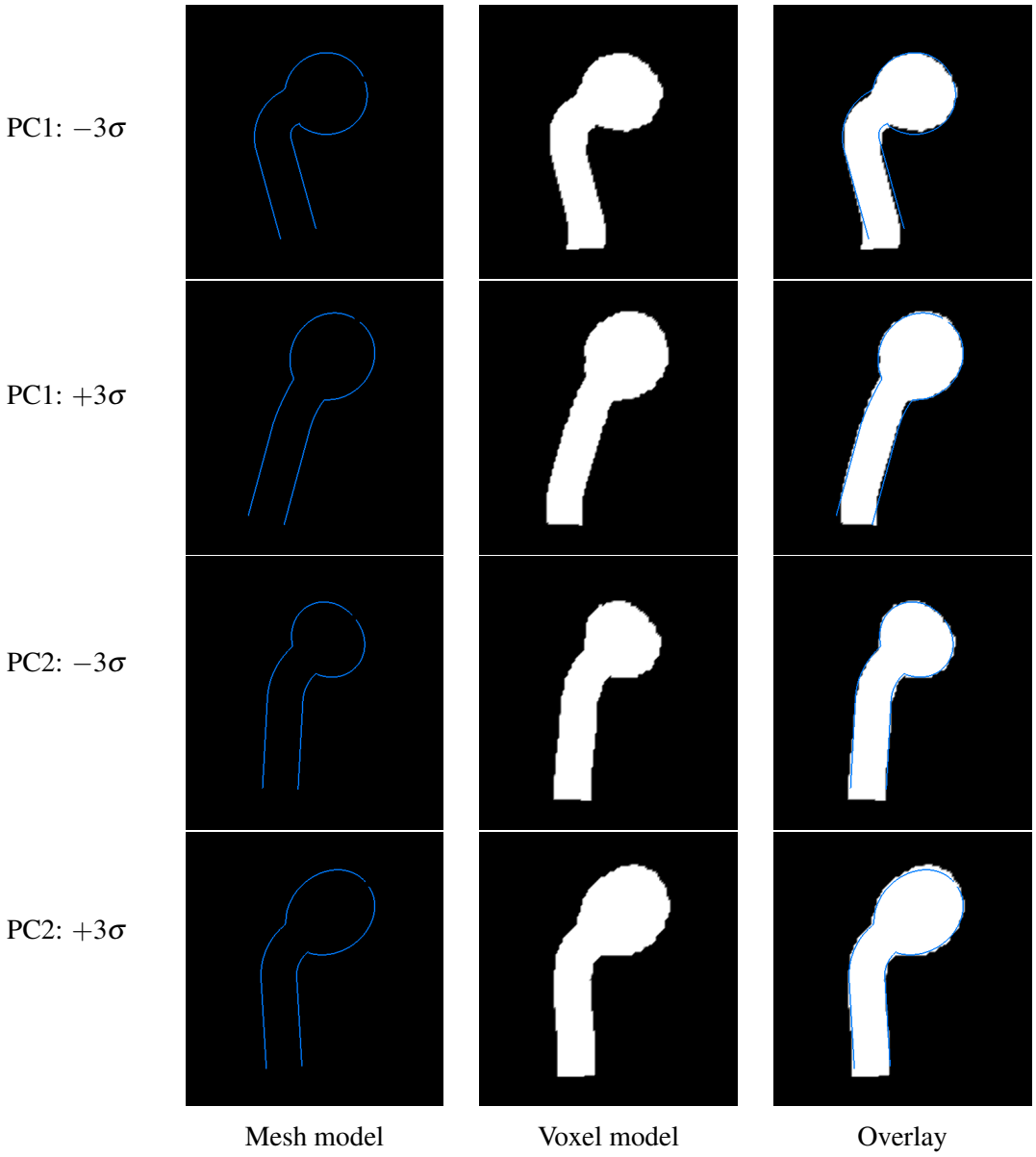


Figure 5.4 Cross-sections of the two principal components of the lollipop mesh-model (left), voxel-model (middle) and an overlay of the two (right). The voxel-model was trained using samples drawn from the analytical mesh model. The top two rows correspond to the first principal component for parameter values -3σ and 3σ, respectively. The bottom two rows correspond to the second principal component for parameter values -3σ and 3σ, respectively.

1000 voxel samples. For a gradient descent-based optimisation, the step-size is set based on the type of local minima encountered in the object at hand; if the step-size is selected too small, the gradient descent is more likely to get trapped in local minima - though the stochastic implementation of the gradient descent algorithm helps to mitigate this weakness;

if selected too large, the algorithm is less likely to converge on an accurate estimate of the target object. The number of iterations and number of voxel samples were selected while keeping in mind the trade-off between the accuracy gained from the additional iterations versus the added computational overhead. The training volumes were pre-processed in the same way as for the lollipop model. The GPMM was defined on a regular voxel grid with dimensions $48 \times 40 \times 208$. The first three principal components of the femur GPMM are shown in Figure 5.5.

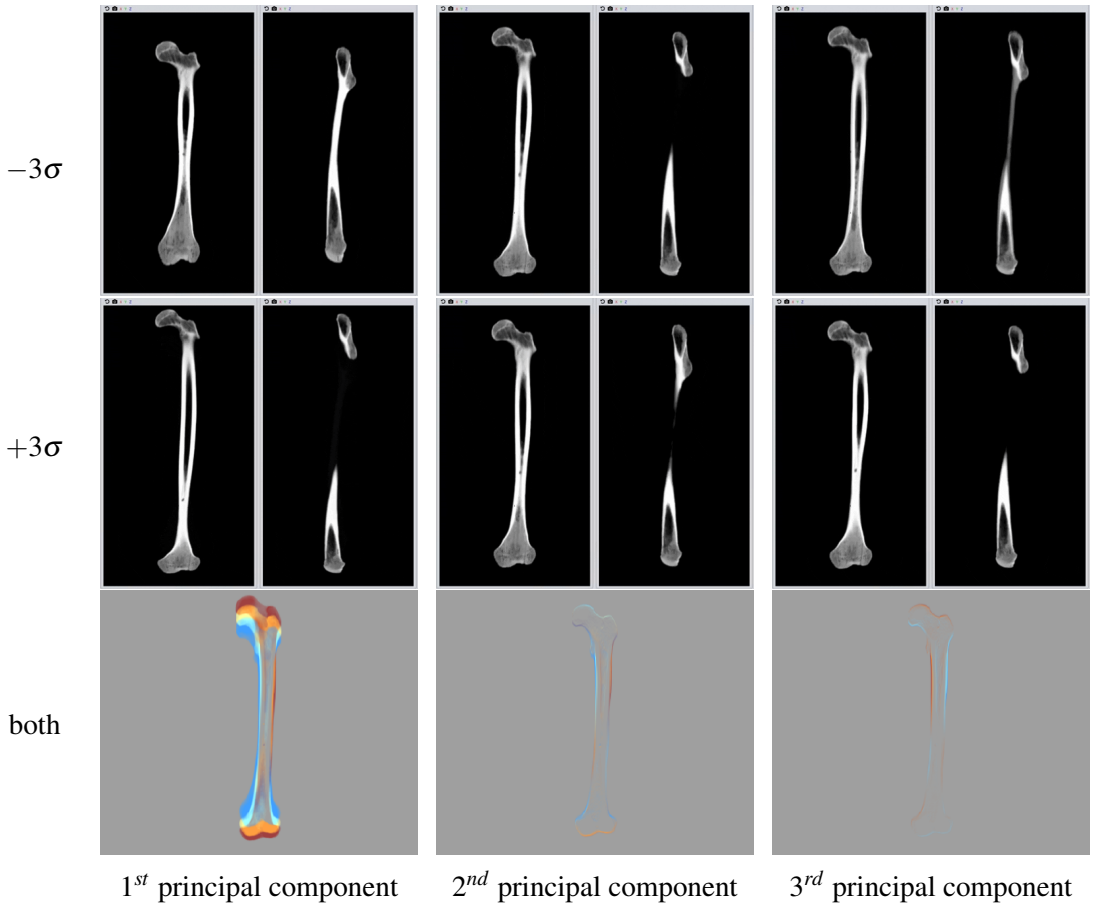


Figure 5.5 The first three principal components of the femur GPMM. The top row shows CT cross-sections (lateral and anterior-posterior) for each principal component with a parameter value of -3σ. The middle row shows CT cross-sections for each principal component with a parameter value of 3σ. Each item in the bottom row shows a difference image of DRRs corresponding to a parameter value of -3σ (blue) and 3σ (orange). The mean is shown in white.

5.3 A Bayesian MCMC approach to 2D-3D reconstruction

When reconstructing volumetric bone structures from a 2D X-ray image, the search for optimal model and pose parameters is particularly challenging. Due to the ill-posed nature of the problem, commonly used gradient-based strategies are often trapped in local optima and a pose initialisation step is therefore necessary to ensure a more robust parameter search. The Markov chain Monte Carlo (MCMC) algorithm is able to mitigate this susceptibility to local optima, since it does not rely on gradients. The MCMC-based approach proposed by [138] was adapted to the 2D-3D reconstruction algorithm being developed, thus allowing manual pose initialisation to be avoided. The subsections to follow describe the relevant details of their approach, and their application to the problem of 2D-3D reconstruction.

5.3.1 Previous work

The authors of [138] adapted the well-known MCMC algorithm to the problem of fitting a three-dimensional GPMM of the face to a single 2D image. Since GPMMs are inherently probabilistic, they framed the fitting problem as a Bayesian inferencing of the posterior distribution $P(\theta|I)$ from a prior distribution $P(\theta)$ provided by the model with parameters θ, and image likelihood $l(\theta;I)$:

$$P(\theta|I) \propto l(\theta;I)P(\theta). \tag{5.4}$$

Bayesian inferencing itself provides a number of advantages, the foremost of which is the ability to model uncertainty in data. This characteristic is well-suited to GPMMs because noisy observations can be used as a prior, and help navigate a non-convex model parameter space (see Chapter 2.4.4). Furthermore, the resulting inference is not a single maximum estimate, but a distribution of solutions. This capability provides the ability to not only propose the most probable single solution, but also provide information on its stability, and level of uncertainty.

The Metropolis-Hastings (MH) algorithm allows sample-wise evaluation of unnormalised posterior distributions. This characteristic is significant because the computation of the normalisation term required for the standard Bayes theorem, $P(I)$, is intractable [61]. Samples are drawn from the posterior distribution of the model parameters, given the target image $P(\theta|\tilde{I})$, using the Metropolis-Hastings (MH) algorithm, which iterates in a two-step process. First, a sample, $\theta_{t+1} \rightarrow \theta^{'}$, is drawn as θ' from a proposal distribution $Q(\theta'|\theta)$. The authors of [138] utilise a mixture of random walks for their proposals, though any method can be used to set the next sample. The random walk proposal is defined as a Gaussian update of

the existing sample

$$Q : \theta' = \theta + d \quad d \sim N(d|0, \sigma) \tag{5.5}$$

where σ is the standard deviation of the Gaussian distribution and d is the step-size. The primary advantage of the random walk is that it is unbiased, and it is this source of randomness that makes the parameter search less susceptible to local minima. Second, the random-walk proposal is verified using the Metropolis-Hastings rule. As the algorithm iterates, a Markov Chain is built up where a probability a is then assigned with which a given sample is considered to be equivalent to one drawn from the target parameter distribution $P(\theta|I)$. This probability is defined as

$$a = \min\{\frac{P(\theta'|\tilde{I})}{P(\theta|\tilde{I})}\frac{Q(\theta|\theta')}{Q(\theta'|\theta)}, 1\}. \tag{5.6}$$

The new state is then verified by evaluating the likelihood of the proposed parameter. Instead of using just a single likelihood function, the authors proposed to cascade the verification step, which is typically comprised of a set of different likelihoods. By doing this, different origins of information such as feature-based landmark distances and image pixel-errors could be incorporated into the Bayesian inference process. Ultimately, the authors determined their approach to perform well for the task of 3D face-model reconstruction from a single image, but this approach is still inefficient for less complex scenarios where the robustness and flexibility offered were unnecessary.

5.3.2 A Bayesian MCMC approach to 2D-3D reconstruction

The 2D-3D reconstruction problem is similar to the face-model fitting problem addressed by [138]. Using the notation established in Chapter 2.4, 2D-3D reconstruction is thus formulated as the Bayesian inferencing of a posterior distribution $P(\mathbf{w}|R_{DRR})$ from a prior distribution $P(\mathbf{w})$, provided by the model, and a likelihood of similarity $l(\mathbf{w}; R_{DRR})$ between the simulated X-ray image R_{DRR} and the target X-ray image R_{XRAY}:

$$P(\mathbf{w}|R_{DRR}) \propto l(\mathbf{w}; R_{DRR})P(\mathbf{w}) \tag{5.7}$$

Ultimately, T parameter samples are drawn, where each sample corresponds to some deformed instance of the deformable model projected to 2D space as a simulated X-ray image $R_{DRR} = \mathscr{P}(\mathbf{v}(\mathbf{w}_t))$. Following the approach of [138], inference is driven by the MH algorithm using the propose-verify scheme. The algorithm updates the estimated model parameters, either by accepting, $\mathbf{w}_{t+1} \rightarrow \mathbf{w}'$, or rejecting $\mathbf{w}_{t+1} \rightarrow \mathbf{w}_t$, a set of model parameters proposed by a Gaussian random walk. Each proposed set of model parameters is then accepted with

probability

$$a = \min\{\frac{P(\mathbf{w}'|R_{XRAY})}{P(\mathbf{w}|R_{XRAY})}\frac{Q(\mathbf{w}|\mathbf{w}')}{Q(\mathbf{w}'|\mathbf{w})}, 1\}. \tag{5.8}$$

The likelihood function serves as the Bayesian equivalent of a similarity measure, estimating the probability that the projected 2D image of the current model instance is the same as the target image. Thus, for image-based registration, any pixel-error measure will readily serve as a likelihood function (see Chapter 2.4.3). Multiple likelihood functions can be combined when additional target X-ray images are available:

$$l(\mathbf{w}|R_{DRR,1},\ldots,R_{DRR,C}) = \prod_{c=1}^{C} l_c(\mathbf{w}|R_{DRR,c}). \tag{5.9}$$

where $R_{DRR,c}$ is the DRR, and l_c the likelihood, for the c^{th} target. In a similar fashion, additional similarity measures could be incorporated. However, these details are beyond the scope of the experimental investigation presented in Chapter 6.

5.3.3 Implementation details

The 2D-3D reconstruction algorithm was developed for use in the experimental investigation presented in Chapter 6. The central aim of the investigation was not to find the best possible reconstruction accuracy, but rather to study relative reconstruction accuracy when comparing different calibration offsets. Therefore, a more complex parameter search strategy, such as one having multiple stages, resolutions, or likelihood functions, was avoided to mitigate introducing additional confounding factors. The Markov chain was initialised as the model mean parameters. Ultimately, for each reconstruction, 10^5 samples were drawn from the model, which, if accepted by the MH rule, updated the estimated CT volume. An L_2 error was used as the likelihood function, which was more readily parallelisable (see Section 5.4).

The random-walk step-size is context-sensitive and as such, step sizes were set empirically for each experiment. All simulated X-ray images were generated using the DRR renderer presented in Chapter 3. All necessary software libraries for Bayesian inferencing, MCMC-based sampling, as well as the propose-verify scheme were provided by the Scalismo framework *v*0.17.

5.4 Parallelisation of the 2D-3D reconstruction algorithm

Due to the iterative nature of image-based 2D-3D reconstruction, and the presence of many local optima, thousands of X-ray images need to be simulated to ensure the parameter space is

sufficiently explored, which causes a computational bottleneck. For clinicians, having to wait for more than a few minutes for the algorithm to finish is impractical [91]. Researchers have, therefore, explored ways to reduce the time it costs to renderer a DRR, either through using more sophisticated rendering techniques, or by optimising existing rendering algorithms for parallel processing on GPUs [101]. Considerable advances have been made, leading to a new computational bottleneck being identified: the process of sampling from the deformable model. Researchers have successfully migrated the model-sampling step onto the GPU - but thus far only for tetrahedral representations of bone. The work of [135] was the first example in literature of a 2D-3D registration algorithm applying hardware acceleration beyond the X-ray simulation step. Since then, [47], and [78] have improved upon this mesh-based approach, but so far none have considered the parallelisation of GPMM-based or voxel-based deformable models. For the 2D-3D reconstruction algorithm being developed, the method for sampling from the femur GPMM (developed in Section 5.2), and part of the likelihood evaluation, was implemented on a GPU.

5.4.1 Implementation details

The algorithm for sampling from a GPMM is well-suited to implementation on GPU, as the operations performed for each voxel are identical (Equation 5.3). Before performing these computations, the necessary mathematical objects are transported to GPU memory (Step 1 of Figure 5.6). These objects include the reference matrix, mean vector, covariance matrix, standard-deviation vector, and initial coefficients vector. Then, for each voxel-thread, a series of steps are followed. First, the values for the mean vector, eigenvalue (standard deviation), eigenvector (deformation vector), and model parameter coefficient relevant to the current voxel-index are looked up (for each principal component). Then, the resultant deformation vector is computed using Equation 5.3, the result of which is tested to determine whether it falls outside the 3D domain. If the deformation vector falls outside the domain it is discarded. Otherwise the 3D coordinate to which it points is computed using simple vector algebra. Finally, the HU value corresponding to the coordinates is looked up (interpolating if necessary) and outputted. The voxels outputted by the GPMM kernel are gathered into a single 3D memory object - a CT volume instance for the current set of model parameters - and subsequently rendered as a simulated set of pose and machine-specific parameters by the DRR rendering kernel (Step 2 of Figure 5.6). A detailed description of the DRR rendering kernel is provided in Chapter 3.5.

In addition to sampling from the GPMM, a portion of the likelihood evaluation step was also migrated to the GPU. Here, the L_2 likelihood function is well-suited to the GPU, whose parallel threads easily map to the identical arithmetic operations (subtraction and

exponentiation) performed for each pixel. The result is a 2D grid of values, which are easily stored in an image-optimised memory object, and transported back to the CPU. The CPU then parses this memory object and sums its pixel-values in order to obtain a single likelihood value.

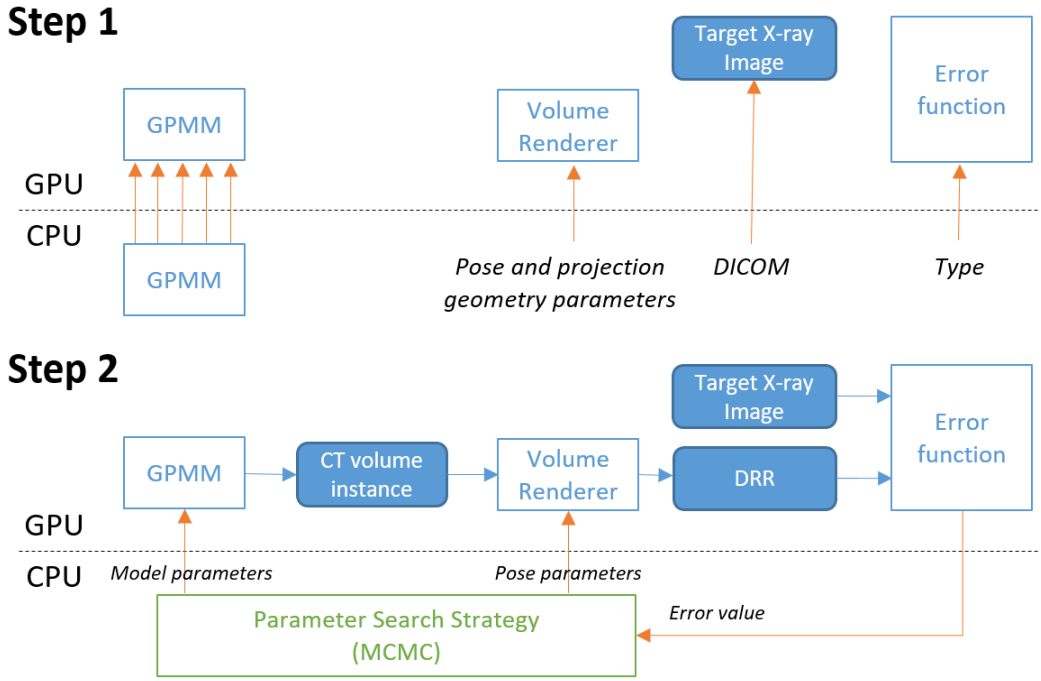


Figure 5.6 The GPU hosts a set of kernels corresponding to the GPMM, volume renderer, and error function. In the first step (Figure 5.6A), the components which comprise the GPMM (reference volume, mean vector, covariance matrix, standard deviations vector, and initial coefficients vector) are loaded into GPU memory together with the volume renderer settings (projection geometry parameters, sampling rate, interpolation mode), error function, and target X-ray image. The second step is iterative (Figure 5.6B), in which the parameter search strategy sends the model and pose parameters for the current iteration to the GPMM kernel and volume renderer kernel, respectively. The GPMM kernel provides a CT volume instance corresponding to the provided model parameters, which is subsequently rendered by the volume renderer kernel as a DRR, for a specific set of pose and machine-specific parameters. The error function kernel compares the DRR to the target X-ray image and yields a grid of error values, which is subsequently sent back to the parameter search strategy - the only component residing in CPU memory.

5.5 Discussion

In this chapter, a novel implementation of a 2D-3D reconstruction algorithm was described. The purpose of the algorithm is to facilitate an investigation into the influence of machine-specific DRR parameters on the accuracy of 2D-3D reconstructions.

First, the femur bone was modelled using a novel application of GPMMs to CT voxel-volumes. GPMMs provide a probabilistic interpretation of the conventional PDMs typically used for orthopaedic 2D-3D reconstruction. A GPMM with an analytically defined covariance matrix was used as a preliminary model with which to establish correspondence between the reference volume and the rest of volumes comprising the training set. The analytical GPMM was subsequently augmented with the statistical insights provided by said training set, thus highlighting one of the key advantages provided by GPMMs. Previously, GPMMs have only been applied to triangular or tetrahedral meshes. The GPMM framework was extended to accommodate voxel-based CT volumes. While the voxel-based modelling approach avoids segmentation errors, and holds promise for multi-object modelling, the size of the CT volume - and thus the number of objects which can be accommodated - remains limited due to the current memory capacity of commercial graphics hardware. However, it is assumed that this issue will become less of a factor in future, as hardware technology continues to develop. The femur training set was limited to a binary intensity distribution, which made the process of establishing correspondence simpler. As a consequence, the resulting deformable model had reduced flexibility, but was assumed sufficient for the forthcoming experimental investigation.

Next, the parameter search strategy component of the algorithm was implemented using a Bayesian MCMC approach. Its primary appeal is the ability to avoid local minima, a major challenge in the field, which may lead to inaccurate reconstructions. Additional advantages are offered, such as the ability to measure the stability of the resulting distribution of solution or the ability to combine multiple likelihood functions (but these extensions were not applied). The primary disadvantage of the MCMC-based approach is that it has a long run-time compared to established approaches (See Chapter 2.4.4). However, a faster run-time can be traded for reconstruction accuracy, since the level of accuracy achieved is proportional to the number of samples that are taken (given sufficient time, the MH algorithm is guaranteed to explore the entire parameter space). In some scenarios, this characteristic may be advantageous, as reconstruction accuracy can be made to scale with the computational resources that are available.

Finally, a significant reduction in computation time is made possible through a hardware-accelerated parallelisation of the model-sampling - and part of the likelihood estimation - processes. Almost the entire 2D-3D reconstruction algorithm was implemented on the GPU.

Preliminary testing demonstrated a considerable reduction in computation time as a result, though a quantitative evaluation will be presented in Chapter 6. In principal, the parts of the Bayesian inferencing framework which are still executed on the CPU could also be migrated to the GPU, leading to additional performance gains. This constitutes future work.

5.6 Conclusion

Having developed a complete 2D-3D reconstruction algorithm, the end-goal of the work is made possible, which is to study the effects of machine-specific calibration errors in the context of 2D-3D reconstruction. Chapter 6 presents such an investigation. First, the femur model is validated in terms of generality and specificity. Then, the reduction in computation time afforded by the parallelisation of the algorithm is measured. Finally, the influence of machine-specific parameters on 2D-3D reconstruction accuracy are investigated in both simulated and real-world settings.

Chapter 6

Evaluating machine-specific calibration for 2D-3D Reconstruction

6.1 Introduction

Machine-specific calibration settings affect both the resulting geometry and intensity profile of simulated X-ray images, and will thus also impact the accuracy of estimated volumes when used in a 2D-3D reconstruction algorithm. However, the importance of correct projection calibration is not discussed in literature, despite the fact that X-ray machines vary significantly with regard to their imaging setup. An experimental investigation was performed in order to better understand the influence of machine-specific parameters on 2D-3D registration accuracy (which includes both rigid alignment and non-rigid reconstruction).

Before the aim could be addressed, it was first necessary to ensure that the deformable model to be used (developed in Chapter 5) performed as expected. The first experiment accomplished this by measuring the ability of the model to approximate the same examples used to train it (specificity) and the model's ability to approximate unseen volumes (generality). Experiment 2 measured the reduction in computation time afforded by a GPU implementation of the deformable model versus one utilising a CPU. Having validated the model, Experiment 3 and 4 employed the Bayesian inferencing framework described in Chapter 5 to measure 2D-3D registration accuracy as a function of the same sequence of parameter offsets used in the investigation described in Chapter 4; Experiment 3 considered simulated target X-ray images while Experiment 4 considered real target X-ray images, which are more difficult to register. In doing so, the effect of confounders such as the model's generality and the presence of foreign objects in the target image could be studied

separately. The chapter concludes with a summary of the insights gained while performing the aforementioned experiments.

6.2 Experiment 1: Femur Model Validation

The basic concept of deformable 3D modelling is to establish the pattern of valid structural variations of a class of volumes, learnt from a training set (Chapter 2.4.1). The statistical analysis performed aimed to validate the femur GPMM trained in Chapter 5.2.3 by providing a parameterisation of this variability.

6.2.1 Methods

Two commonly used metrics to evaluate mesh-based shape models were adapted to validate the voxel-based GPMM in terms of generality and specificity [152, 112]. Specificity quantifies the ability of the model to generate instances similar to those available within the training set. Generality is the ability of the model to generate instances not explicitly provided by the training set, which is the goal of model-building [152]. Furthermore, it is important to measure this property because if the model is over-fitted to the training set, it may have excellent specificity but have a weakened ability to generalise to unseen examples. In the mesh-based paradigm, specificity is computed as the average distance (root-mean-square error (RMSE) or mean-absolute-distance (MAD) between randomly generated shapes and their nearest match in the training set. Generality is computed as the average distance between a certain number of unseen examples and their respective model reconstructions [152].

A similar protocol for computing specificity and generality was followed, except that image-based error using mutual information[1] (MI) as well as RMSE was computed, instead of Euclidean distance. The RMSE values were normalised such that they reflect greyscale intensity values instead of HUs. Specificity and generality were considered in relation to the number of images that the model was trained with. Doing so allows one to gauge the extent to which additional training examples improve the generality of the model, while simultaneously ensuring that the model has not been over-fitted.

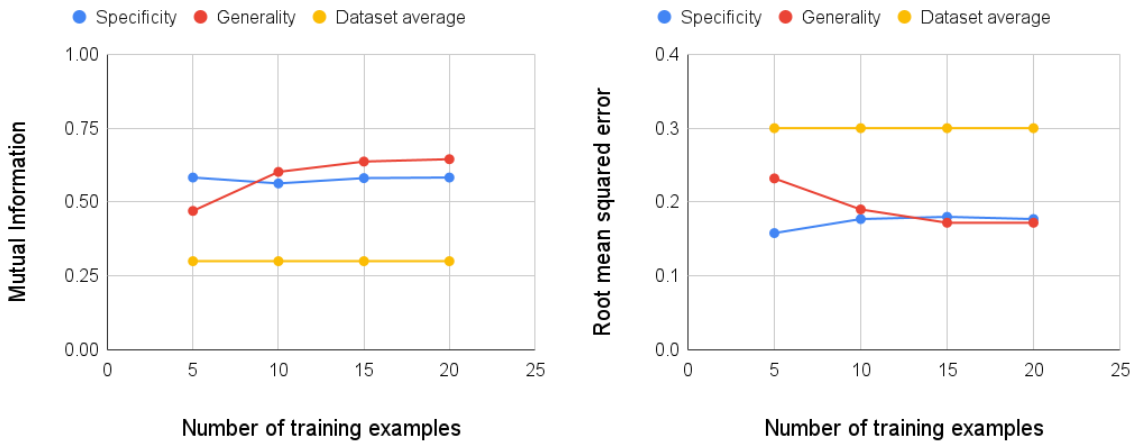
The accuracy of the registration procedure depends strongly on the choice of optimisation strategy and similarity measure. If the same registration procedure was used as the one for gauging model quality, the results would be biased, since one might merely over-fit the model to the same measure used to gauge its quality. Therefore, a different similarity

[1] Mutual information measures the amount of information that two data volumes in question contain about each other. The similarity is maximal when two volumes are identical, and decreases to zero in proportion to dissimilarity [193].

measure was employed to establish correspondence among the training volumes (MSE) than the ones used for model validation (RMSE and MI). In both cases, a stochastic gradient descent strategy was used, having a step-size of 0.05 and 350 iterations. Values for RMSE and MI were plotted as a function of the number of images that the model was trained with, $n = \{5, 10, 15, 20\}$. The Scalismo framework v0.17 [159] provides necessary tools for training a GPMM using surface or volumetric meshes [159]. These methods were adapted to accommodate voxel-based volumes.

6.2.2 Results and Discussion

The specificity and generality of the GPMM was plotted as a function of the number of training examples used, measured using MI in Figure 6.1a, and RMSE in Figure 6.1b. The mean MI and mean RMSE of the images that comprise the training dataset were included as a baseline reference. Furthermore, it should be noted that larger MI values reflect a higher similarity, while the inverse is true for RMSE values.



(a) (b)

Figure 6.1 Specificity and generality, measured in terms of mutual information (MI) and root mean squared error (RMSE), as a function of the number of training examples used to build the model. Larger values are favourable for MI (shown on the left) while lower values are favourable for RMSE (shown on the right). For both plots, the "dataset average" indicates the average MI or RMSE value of the set of training examples (aligned but not deformed), which serves as an upper or lower bound against which to judge the achieved specificity and generality values of the trained models.

By qualitatively evaluating the trend of MI values in Figure 6.1a and comparing them to the RMSE values in Figure 6.1b, the optimal number of training examples is assumed to

be around 15, providing an acceptable compromise between the specificity and generality values. These two metrics converge at around $8-10$ training examples - when considering both RMSE and MI values - and possibly diverge again at around 20 training examples (only in terms of MI), at which point the model starts to become over-fitted. However, it should be noted that this optima will not necessarily be valid for larger volume dimensions, or when training with non-binary volumes. The specificity falls within the average MI of the training examples, which suggests that randomly generated images are fairly typical when compared to the training dataset, though slightly less varied. Furthermore, as predicted, the value for generality is better than for specificity, since, in this case, the similarity measure was optimised.

In, conclusion, the femur GPMM was validated using conventional mesh model metrics, adapted to voxel-based volumes. First, through visual inspection and using the specificity metric, it was confirmed that randomly generated model deformations maintain their validity as femur structures. Second, the model was fitted to unseen femur targets with an average RMSE value of 0.17, and an average MI value of 0.64. The experimental results indicate that voxel-based models, in principle, are a valid alternative to conventional mesh-based approaches. However, given the context-specific nature of these results, robust conclusions cannot be made since no other study exists with which to compare them to. Regardless, the model was assumed to be valid for use in the subsequent experimental investigations.

6.3 Experiment 2: Hardware-accelerated model sampling

Previously, researchers have successfully parallelised a model-sampling approach for 2D-3D reconstruction [47]. However, so far none have done so for GPMM-based or voxel-based deformable models. In this experiment, the reduction in computation time afforded by a GPU implementation of a voxel-based GPMM was investigated. The size of the GPMM's covariance matrix grows exponentially with the number of examples used to train it (the rank of the model). The size and manner in which memory objects are accessed by the GPU are known to influence computation time [3]. Furthermore, the number of instructions that need to be executed to obtain each voxel deformation is also proportional to rank, making it an interesting model property against which to measure computation time.

6.3.1 Materials and Methods

The rate at which samples can be drawn and rendered as a simulated X-ray image was measured using the GPMM that was validated in Experiment 1. The GPMM, having a

rank of 19 (corresponding to the number of training volumes), was used to generate 5 new GPMMs; the first has 18 rows truncated from its covariance matrix ($\bar{v}_{rank=1}$) and each subsequent GPMM had two fewer rows truncated from its covariance matrix: $\bar{v}_{rank}$ where rank $= \{1,3,5,7,9\}$. Then, for each of the 5 models, the following steps were followed. Model instances, each having a random pose and model parameter configuration, were continuously generated and rendered, resulting in a sequence of simulated X-ray images. The number of simulated X-ray images rendered was recorded each second, for 60 seconds, and averaged, resulting in an average frames-per-second (FPS) value. Finally, the process was repeated for a CPU implementation[2]. The resulting set of 2×5 data-points were plotted. Furthermore, the number of parallel threads used by GPMM kernel are roughly equivalent to those of the DRR renderer kernel (in the order of 10^5). The results were computed using an Intel Xeon 3.5 GHz CPU with a 16-GB RAM, and a NVIDIA GTX 970 GPU with 8 GB of RAM.

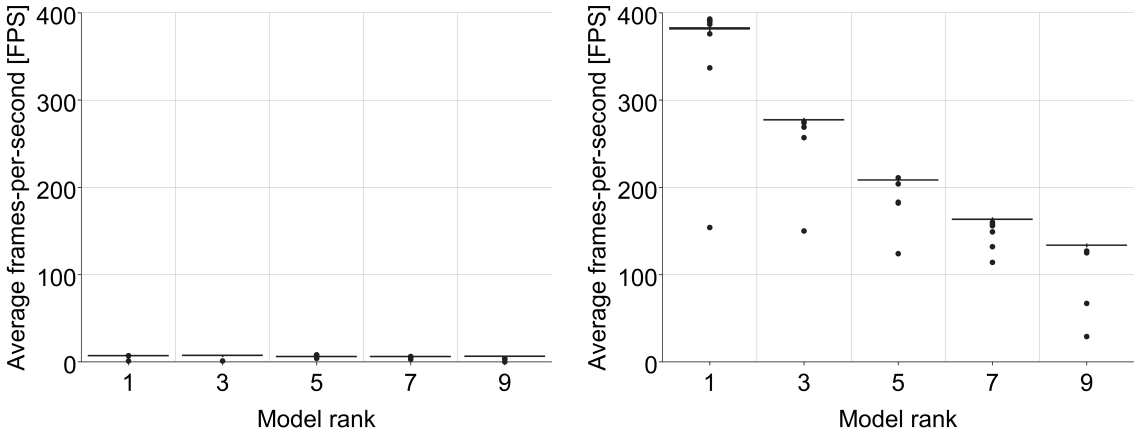
6.3.2 Results and discussion

Figures 6.2a and 6.2b show the average FPS as a function of the GPMM rank for the CPU and GPU, respectively. It can be seen that the time it takes the GPU to sample a deformed instance from the model and generate a simulated X-ray image scales exponentially as the rank of the model increases linearly. However, in the case of the CPU, the rank of the model makes little difference to the resulting FPS. This is because the computational bottleneck of the DRR rendering step overshadows any reduction in computation time that would be afforded by using a smaller model. Generally, the FPS variance for both CPU and GPU implementations remained stable as a function of model rank. However, outliers corresponding to bouts of poorer performance did occur regardless of model rank, or computational architecture. These are likely due to the operating system (see Chapter 4.5). Finally, it is worth reiterating that the number of voxels comprising the model correspond to the number of parallel threads assigned to the GPMM kernel running on the GPU. Therefore, if the number of desired parallel threads exceeds that capacity of the GPU used (typically in the order of 10^4), the GPU will split the computations into separate batches, resulting in an additional computation time overhead.

A hardware-accelerated implementation of the 2D-3D reconstruction algorithm enabled a hundred-fold reduction in computation time, compared to a CPU-only implementation. When compared to the results presented in Chapter 4.5, the process of sampling from the model lowered the average FPS by at least 8% - even when the model has only a single

[2]The CPU implementation was not explicitly optimised, but rather an "out-of-the-box" OpenCL method enabling compilation and execution using the CPU.

principal component. It is thus evident that the model-sampling step has taken the place of the X-ray simulation step as the computational bottleneck of the 2D-3D reconstruction algorithm. Other than the rank of the model, another likely contributor to the bottleneck is the larger number of multi-dimensional memory objects that are accessed by the GPMM kernel. Future research efforts should consider memory access optimisation, which would further reduce the computation time of the 2D-3D reconstruction algorithm [3].



(a) CPU (b) GPU

Figure 6.2 Two box-plots showing the difference in computation time, measured in FPS, of the 2D-3D reconstruction algorithm when executed on the CPU (left) and GPU (right). One "frame" is equivalent to the process of obtaining a deformable model instance and rendering it as a simulated X-ray image.

6.4 Experiment 3: 2D-3D registration of simulated targets

The experimental investigation in Chapter 4 determined that machine-specific calibration parameters have a notable impact on DRR fidelity. However, the ability to simulate X-ray images is only used as a means for reconstructing 3D bone structures. Therefore the aim of this experiment was to investigate the influence of the same parameters on 2D-3D reconstruction accuracy, namely the source-to-object distance, d_{S2O}, and incident energy, I_0. The d_{S2O} parameter is the distance from the X-ray source to the imaged object, and affects both the global scale and local depth-dependent distortions of the resulting 2D projection. The I_0 parameter is the incident energy of each modelled X-ray and is directly proportional to the resulting pixel-intensities of the DRR. The experiment focused on simulated registration targets rather than real ones. In this simplified context a perfect registration was possible, since the 3D object was used to obtain the 2D target using the DRR renderer. Doing so

allowed the added complexity of secondary imaging effects, which are not accounted for by the renderer, or the inability of the deformable model to fully account for 3D appearance, to be avoided. These complexities were addressed in Experiment 4.

6.4.1 Materials and Methods

Two types of registration were considered: rigid 2D-3D alignment - where only pose parameters were searched for - and non-rigid 2D-3D reconstruction - where pose and deformable model parameters were searched for simultaneously. For the 2D-3D alignment, the femur CT image described in Chapter 5.2.3 was used as the 3D object. Ten DRRs were generated (shown in Figure 6.3a) using sets of randomly generated pose parameters (listed in Appendix A). The parameters were generated using a uniform distribution and bounded to reflect the kinds of deviations that would be encountered in a clinical setting; rotation angles ranged between -45 and $+45$ degrees, while 2D translations (x and y) ranged between -250 and $+250$ mm. Each DRR was in turn used as a target image for the parameter search. For the non-rigid reconstruction part of the experiment, the model validated in Experiment 1 was used both to generate reconstruction targets, and as the deformable model. Ten DRRs - shown in Figure 6.3b - were generated using randomly generated sets of pose and deformable model parameters, and in turn used as a target image for the parameter search. The parameters were generated again using a uniform distribution, and ranged between -3σ and $+3\sigma$, corresponding to the 5 principal components of the deformable model.

For either type of registration, alignment or reconstruction, a set of 10 registrations were performed, each corresponding to a different DRR renderer calibration. The DRR renderer was calibrated first for a range of incident energy offsets, $I_0 = \{-1.0, -0.5, 0.0, 0.5, 1.0\}$, while the S2O distance was set to its correct value, $987mm$, and then for a range of S2O distance offsets, $d_{S2O} = \{-250, -125, 0, 125, 250\}$, while the incident energy was set to the calibrated value determined in Chapter 4. Thus, a total of 200 registrations were performed (10 different target images $\times$ 10 different calibrations $\times$ 2 types of registration). The conventional approach to 2D-3D reconstruction is to actively optimise the model parameters and cease to do so after a predefined similarity threshold is reached. However, to enable a comparison between calibration offsets, an identical pre-determined number of samples were taken for each reconstruction. An MCMC-based parameter search strategy was used, employing the MH algorithm. Following the approach of [138], 10^5 samples were drawn while the pose step-size and model step-size were determined empirically, having a value of 5 and 20, respectively. These hyper-parameters were kept constant, and the algorithm was kept simple, using a single L_2 likelihood function for the error measure.

The accuracy of a given alignment or reconstruction was measured as an overlap percentage of the "moving" DRR image with the target DRR using 2D Dice-coefficients (DCs) [155]. Each pixel of both the moving and target images were therefore designated a label corresponding either to bone or non-bone background. In addition, 3D-DCs were computed between the ground-truth volumes (used to generate the ground-truth DRR) and the estimated volumes. Doing so provided insight into the potential incongruities between the 2D and 3D domain, due to the fact that multiple solutions may exist when projecting 3D information to 2D. Since the aim of the experiment was to study the effect that DRR calibration may or may not have on registration accuracy, the *relative* registration accuracy among a set of calibrations was considered more important than the degree of accuracy of any one registration, or the specific parameter search strategy that was used. For this reason, the way in which the MCMC algorithm explores the parameter-space was kept relatively straightforward, avoiding any confounding factors that may be introduced by more advanced strategies such as multi-scale, multi-stage, or multi-likelihood approaches.

Finally, due to the limited sample-size, median values were considered to be more important since they are significantly more robust to noise than mean values[3].

6.4.2 Results and discussion

Two-dimensional-to-three-dimensional registration accuracy was plotted as a function of d_{S2O} offset, shown in Figure 6.4; and I_0 offset, shown in Figure 6.5. For the d_{S2O}, both the alignment and reconstruction accuracy corresponding to the correct calibration value resulted in the greatest registration accuracy. A linear decrease of registration accuracy in proportion to the calibration offset magnitude was observed, where an offset of 125 mm translated to a 2D-DC loss of $2-8\%$ for 2D-3D alignment, and $4-10\%$ for 2D-3D reconstruction. This result is consistent with the pattern observed for DRR fidelity in Chapter 4. Offsets which resulted in smaller S2O distances resulted in a greater loss of registration accuracy, which was also in accordance with the DRR fidelity results. When d_{S2O} is reduced, object projections become larger. In turn, anisotropic distortions are pronounced and cause greater disparities with the ground-truth than if d_{S2O} is increased. In the latter case, projections approximate a parallel projection - as was already the case along the scan-direction. The same trends were observed for 2D-3D reconstruction. For the latter, increased variance was observed as a larger inter-quartile range (IQR) for S2O reduction. The general registration accuracy for 2D-3D alignment was considerably better compared to 2D-3D reconstruction,

[3]The median has a "breakdown-point" of 50%, which means that more than half of a set of measurement samples need to be outliers before the median can be moved outside the range of non-outliers. By contrast, the mean has a breakdown-point of 0, which means that as a single large outlier can distort it significantly.

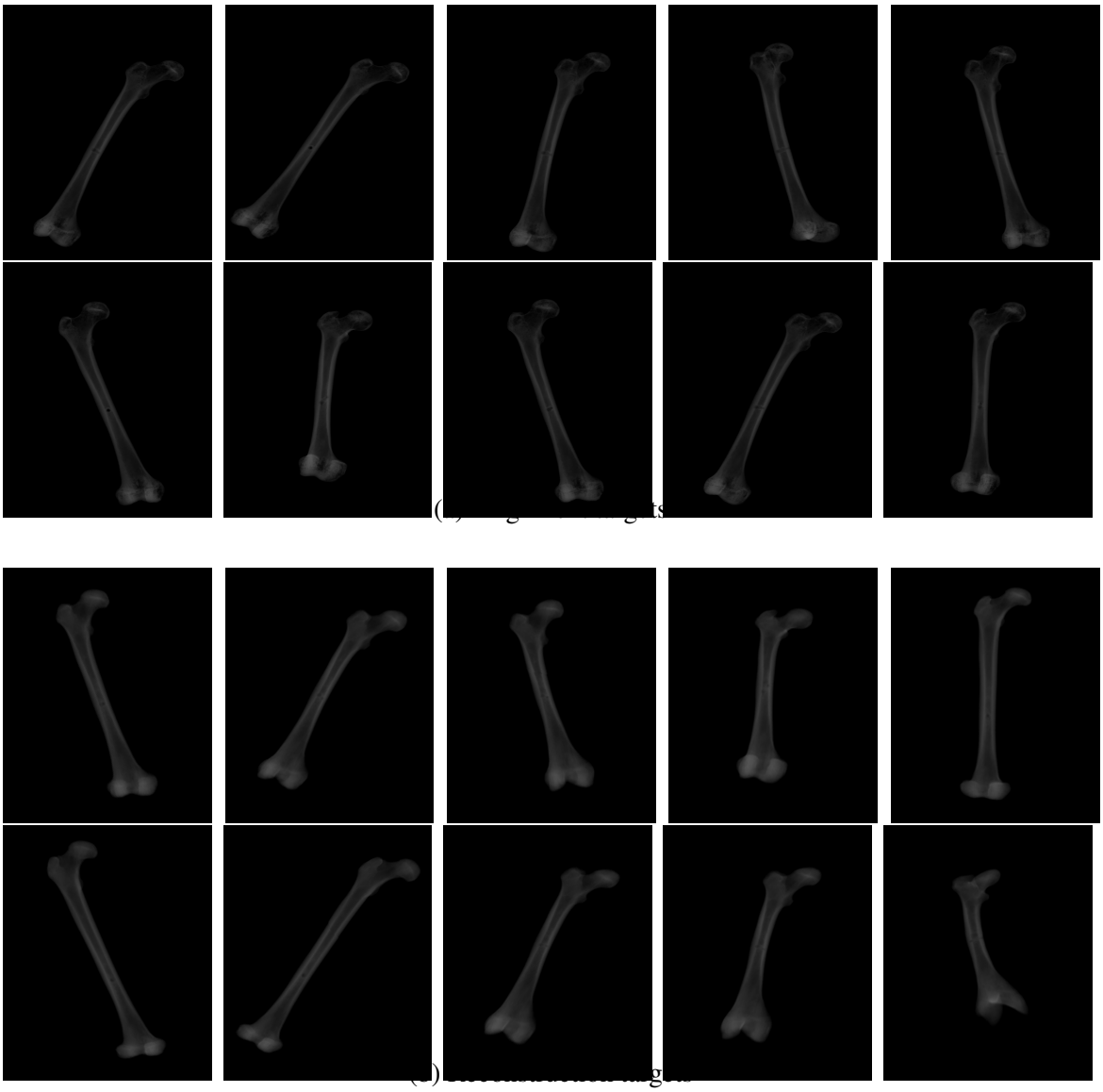


Figure 6.3 Simulated targets for 2D-3D alignment (above) and 2D-3D reconstruction (below). The targets were generated using sets of randomly generated pose parameters, which include roll, pitch, yaw, translation in the x-direction and translation in the y-direction, and model parameters, which correspond to the 5 principal components of the femur model.

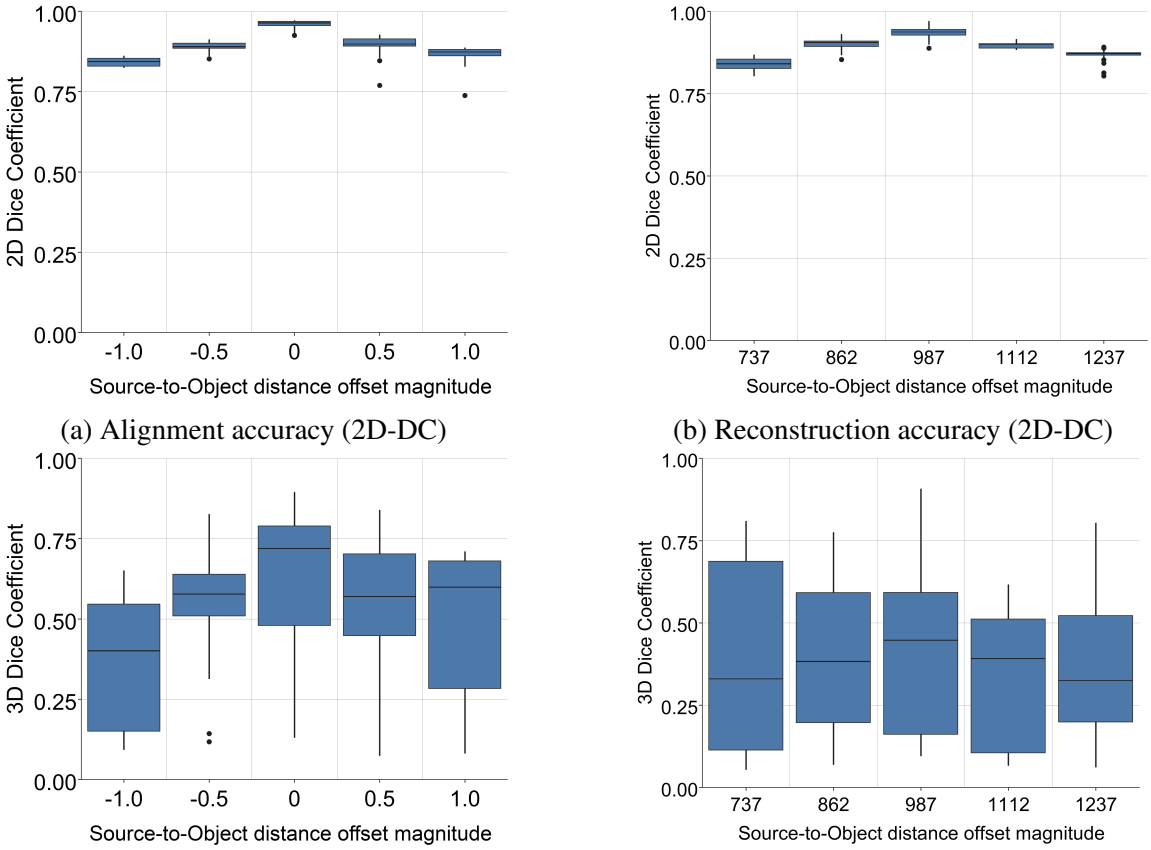
which was to be expected since there were fewer parameters to optimise. Aside from more subtle local depth-dependent distortions, the d_{S2O} had a scaling effect on the imaged volume, which amplified the ambiguity of the 2D projections. This characteristic was compounded in the case of non-rigid reconstruction because the model was able to compensate for this scaling mismatch, for example, by varying the length of the reconstructed femur (the first principal component of the model). Therefore, one may surmise that d_{S2O} mis-calibrations

increased the likelihood of the parameter-search getting trapped by local minima. The 3D-DC results were more noisy than for the 2D-DC results, which was expected due to the additional dimension that is considered. For example, it was possible for objects to appear correctly aligned when in fact they were not, since a single 3D pose can account for multiple 2D projections. While the results were very noisy, if one considers median accuracy as well as the best-case accuracy, the 2D-3D reconstruction results showed the same trends as for the 2D-DCs in all but one of the calibration offset magnitudes. The reconstruction results were even more noisy. However, the best-case DC values are still consistent with the trend observed for the alignment results.

Despite varying the different calibration offset magnitudes for I_0, approximately the same accuracy and variance was recorded for each (around 96%), which was comparable to the accuracy attained for the correctly calibrated d_{S2O} parameter. The results for I_0 therefore indicate that I_0 calibration had little to no effect on registration accuracy. In some cases, an incorrectly calibrated I_0 resulted in superior accuracy. In these cases, the difference in intensity values between the target and the moving image were exaggerated, resulting in a "steeper" L_2 likelihood function. It would therefore be possible for random pose or model parameter perturbations to be accepted more frequently provided that the step-size is not too large (see Equation 5.6). In these cases, more accurate registrations may be obtained using fewer iterations. While local minima would also be exaggerated, their ability to impede the optimisation may still be overcome by the stochasticity of the algorithm. Nonetheless, for the 3D-DCs, the best-case accuracy was still achieved when I_0 was correctly calibrated.

Aside from relative registration accuracy, variance in accuracy was also considered. Variance here can be interpreted as the capacity to consistently determine an accurate solution, irrespective of the registration target. This quality will be referred to as robustness. The registration accuracy variance corresponding to the correctly calibrated renderer (shown in Figure 6.4) was noticeably smaller compared to others. Furthermore, this observation held for both geometry and intensity-related mis-calibrations. In the case of d_{S2O}, the DC is intrinsically reduced because, due to the distance-dependent magnification of 2D projection, the area covered by the silhouette cannot be equal to that of the correctly calibrated target (when the correct pose and model parameters are used). However, in the case of I_0, the area covered by the silhouette only changed because of differing pose and model parameters. It was therefore also reasonable to surmise that the correctly calibrated renderer had a positive impact on the robustness of the parameter search strategy.

The sample-size[4] of any one registration accuracy measurement was limited to $n = 10$ simulated targets due to computational resource constraints, and $n = 5$ real targets due to data availability constraints. It was assumed that these quantities were sufficient for drawing conclusions about the accuracy and robustness of the algorithm, though strong conclusions are best avoided - especially since other studies with which to compare these results do not exist.



(a) Alignment accuracy (2D-DC)

(b) Reconstruction accuracy (2D-DC)

(c) Alignment accuracy (3D-DC)

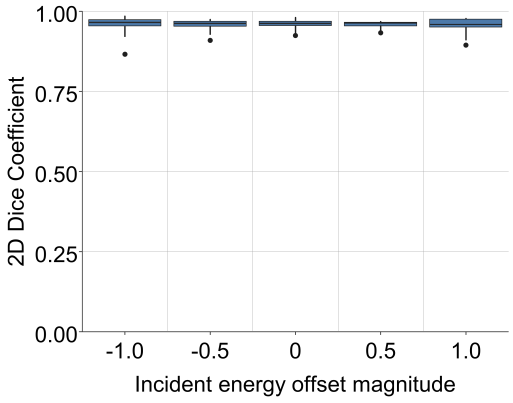
(d) Reconstruction accuracy (3D-DC)

Figure 6.4 Box-plots for the simulated targets, showing 2D-3D alignment accuracy (left column) and reconstruction accuracy (right column) when measured using a 2D Dice coefficient (DC) (top row) and 3D-DC (bottom row), both as a function of S2O distance d_{S2O} offset.

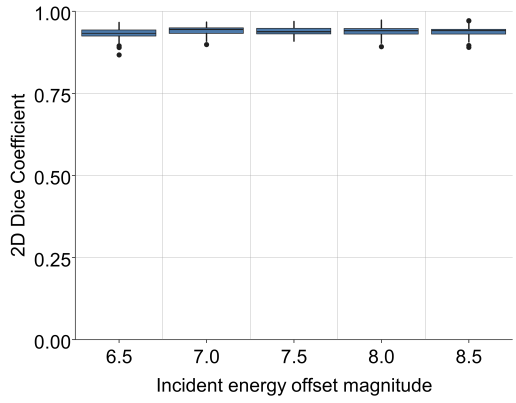
6.5 Experiment 4: 2D-3D registration of real targets

The goal of this experiment was similar to that of the previous one; to study the influence DRR renderer mis-calibration on registration accuracy, for two registration scenarios: 2D-3D

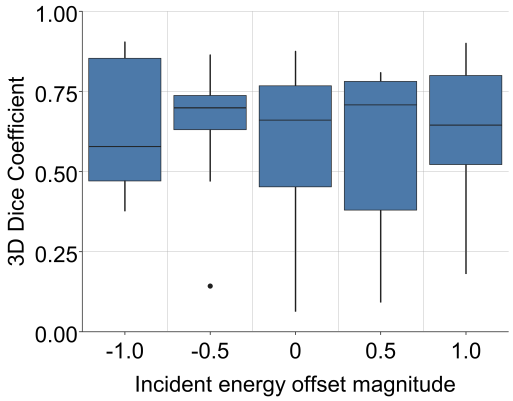
[4]Note that here "sample-size" refers to the number of measurements taken, not the number of Markov chain samples.



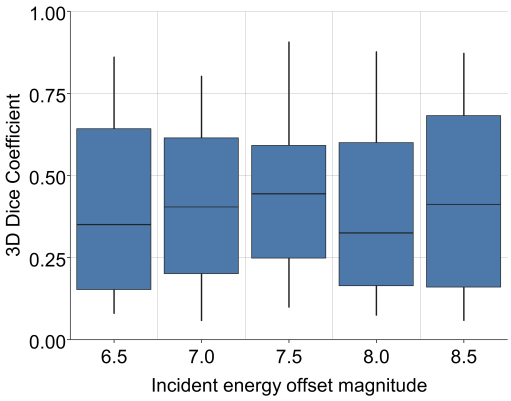
(a) Alignment accuracy (2D-DC)



(b) Reconstruction accuracy (2D-DC)

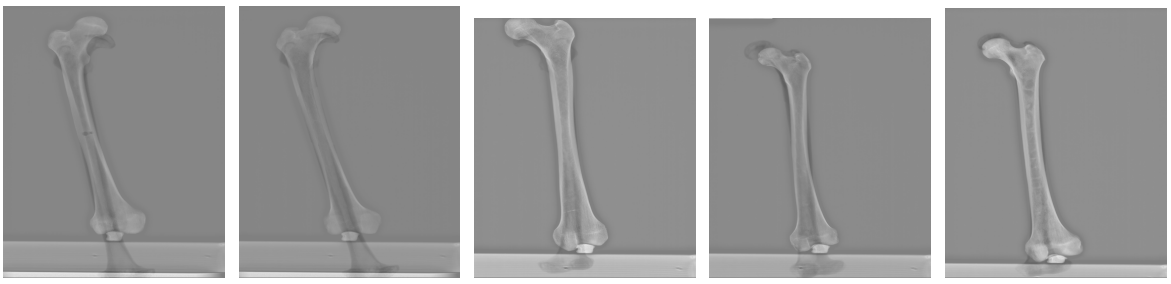


(c) Alignment accuracy (3D-DC)



(d) Reconstruction accuracy (3D-DC)

Figure 6.5 Box-plots for the simulated targets, showing 2D-3D alignment accuracy (left column) and reconstruction accuracy (right column) when measured using a 2D Dice coefficient (DC) (top row) and 3D-DC (bottom row), both as a function of incident energy I_0 offset.



(a) Femur 1 (b) Femur 2 (c) Femur 3 (d) Femur 4 (e) Femur 5

Figure 6.6 Difference images between each of the 10 simulated X-ray image targets and their respective estimates, when using the correct machine-specific parameter calibration.

alignment, and 2D-3D reconstruction. However, instead of using DRRs as targets, real X-ray images of dry-bone femurs were used. While soft-tissue noise, and significant occlusions were still controlled for, objects other than the bone-of-interest, and photon noise were

present, adding complexity to the problem, and thereby more closely resembled a real-world scenario. In contrast to Experiment 3, 3D-DCs were not considered. In this specific experimental setup, their computation risked introducing additional confounders as a result of the additional 3D-3D alignment step that would first have to be performed after each 2D-3D registration. The selection of alignment algorithm and error measure for this task - as well as the selection of algorithm hyper-parameters - prevented any insights or conclusions made from being generalised.

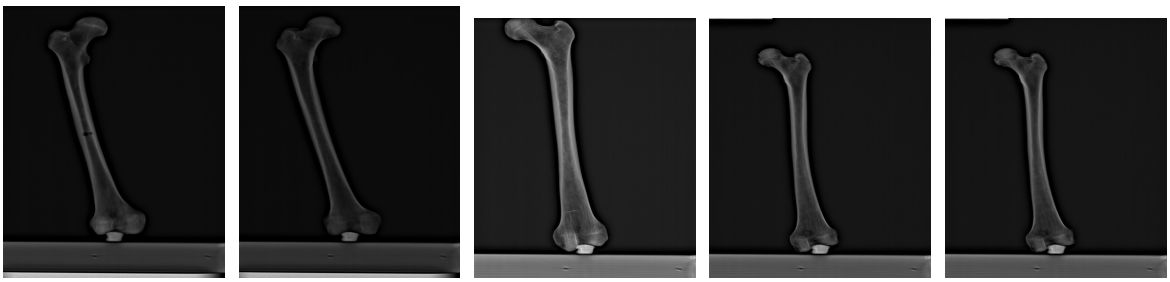
6.5.1 Materials

For the rigid alignment, a dataset comprising 5 sets of 1 × 3D CT image and 1 × corresponding 2D X-ray image (lateral orientation) was used. A total of 5 previously imaged adult femur bone CT volumes were obtained with ethical approval from the Faculty of Health Sciences Human Research Ethics Committee (HREC) of the University of Cape Town (reference HREC 258/2016). Each femur bone was obtained from a dissected leg, with the surrounding tissues removed, collected from a cadaveric specimen at the Division of Clinical Anatomy and Biological Anthropology, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, South Africa. Having been anonymised, information regarding age-range and sex were unknown. The 2D X-ray images were imaged using an EOS X-ray imaging system [48] at IMT Atlantique, INSERM in Brest, France, each comprising between 1764 × 2925 and 1764 × 3037 pixels with a pixel-spacing of 0.179363. Each X-ray image was saved in DICOM format, along with machine-specific parameters. The CT images were also imaged at IMT Atlantique, INSERM in Brest, France, each comprising between 626 and 691 slices, each comprising 768 × 768 pixels. Each CT image has a uniform trans-axial pixel size of 0.260417 mm, and slice-spacing of 0.8 mm, thus resulting in a voxel size of 0.260417 × 0.260417 × 0.8 mm. The imaged femur bones were originally placed in a water-bath to simulate the presence of soft tissues surrounding the bone. Since the research-at-hand only considered the simplified setting of dry-bone specimens, the femur region in each CT slice was hand-segmented. A CT volume was extracted for each image and cropped to only include the bounding box of the femur. Regions which previously corresponded to water were assigned a HU value of -1000, which corresponds to air. Finally, the CT volumes were each subsequently resampled to a uniform voxel size of 1 mm using a Lanczos filter, resulting in a set of volumes with dimensions ranging between 90 × 70 × 414 mm and 200 × 200 × 553 mm. All image processing was performed using the Amira suite [158].

The model validated in Experiment 1 was used as the deformable model for the non-rigid reconstruction part of the experiment. The set of 5 previously described X-ray images were used as reconstruction targets.

6.5.2 Methods

The same procedure was followed as was used for the previous investigation, but instead of 10 randomly generated DRRs as targets, 5 real X-ray images were used. In each X-ray image, one of 5 dry-bone femur specimens was imaged (Figure 6.7). Each femur was also imaged using CT, which served as an input to DRR renderer for rigid alignment, and for projecting the ground-truth when performing the non-rigid reconstruction. Where appropriate, observations were made in reference to Experiment 3 in order to gain insight into the additional complexity of working with real-world images.



(a) Femur 1 (b) Femur 2 (c) Femur 3 (d) Femur 4 (e) Femur 5

Figure 6.7 Five real target X-ray images, taken from 5 dry-bone femur specimens. Each femur was also imaged using CT.

6.5.3 Results and discussion

Two-dimensional-to-three-dimensional registration accuracy as a function of parameter calibration offset are provided for S2O distance d_{S2O} and incident energy I_0 in Figure 6.9. Difference images for the correct parameter calibrations are provided in Figure 6.8. As was the case for the simulated targets used in Experiment 3, the best registration accuracies were obtained for the correct calibration, with linearly decreasing accuracy as the calibration offset magnitude was increased. This observation was true when considering median, IQR, best-case accuracy, and worst-case accuracy[5]. An offset of 125 mm translated to a 2D-DC loss of $3-14\%$ for 2D-3D alignment, and $2-7\%$ for 2D-3D reconstruction. However, larger errors and variances were recorded for a reduction - rather than an increase - in the d_{S2O}. This observation indicates that the distortion due to anisotropic magnifications caused a greater discrepancy (compared to the correct calibration) than for the parallel projection that was approximated as d_{S2O} was increased. Also as before, the 2D-3D alignment had considerably higher accuracy than the 2D-3D reconstruction results, since the algorithm only had pose

[5]For the worst-case accuracy, outliers were not considered.

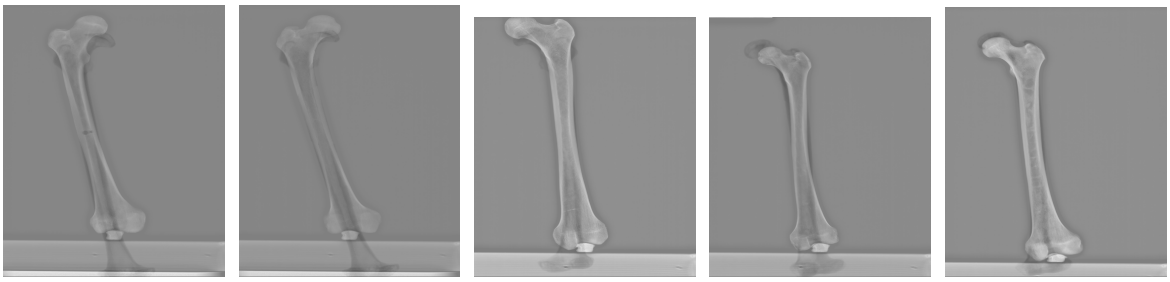
parameters and no model parameters to optimise. The same trends were noted for the 2D-3D reconstruction results as for the simulated results. This observation was true for the median, IQR, and worst-case results, but curiously not when an offset magnitude of $+150mm$ for d_{S2O}. In this case, the best-case accuracy and IQR was superior. It should be noted, however, that despite the insights made possible through using this unique CT-X-ray-image-pair dataset, only a limited number of samples (N=5) were available. A single sample could therefore have a disproportionate effect on group-level statistics, and robust conclusions are best avoided until more data becomes available. Nonetheless, as a first exploration, valuable insights were made possible.

The I_0 parameter had little to no effect on registration accuracy, whether for an alignment or reconstruction, and the reconstruction results produced worse accuracies - as was expected - due to the additional model parameters that had to be optimised for. These observations were also made for the simulated targets of Experiment 3. Interestingly, the results for the real targets were seemingly superior to those of the simulated targets - both in terms of Dice coefficients and variance. However, it is likely that an insufficient number of samples were taken for the precise distribution of registration accuracies to be measured per offset (10 samples as opposed to 5 samples). Furthermore, it was sometimes the case that incorrect calibrations produced better registration accuracy than for the correct calibration, lending further credence to the "exaggerated slope hypothesis" discussed in Experiment 3. Nonetheless, the best-case reconstruction accuracies were recorded when using the correct I_0 calibration.

In general, the correctly calibrated alignment and reconstruction results also had a smaller variance as before, indicating that the parameter search was able to perform in a more robust manner. The presence of other objects proved to be a major challenge for the 2D-3D reconstruction. In each of the reconstructions, shown in Figure 6.8, the algorithm overestimated the length of the femur, which is likely due to the platform on which the dry-bone femurs were mounted. One can surmise that the same issue would be encountered - and compounded - when soft-tissue and neighbouring bones were also present in the target X-ray image.

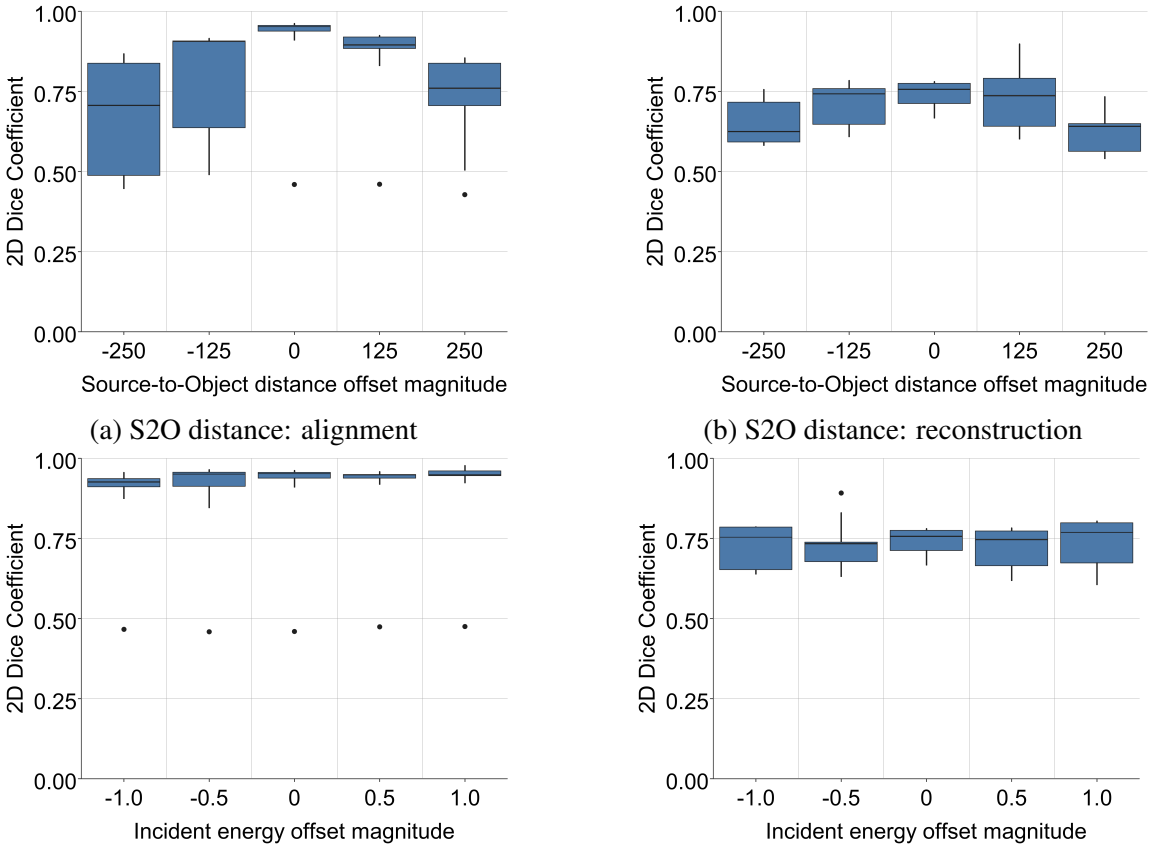
6.6 Discussion

The femur GPMM was validated using conventional mesh model metrics, adapted to voxel-based volumes. First, through visual inspection and using the specificity metric, it was confirmed that randomly generated model deformations maintain their validity as femur structures. Second, the model was fitted to unseen femur targets with an average RMSE value



(a) Femur 1 (b) Femur 2 (c) Femur 3 (d) Femur 4 (e) Femur 5

Figure 6.8 Difference images between each of the 5 real X-ray image targets and their respective estimates, when the correct machine-specific parameter calibration were used.



(a) S2O distance: alignment

(b) S2O distance: reconstruction

(c) Incident energy: alignment

(d) Incident energy: reconstruction

Figure 6.9 Two-dimensional dice coefficient measurements when using real target X-ray images. The top row corresponds to the S2O distance parameter, d_{S2O} while the bottom row corresponds to the modelled incident energy parameter, I_0. The left column corresponds to the results for 2D-3D alignment (rigid 2D-3D registration), while the right column corresponds to the results for 2D-3D reconstruction (non-rigid 2D-3D registration).

of 0.17, and an average MI value of 0.64. The experimental results indicate that voxel-based models, in principle, are a valid alternative to conventional mesh-based approaches. However, given the context-specific nature of these results, robust conclusions are best avoided since no other study exists with which to compare them to. Regardless, the model was assumed to be valid for use in the subsequent experimental investigations.

The validated model was used to measure 2D-3D registration accuracy as a function of parameter offsets. The same sequence of parameter offsets for the investigation in Chapter 4 was used. First, simulated X-ray image targets were considered before a similar investigation was performed, using real targets. Doing so allowed the effect of confounders such as the model's generality and the presence of foreign objects in the target image to be studied separately. Secondary imaging effects were not accounted for by the renderer. Therefore, inherent discrepancies are always present between a real X-ray image and its DRR estimate - even if the CT volume is perfectly registered. Moreover, for the case of deformable registration, the model is unable to account for all possible variations in bone appearance, and discrepancies will therefore exist for unseen targets (the reader is referred to Chapter 2.4.4 for an in-depth discussion on this topic). When comparing the accuracies resulting from different offset magnitudes, relative accuracy was considered more important than absolute accuracy since it provides better insights regarding correct calibration than the quality of the parameter search strategy itself (the parameter search strategy was kept constant for each of the calibration offsets). For the d_{S2O}, both the alignment and reconstruction accuracy corresponding to the correct calibration value resulted in the greatest registration accuracy. A linear decrease of 2D-3D registration accuracy in proportion to the calibration offset magnitude was observed. An offset of 125 mm translated to a 2D-DC loss of $2-8\%$ for 2D-3D alignment, and $4-10\%$ for 2D-3D reconstruction. This result was consistent with the pattern observed for DRR fidelity in Chapter 4. Offsets which resulted in smaller S2O distances resulted in a greater loss of registration accuracy, which was also in accordance with the DRR fidelity results.

Interestingly, the I_0 calibration did not have a discernable impact on the level of either alignment or reconstruction accuracy. However, it is worth noting that the choice of likelihood function (L_2 error) may have influenced this result. Since the L_2 error only measures intensity differences, varying I_0 - which acts as a linear scaling of each DRR pixel intensity - would not change the monoticity of the distribution of likelihood values. In some cases it may have been possible for the amplified differences to better serve the parameter-search because the likelihood values would appear less "flat" (for more insight, see the discussion on monotocity presented in Chapter 2.4.4). This observation also highlights the interplay between choice of likelihood function and the step-size of the random-walk.

For the simulated targets, 3D-DCs were computed between the ground-truth volume (used to generate the ground-truth DRR) and the estimated volume. The aim of computing these values was to observe to what extent errors arising from the ambiguity of 2D projections are propagated into 3D. The results, however, were very noisy for both the S2O distance and the incident energy parameters. It was clear that despite reasonably accurate 2D registrations, their corresponding 3D registrations were significantly worse. This result highlighted the difficulty of the 2D-3D reconstruction problem, and provides a strong argument for incorporating additional X-ray views when they are available. Future investigations may benefit from increasing the number of samples comprising the Markov chain.

The same general pattern of decreasing 2D-3D reconstruction accuracy was evident when using real X-ray image targets. An offset of 125 mm translated to a 2D-DC loss of $3-14\%$ for 2D-3D alignment, and $2-7\%$ for 2D-3D reconstruction. Compared to simulated targets, real X-ray images are more difficult to optimise for, because of the presence of photon noise, and foreign objects such as the adhesive used to hold the femurs in place, or the surface upon which the femurs were placed. Moreover, innate discrepancies will always exist between a DRR of a perfectly aligned CT volume and the real-X-ray image due to the secondary imaging effects which were not accounted for by the renderer. For the case of deformable registration, the model is unable to account for all possible variations in bone appearance, causing additional discrepancies to exist for unseen targets. In addition, while real X-ray image targets were used, these were taken of dry-bone specimens (no soft tissue present). Moreover, since the safety of a living subject did not have to be considered, it was possible to image the dry-bone specimens with an energy level that would otherwise be considered excessive for living subjects. It should therefore be noted that the "real" X-ray image targets that were used do not fully represent the complexity of a real-world clinical scenario.

Two strengths stand out when considering the investigations that were performed. First, the DRR renderer that was developed in Chapter 3 made it possible to investigate important variables of the image formation process employed by X-ray machines. Second, the ability to using real X-ray images as ground-truth, has not previously been described in literature. While the renderer and the unique dataset enabled novel insights into the importance of machine-specific calibration, the investigations were limited by a number of factors. First, only two parameters were considered, namely d_{S2O} and I_0; other parameters such as d_{S2D} or the contrast enhancement parameters ("window" and level") are also of interest. The contrast-enhancement parameters in particular - which pertain primarily to intensity information - in particular are worthy of an investigation as it may provide insight into the minimal impact on registration accuracy that was recorded for I_0. Another limiting factor is that registration accuracy was only considered with respect to binary overlap (Dice coefficients) and not with

regard to the accuracy of the estimated intensity information. Finally, it should be noted that despite the insights made using the unique dataset of CT-X-ray image pairs, only a limited number of samples (N=5) were available. Small samples therefore had a disproportionate effect on group-level statistics, and robust conclusions were not possible. Nonetheless, as a first exploration into this research question, this dataset proved to be very valuable.

6.7 Conclusion

A set of experiments were performed to evaluate the role of machine-specific DRR calibration in the context of 2D-3D reconstruction, which comprises three components, an X-ray simulation method, a deformable model, and a parameter search strategy. Previously, Chapter 4 presented a set of experiments validating the X-ray simulation component (developed in Chapter 3), and evaluating machine-specific calibration for X-ray simulation itself.

The first experiment of this chapter describes experiments that were performed to validate the deformable femur model component developed in Chapter 5. The second experiment measured the reduction in computation time afforded through hardware-based acceleration of the deformable model. The third and fourth experiments utilised the parameter search strategy (also developed in Chapter 5) to evaluate machine-specific calibration in the context of 2D-3D reconstruction. While this chapter includes a discussion of its experimental results, Chapter 7 provides a more in-depth discussion of these results in the context of the field of orthopaedic 2D-3D reconstruction together with conclusions and insights gained from other chapters.

Chapter 7

Discussion

7.1 Introduction

In the preceding chapters, a number of novel contributions were made to the field of 2D-3D reconstruction. This chapter provides a discussion of the work that was done in relation to the state-of-the-art with regard to its strengths, weaknesses, and limitations. Persisting challenges and opportunities for future research are also highlighted. In the sections to follow, each component comprising the 2D-3D reconstruction algorithm is addressed separately, followed by concluding remarks. It should be noted that an in-depth discussion of the state-of-the-art is provided in Chapter 2.5 and is therefore not included here.

7.2 Digitally Reconstructed Radiograph renderer

While computationally expensive, the ability to generate simulated X-ray images with a DRR renderer significantly improves the accuracy of 2D-3D registration compared to feature-based approaches. Over time, a variety of DRR rendering methods have been proposed (see Chapter 2.4.2). However, almost all have focused on reducing computation time, and none have considered implicit X-ray machine-specific calibration settings which affect both the resulting geometry and intensity profile of the simulated X-ray image.

A standard volume rendering framework was adapted to the unique requirements of X-ray imaging. Such an approach allows machine-specific X-ray projection geometry to be accurately and efficiently modelled, and allows the renderer to be calibrated to settings readily available in the standard DICOM header of an X-ray image. Furthermore, a significant reduction in computation time was made possible through a hardware-accelerated parallelisation of the rendering algorithm. A computation time reduction factor of as much

as 14 was achieved, while factors of up to 25 are reported in literature [101]. However, a further reduction in computation time was achieved by extending the GPU implementation to the rest of the 2D-3D reconstruction algorithm (see Section 7.4.1).

Two machine-specific parameters were the focus of an experimental investigation: the distance from the X-ray source to the object to be imaged (source-to-object distance), and modelled energy intensity of the X-ray source (incident energy). A set of simulated X-ray images were rendered as each correctly calibrated parameter was offset by an increasing magnitude. The resulting rendering errors were then measured through comparison to a ground-truth X-ray image - with known calibration parameters - of an object of known geometry and density. A set of simulated X-ray images were rendered as each correctly calibrated parameter was offset by an incrementing magnitude. The resulting rendering errors were then measured through comparison to a ground-truth X-ray image - with known calibration parameters - of an object of known geometry and density. The results demonstrated a 2D Dice overlap error of 0.4% for every 1-cm offset of the S2O distance, and a pixel intensity error of 1% for every 0.1 point offset in the incident energy coefficient. While the results of the 2D-3D reconstruction experiments suggest that intensity-based discrepancies have lesser importance compared to geometric discrepancies, they will certainly have more of an impact in more noisy contexts such as when soft tissues and other bone structures are present. Furthermore, the renderer does not model secondary imaging effects, such as scatter radiation, veiling glare and beam-hardening, and this can result in pixel intensity errors. Experimental results demonstrated that the linear slope of density-dependent intensity values was over-estimated and it was concluded that beam-hardening is most likely the cause. By accounting for secondary imaging effects, as proposed by [147], such undesirable consequences can be avoided. Moreover, combining such a first-principles approach with physics-based deep learning methods [176] could alleviate the need for simulating a large number of X-ray images in order to ensure the parameter-space is sufficiently explored.

The iterative nature of DRR-based 2D-3D reconstruction necessitates that thousands of X-ray images be simulated to ensure the parameter-space is sufficiently explored. As a future research avenue, physics-based deep learning methods could help alleviate the need for simulating a large number of X-ray images with different parameter configurations [176]. Combining PBDL methods with a first-principles-based approach such as the one proposed by [147] is a worthwhile avenue of future research.

7.2.1 Advanced compositing functions for improved registration

An interesting avenue for future research is the potential benefit of using alternative compositing functions for the registration process. A rudimentary example is a simple threshold along

each ray to remove tissue noise, making the algorithm more robust. More advanced transfer functions - such as those that consider the distribution of intensity values along each ray (histogram-based statistics) - hold much promise. For example, preliminary investigations show that material classes such as cortical bone, cancellous bone, and soft tissue can potentially be distinguished using skewness (the third standardised moment of the distribution). If different types of bone regions can thus be segmented "on the fly", contours residing in different regions may be registered in tandem - making the parameter search strategy more robust. Furthermore, the DRR render could, in principle, leverage information about the location of these regions as prior information for inferring neighbouring pixel values.

7.2.2 Choice of machine-specific parameters

In the presented work, only two machine-specific parameters were the focus of experimental investigation. One of these two parameters, the S2O distance, was automatically read from the standard DICOM header of the X-ray images. However, the renderer, in principle, can be made to adapt readily to any other relevant parameters provided in the header.

An X-ray phantom was used to manually calibrate the incident energy parameter. This was accomplished by minimising the difference in pixel intensity of pixel neighbourhoods - on both the DRR and imaged phantom - corresponding to known-density regions in the phantom volume. However, such a calibration will only be valid for target X-ray images obtained using the exact same X-ray machine settings for peak kilo-voltage, milli-Ampere-second, and exposure time, since these all have an effect on the magnitude of incident energy. Future research might use the Bayesian inferencing strategy described in Chapter 5 to search for an optimal incident energy value in parallel to the existing parameters. Furthermore, the incident energy value, or at least its upper and/or lower bound could - in principle - be derived from the aforementioned X-ray machine settings as they are provided in the standard DICOM header. Once determined, such information can be incorporated as prior knowledge in the Bayesian inferencing strategy - either by determining whether a random proposal is accepted or rejected, or by informing the proposal function used to draw candidate values. The fact that these parameters are considered independently allows their respective evaluations to be performed in parallel, which scales more readily, especially if hardware-based parallelisation is employed.

The DICOM header of an X-ray image also include two parameters concerning contrast enhancement post-processing of the X-ray image namely "window-centre" and "window-width" (see Chapter 3.4.3). A similar mis-calibration-based investigation for these parameters will provide additional insights into their importance in the context of 2D-3D reconstruction. For example, the parameter search strategy may be made more robust by computing image

similarity as an aggregation of likelihood terms where each term corresponds to the same target X-ray view, but having been post-processed using different contrast enhancement settings.

7.3 Deformable models

Volumetric modelling of CT information has previously been accomplished using tetrahedral-based statistical shape and intensity models (SSIMs). While SSIMs are robust to artefacts and noise, they require a large amount of training data in order to obtain a model capable of expressing all possible target appearances. Gaussian process morphable models provide a number of advantages over this conventional approach, but have thus far only been used for surface-based models. The foremost advantage of GPMMs is their ability to intuitively and analytically incorporate expert domain knowledge using covariance functions. Thus, sufficiently generalised models can be constructed - even when only a few training examples are available. Alternatively, researchers have explored the use of deep-learning-based data augmentation techniques to artificially increase the quantity of available training data [177, 116].

The femur bone was modelled using a novel application of GPMMs to CT voxel-volumes. First, the proposed method for constructing a voxel-based GPMM was validated using instances generated using an analytically defined "lollipop" model. A GPMM with an analytically defined covariance matrix was used as a preliminary model with which to establish correspondence between the reference volume and the rest of volumes comprising the training set. The analytical GPMM was subsequently augmented with the statistical insights provided by said training set. The proposed modeling approach resulted in similar principal components as the original model. The femur GPMM was trained in a similar fashion - limiting the examples comprising the training set to a binary intensity distribution. In doing so, the 3D-3D registration step used to establish correspondence among the training examples was made much simpler, by avoiding the complex local minima present in a full-intensity CT. However, when sampling from the model, each sample deformation field is not applied to the binary reference volume, but rather to the original CT volume from which the reference was derived - resulting in an intensity distribution more representative of actual CT. As a consequence, the resulting deformable model has reduced flexibility. The specificity and generality of the model was validated using conventional mesh-based metrics adapted to the voxel-based setting.

As future work, it would be worth extending the GPMM to explicitly model intensity information with an additional covariance matrix. In principal, an independent intensity

model component will enable more accurate reconstructions. While some authors have studied the independent or combined approaches in the context of 2D-3D bone reconstruction [136], further study of the relationship between the shape and intensity subspaces would be a valuable contribution to the field.

While the voxel-based modelling approach avoids segmentation errors, and holds promise for multi-object modelling, the size of the CT volume - and thus the number of objects which can be accommodated - remains limited due to the current memory capacity of commercial graphics hardware. However, it is assumed that this issue will become less of a factor in the future, as the technology continues to be developed.

Finally, while PCA-based modelling approaches have been largely successful, it is worth noting that they assume anatomical structures can be modelled with sufficient accuracy as linear combinations of deformations. Some authors have highlighted the limitations of this assumption (such as a susceptibility to outliers) and proposed non-linear modelling approaches [96, 97, 167, 4]. Furthermore, the modeling of one type of bone structure does not readily transfer to others as accurate GPMM-based modelling relies heavily on the provided anatomy-specific training examples.

7.4 Parameter search strategy

When reconstructing volumetric bone structures from a 2D X-ray image, linear and non-linear methods have been proposed for determining the set of patient-specific parameters (see Chapter 2.4.4). Gradient-based strategies are often trapped in local optima and a pose initialisation step is therefore necessary to ensure a more robust parameter search. Pose initialisation is typically done separately, and often requires manual input in the form of landmark annotations.

The Bayesian MCMC-based approach proposed by [138] was adapted to the context of 2D-3D reconstruction. In doing so, manual pose initialisation was avoided - since the algorithm does not rely on gradients and this mitigates susceptibility to local optima.

Given enough time, the entire parameter space will be searched and so the global optima is guaranteed to be found eventually - which is not the case for gradient-based approaches. The level of accuracy thus scales with the available computational resources. However, compared to other approaches, the chosen approach has a much longer computation time, which is impractical for clinical use.

For the experimental investigation, the influence of the same two machine-specific parameters on 2D-3D reconstruction accuracy were investigated, namely the S2O distance and the incident energy of the modelled X-rays. First, having generated a set of random DRRs,

each was used as a simulated registration target while offsetting the correctly calibrated parameters in the same fashion as for the first experiment (see Section 7.2). A 2D Dice overlap error of at least 0.3% was estimated for every 1 cm with which the S2O distance was offset. The experiment was then repeated while using real X-ray images of dry-bone femurs as targets, which demonstrated a 2D Dice overlap error of at least 0.4% for every 1 cm with which the S2O distance was offset. The incident energy parameter had no discernible effect on the resulting registration accuracy for either the simulated or real-world targets. As a whole, the experimental results demonstrate that machine-specific calibration has a noticeable impact on the accuracy of simulated X-ray images and, in the case of S2O distance, also on the accuracy of 2D-3D reconstructions.

In this work, the parameter search strategy was used as a means to compare relative registration accuracy among a set of different DRR renderer calibrations. The registration accuracy corresponding to each calibration setting was considered with respect to a single solution. However, an advantage of a Bayesian approach to the parameter search - which was not utilised in this work - is its ability to instead provide a distribution of solutions. Thus, a more nuanced understanding of the effect that machine-specific mis-calibration has on parameter search robustness is possible in future work.

The L_2 error was used exclusively as the likelihood function of parameter search strategy. However, the chosen approach allows for multiple image-based - as well as feature-based - similarity measures to be combined together, thus providing a means to incorporate different sources of information and also to better navigate the parameter space.

Researchers in the field of 2D-3D reconstruction have become increasingly focused on reconstruction from a single X-ray image in order to reduce the dose of ionising radiation a patient would receive [182]. The results of the 2D-3D reconstruction experiments performed in this work highlight the difficulty of imposing this constraint; the ambiguity of solutions resulting from the process of 2D projection is compounded. For example, the first principal component of the femur model, which varies in the bone-length direction, is very difficult to distinguish from the pose parameter affecting pitch. Here, a second X-ray view is very valuable because it allows one to decouple the effects of these two parameters. The focus of research should be informed by the cost-benefit trade-off of increased radiation dosage versus increased accuracy, which in turn depends on the application for which 2D-3D reconstruction is employed. Another point worth noting is that low-dose X-ray machines, such as the one provided by EOS, acquires two X-ray images by default.

The scope of the 2D-3D reconstruction experiments that were performed were limited to relative registration accuracy among sets of different DRR renderer calibrations. It is

therefore not possible to comment on the performance of the chosen approach in relation to existing approaches, but would be a valuable line of future investigations.

7.4.1 Parallelisation of entire DRR-based 2D-3D reconstruction algorithm

A significant reduction in computation time was made possible through a hardware-accelerated parallelisation of the model-sampling and likelihood estimation step. A hundred-fold reduction in computation time was achieved, compared to a CPU-only implementation. When compared to the computation time results pertaining solely to the rendering step, the process of sampling from the model lowered the average FPS by at least 8% - even when the model comprising a single principal component was used. Such an approach is limited predominantly by GPU memory capacity, though it is assumed that this limitation will be alleviated as the technology continues to develop.

Previously, [47], parallelised the model-sampling method for tetrahedral meshes. However, they did not report on the reduction in computation when compared to CPU. Their work is the only example in literature of extending the parallelisation of the X-ray simulation method to the rest of the 2D-3D reconstruction algorithm.

In this work, almost the entire 2D-3D reconstruction algorithm was implemented on the GPU. In principal, the parts of the Bayesian inferencing framework which are still executed on the CPU could in also be migrated to the GPU, leading to additional performance gains. In pursuit of further improving computation time, future research should focus on porting the parameter search algorithm onto the GPU. By effectively executing the entire parameter search on the GPU, memory exchanges between the CPU and GPU during this process - which are a major contributor to computation time - would be eliminated.

7.5 Conclusion

The field of 2D-3D reconstruction has made it possible for 3D bone structures, conventionally derived from 3D modalities such as CT, to be derived from 2D modalities such as X-ray imaging. It has thus provided a truly valuable clinical contribution as interventions such as implant design and postoperative evaluation are made more accessible, less expensive and, in some cases, reducing the dose of ionising radiation to the patient.

In this work, a novel 2D-3D reconstruction algorithm was developed, comprising a volume rendering approach to X-ray simulation, a voxel-based deformable femur model, and a Bayesian inferencing framework - all accelerated using a graphics processing unit. The

experimental results support the hypothesis that machine-specific calibration - a commonly overlooked consideration - has a noticeable impact not only on the accuracy of simulated X-ray images, but also on the accuracy of 2D-3D reconstructions. Furthermore, an extensive literature review was presented, as well as an in-depth commentary on the work done and suggestions for future research. In summary, the work presented in this thesis provides valuable insight into previously overlooked facets of orthopaedic 2D-3D reconstruction, and suggests that machine-specific calibration should be considered carefully when performing 2D-3D reconstruction.

Bibliography

[1] Agricola, R., Reijman, M., Bierma-Zeinstra, S., Verhaar, J., Weinans, H., and Waarsing, J. (2013). Total hip replacement but not clinical osteoarthritis can be predicted by the shape of the hip: a prospective cohort study. *Osteoarthritis and Cartilage*, 21(4):559–564.

[2] Ahmad, O., Ramamurthi, K., Wilson, K. E., Engelke, K., Prince, R. L., and Taylor, R. H. (2010). Volumetric DXA (VXA): a new method to extract 3D information from multiple in vivo DXA images. *Journal of Bone and Mineral Research*, 25(12):2744–2751.

[3] Akenine-Moller, T., Haines, E., and Hoffman, N. (2019). *Real-time rendering*. AK Peters/CRC Press.

[4] Armstrong, J. R., Campbell, J. Q., and Petrella, A. J. (2021). A comparison of cartesian-only vs. cartesian-spherical hybrid coordinates for statistical shape modeling in the lumbar spine. *Computer Methods and Programs in Biomedicine*, 204:106056.

[5] Arvo, J. and Kirk, D. (1990). Particle transport and image synthesis. In *Proceedings of the 17th annual Conference on Computer Graphics and Interactive Techniques*, pages 63–66.

[6] Aubert, B., Vergari, C., Ilharreborde, B., Courvoisier, A., and Skalli, W. (2016). 3D reconstruction of rib cage geometry from biplanar radiographs using a statistical parametric model approach. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 4(5):281–295.

[7] Aubin, C. E., Descrimes, J. L., Dansereau, J., Skalli, W., Lavaste, F., and Labelle, H. (1995). Geometrical modeling of the spine and the thorax for the biomechanical analysis of scoliotic deformities using the finite element method. *Ann. Chir.*, 49(8):749–761.

[8] Bai, C., Shao, L., Da Silva, A. J., and Zhao, Z. (2003). A generalized model for the conversion from CT numbers to linear attenuation coefficients. *IEEE Transactions on Nuclear Science*, 50(5):1510–1515.

[9] Baka, N., de Bruijne, M., van Walsum, T., Kaptein, B., Giphart, J., Schaap, M., Niessen, W. J., and Lelieveldt, B. P. (2012). Statistical shape model–based femur kinematics from biplane fluoroscopy. *IEEE Trans Med Imag*, 31(8):1573–1583.

[10] Baka, N., Kaptein, B. L., de Bruijne, M., van Walsum, T., Giphart, J. E., Niessen, W. J., and Lelieveldt, B. P. F. (2011). 2D–3D shape reconstruction of the distal femur from stereo X–ray imaging using statistical shape models. *Med. Image Anal.*, 15(6):840–850.

[11] Ballard, D. H., Hayhoe, M. M., Pook, P. K., and Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4):723–742.

[12] Benameur, S., Mignotte, M., Parent, S., Labelle, H., Skalli, W., and de Guise, J. (2003). 3D/2D registration and segmentation of scoliotic vertebrae using statistical models. *Comput. Med. Imaging Graph.*, 27(5):321–337.

[13] Bernard, F., Salamanca, L., Thunberg, J., Tack, A., Jentsch, D., Lamecker, H., Zachow, S., Hertel, F., Goncalves, J., and Gemmar, P. (2017). Shape-aware surface reconstruction from sparse 3D point-clouds. *Medical Image Analysis*, 38:77–89.

[14] Besl, P. J. and McKay, N. D. (1992). Method for registration of 3D shapes. In *Robotics–DL tentative*, pages 586–606. International Society for Optics and Photonics.

[15] Bhunre, P. K., Leow, W. K., and Howe, T. S. (2007). Recovery of 3D pose of bones in single 2D X-ray images. In *Applications of Computer Vision, 2007. WACV'07. IEEE Workshop on*, pages 48–48. IEEE.

[16] Birkfellner, W., Seemann, R., Figl, M., Hummel, J., Ede, C., Homolka, P., Yang, X., Niederer, P., and Bergmann, H. (2005). Wobbled splatting – a fast perspective volume rendering method for simulation of X–ray images from CT. *Physics in medicine and biology*, 50(9):N73–N84.

[17] Birkfellner, W., Wirth, J., Burgstaller, W., Baumann, B., Staedele, H., Hammer, B., Gellrich, N. C., Jacob, A. L., Regazzoni, P., and Messmer, P. (2003). A faster method for 3d/2d medical image registration—a simulation study. *Physics in Medicine & Biology*, 48(16):2665.

[18] Boisvert, J., Cheriet, F., Pennec, X., Labelle, H., and Ayache, N. (2008). Articulated spine models for 3D reconstruction from partial radiographic data. *IEEE Trans. Biomed. Eng.*, 55(11):2565–2574.

[19] Bonaretti, S., Seiler, C., Boichon, C., Reyes, M., and Büchler, P. (2014). Image–based vs. mesh–based statistical appearance models of the human femur: implications for finite element simulations. *Medical Engineering & Physics*, 36(12):1626–1635.

[20] Boussaid, H., Kadoury, S., Kokkinos, I., Lazennec, J. Y., Zheng, G., and Paragios, N. (2011). 3D model–based reconstruction of the proximal femur from low–dose biplanar X–ray images. In *The 22nd British Machine Vision Conference–BMVC 2011*, pages 1–10.

[21] Bresenham, J. E. (1965). Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30.

[22] Bryan, R., Mohan, P. S., Hopkins, A., Galloway, F., Taylor, M., and Nair, P. B. (2010). Statistical modelling of the whole human femur incorporating geometric and material properties. *Medical Engineering & Physics*, 32(1):57–65.

[23] Caponetti, L. and Fanelli, A. M. (1990). 3D bone reconstruction from two X–ray views. In *Proceedings of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.

[24] Card, M. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.

[25] Card, S. K., Moran, T. P., and Newell, A. (2018). *The psychology of human-computer interaction*. CRC Press.

[26] Castro-Mateos, I., Pozo, J. M., Pereañez, M., Lekadir, K., Lazary, A., and Frangi, A. F. (2015). Statistical interspace models (sims): application to robust 3d spine segmentation. *IEEE Transactions on Medical Imaging*, 34(8):1663–1675.

[27] Cerveri, P., Sacco, C., Olgiati, G., Manzotti, A., and Baroni, G. (2017). 2D/3D reconstruction of the distal femur using statistical shape models addressing personalized surgical instruments in knee arthroplasty: A feasibility analysis. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 13(4).

[28] Chaibi, Y., Cresson, T., Aubert, B., Hausselle, J., Neyret, P., Hauger, O., de Guise, J. A., and Skalli, W. (2012). Fast 3D reconstruction of the lower limb using a parametric model and statistical inferences and clinical measurements calculation from biplanar X–rays. *Comput. Methods Biomech. Biomed. Engin.*, 15(5):457–466.

[29] Chen, C. and Zheng, G. (2014). Fully automatic segmentation of AP pelvis X–rays via random forest regression with efficient feature selection and hierarchical sparse shape composition. *Computer Vision and Image Understanding*, 126:1–10.

[30] Chen, F., Liu, J., Zhao, Z., Zhu, M., and Liao, H. (2017). Three-dimensional feature-enhanced network for automatic femur segmentation. *IEEE Journal of Biomedical and Health Informatics*, 23(1):243–252.

[31] Chen, X., Graham, J., Hutchinson, C., and Muir, L. (2013). Automatic generation of statistical pose and shape models for articulated joints. *IEEE Transactions on Medical Imaging*, 33(2):372–383.

[32] Chintalapani, G., Ellingsen, L. M., Sadowsky, O., Prince, J. L., and Taylor, R. H. (2007). Statistical atlases of bone anatomy: construction, iterative improvement and validation. *Med. Image Comput. Comput. Assist. Interv.*, 10(Pt 1):499–506.

[33] Chintalapani, G., Sadowsky, O., Ellingsen, L. M., Prince, J. L., and Taylor, R. H. (2009). Integrating statistical models of bone density into shape based 2D–3D registration framework. In *MICCAI 2009 Workshop: Probablistic Models for Medical Image Analysis*, pages 151–161.

[34] Chu, C., Takao, M., Ogawa, T., Yokota, F., Sato, Y., and Zheng, G. (2016). Statistical shape modeling of compound musculoskeletal structures around the thigh region. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 885–888. IEEE.

[35] Chui, H., Rambo, J., Duncan, J., Schultz, R., and Rangarajan, A. (1999). Registration of cortical anatomical structures via robust 3D point matching. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 168–181. Springer.

[36] Cook, R. L. (1986). Stochastic sampling in computer graphics. *ACM Transactions on Graphics (TOG)*, 5(1):51–72.

[37] Cook, R. L., Porter, T., and Carpenter, L. (1984). Distributed ray tracing. In *Proceedings of the 11th annual Conference on Computer Graphics and Interactive Techniques*, pages 137–145.

[38] Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):681–685.

[39] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape Models–Their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59.

[40] Cresson, T., Branchaud, D., Chav, R., Godbout, B., and de Guise, J. A. (2010). 3D shape reconstruction of bone from two X–ray images using 2D/3D non–rigid registration based on moving least–squares deformation. In *Medical Imaging 2010: Image Processing*.

[41] Dey, J. and Napel, S. (2006). Targeted 2D/3D registration using ray normalization and a hybrid optimizer. *Medical Physics*, 33(12):4730–4738.

[42] Dong, S., Kettenbach, J., Hinterleitner, I., Bergmann, H., and Birkfellner, W. (2008). The Zernike expansion–an example of a merit function for 2D/3D registration based on orthogonal functions. In *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pages 964–971. Springer.

[43] Dorgham, O. M., Laycock, S. D., and Fisher, M. H. (2012). GPU accelerated generation of digitally reconstructed radiographs for 2D/3D image registration. *IEEE Trans. Biomed. Eng.*, 59(9):2594–2603.

[44] Duchon, C. E. (1979). Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology and Climatology*, 18(8):1016–1022.

[45] Dumas, R., Blanchard, B., Carlier, R., de Loubresse, C. G., Le Huec, J. C., Marty, C., Moinard, M., and Vital, J. M. (2008). A semi–automated method using interpolation and optimisation for the 3D reconstruction of the spine from bi–planar radiography: a precision and accuracy study. *Med. Biol. Eng. Comput.*, 46(1):85–92.

[46] Ehlke, M., Frenzel, T., Ramm, H., Shandiz, M. A., Anglin, C., and Zachow, S. (2015). Towards robust measurement of pelvic parameters from AP radiographs using articulated 3D models. In *Computer Assisted Radiology and Surgery (CARS)*.

[47] Ehlke, M., Ramm, H., Lamecker, H., Hege, H. C., and Zachow, S. (2013). Fast generation of virtual X–ray images for reconstruction of 3D anatomy. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2673–2682.

[48] EOS Imaging (2022). EOS Imaging: A Leader in Low-dose 2D and 3D Medical Imaging.

[49] Fleute, M., Lavallée, S., and Julliard, R. (1999). Incorporating a statistically based shape model into a system for computer–assisted anterior cruciate ligament surgery. *Med. Image Anal.*, 3(3):209–222.

[50] Fouefack, J.-R., Borotikar, B., Douglas, T. S., Burdin, V., and Mutsvangwa, T. E. M. (2020). Dynamic multi-object gaussian process models. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 755–764, Cham. Springer International Publishing.

[51] Fritscher, K. D., Agnes, G., and Rainer, S. (2007). 3D image segmentation using combined shape–intensity prior models. *Int. J. Comput. Assist. Radiol. Surg.*, 1(6):341–350.

[52] Galibarov, P., Prendergast, P., and Lennon, A. (2010). A method to reconstruct patient–specific proximal femur surface models from planar pre–operative radiographs. *Medical Engineering & Physics*, 32(10):1180–1188.

[53] Gee, A. H. and Treece, G. M. (2014). Systematic misregistration and the statistical analysis of surface data. *Medical Image Analysis*, 18(2):385–393.

[54] Gheno, R., Nectoux, E., Herbaux, B., Baldisserotto, M., Glock, L., Cotten, A., and Boutry, N. (2012). Three–dimensional measurements of the lower extremity in children and adolescents using a low–dose biplanar X–ray device. *European Radiology*, 22(4):765–771.

[55] Glover, G. and Pelc, N. (1980). Nonlinear partial volume artifacts in X-ray computed tomography. *Medical Physics*, 7(3):238–248.

[56] Grassi, L., Hraiech, N., Schileo, E., Ansaloni, M., Rochette, M., and Viceconti, M. (2011). Evaluation of the generality and accuracy of a new mesh morphing procedure for the human femur. *Medical Engineering & Physics*, 33(1):112–120.

[57] Gregory, J. S. and Aspden, R. M. (2008). Femoral geometry as a risk factor for osteoporotic hip fracture in men and women. *Medical Engineering & Physics*, 30(10):1275–1286.

[58] Gunay, M., Shim, M. B., and Shimada, K. (2007). Cost– and time–effective three–dimensional bone–shape reconstruction from X–ray images. *Int. J. Med. Robot.*, 3(4):323–335.

[59] Hanaoka, S., Fritscher, K., Welk, M., Nemoto, M., Masutani, Y., Hayashi, N., Ohtomo, K., and Schubert, R. (2011). 3-D graph cut segmentation with Riemannian metrics to avoid the shrinking problem. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 554–561. Springer.

[60] Haq, R., Cates, J., Besachio, D. A., Borgie, R. C., and Audette, M. A. (2015). Statistical shape model construction of lumbar vertebrae and intervertebral discs in segmentation for discectomy surgery simulation. In *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pages 85–96. Springer.

[61] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.

[62] Haverkamp, D. J., Schiphof, D., Bierma-Zeinstra, S. M., Weinans, H., and Waarsing, J. H. (2011). Variation in joint shape of osteoarthritic knees. *Arthritis & Rheumatology*, 63(11):3401–3407.

[63] Heimann, T. and Meinzer, H. P. (2009). Statistical shape models for 3D medical image segmentation: a review. *Medical Image Analysis*, 13(4):543–563.

[64] Herrera, I., Buchart, C., Aguinaga, I., and Borro, D. (2014). Study of a ray casting technique for the visualization of deformable volumes. *IEEE Trans Vis Comput Graphics*, 20(11):1555–1565.

[65] Humbert, L., De Guise, J. A., Aubert, B., Godbout, B., and Skalli, W. (2009). 3D reconstruction of the spine from biplanar X–rays using parametric models based on transversal and longitudinal inferences. *Med. Eng. Phys.*, 31(6):681–687.

[66] Humbert, L., Whitmarsh, T., De Craene, M., del Río Barquero, L. M., Fritscher, K., Schubert, R., Eckstein, F., Link, T., and Frangi, A. F. (2010). 3D reconstruction of both shape and bone mineral density distribution of the femur from DXA images. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 456–459.

[67] Hurvitz, A. and Joskowicz, L. (2008). Registration of a CT–like atlas to fluoroscopic X–ray images using intensity correspondences. *International Journal of Computer Assisted Radiology and Surgery*, 3(6):493–504.

[68] Jacobs, F., Sundermann, E., De Sutter, B., Christiaens, M., and Lemahieu, I. (1998). A fast algorithm to calculate the exact radiological path through a pixel or voxel space. *CIT. Journal of Computing and Information Technology*, 6(1):89–94.

[69] Jerbi, T., Burdin, V., Leboucher, J., Stindel, E., and Roux, C. (2013). 2D–3D frequency registration using a low–dose radiographic system for knee motion estimation. *IEEE Trans. Biomed. Eng.*, 60(3):813–820.

[70] Jia, X. and Jiang, S. B. (2015). *Graphics Processing Unit–Based High Performance Computing in Radiation Therapy*. CRC Press.

[71] Jia, X., Wei, W., and Jia, K. (2012). A GPU–based DRR generation method using cubic window. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH–MSP), 2012 Eighth International Conference on*, pages 403–406. IEEE.

[72] Kadoury, S., Cheriet, F., and Labelle, H. (2009). Personalized X–ray 3D reconstruction of the scoliotic spine from hybrid statistical and image–based models. *IEEE Trans. Med. Imaging*, 28(9):1422–1435.

[73] Kang, X., Yau, W.-P., and Taylor, R. H. (2016). Simultaneous pose estimation and patient-specific model reconstruction from single image using maximum penalized likelihood estimation (MPLE). *Pattern Recognition*, 57:61–69.

[74] Karade, V., Vikas, K., and Bhallamudi, R. (2014). 3D femur model reconstruction from biplane X–ray images: a novel method based on Laplacian surface deformation. *Int. J. Comput. Assist. Radiol. Surg.*, 10(4):473–485.

[75] Kay, T. L. and Kajiya, J. T. (1986). Ray tracing complex scenes. *ACM SIGGRAPH computer graphics*, 20(4):269–278.

[76] Khamene, A., Bloch, P., Wein, W., Svatos, M., and Sauer, F. (2006). Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy. *Medical Image Analysis*, 10(1):96–112.

[77] Khronos, G. (2012). The openCL specification, version 1.2.

[78] Klima, O., Kleparnik, P., Spanel, M., and Zemcik, P. (2016). Intensity–based femoral atlas 2D/3D registration using Levenberg–Marquardt optimisation. In *Proc. of SPIE Medical Imaging*, volume 9788F, pages 1–12.

[79] Knaan, D. and Joskowicz, L. (2003). Effective intensity-based 2D/3D rigid registration between fluoroscopic X-ray and CT. In *Lecture Notes in Computer Science*, pages 351–358. Springer Berlin Heidelberg.

[80] Lacroute, P. and Levoy, M. (1994). Fast volume rendering using a shear–warp factorization of the viewing transformation. In *Proceedings of the 21st annual Conference on Computer Graphics and Interactive Techniques*, pages 451–458. ACM.

[81] Lam, K. C., Gu, X., and Lui, L. M. (2015). Landmark constrained genus–one surface Teichmüller map applied to surface registration in medical imaging. *Medical Image Analysis*, 25(1):45–55.

[82] Lamecker, H., Wenckebach, T. H., and Hege, H. C. (2006). Atlas-based 3D–Shape reconstruction from X–Ray images. In *18th International Conference on Pattern Recognition (ICPR'06)*, pages 371–374.

[83] Langton, C., Pisharody, S., and Keyak, J. (2009). Generation of a 3D proximal femur shape from a single projection 2D radiographic image. *Osteoporosis International*, 20(3):455–461.

[84] Laporte, S., Skalli, W., De Guise, J. A., Lavaste, F., and Mitton, D. (2003). A biplanar reconstruction method based on 2D and 3D contours: Application to the distal femur. *Comput. Methods Biomech. Biomed. Engin.*, 6(1):1–6.

[85] LaRose, D. A. (2001). *Iterative X–ray/CT registration using accelerated volume rendering*. PhD thesis, Citeseer.

[86] Le Bras, A., Laporte, S., Bousson, V., Mitton, D., De Guise, J. A., Laredo, J. D., and Skalli, W. (2004). 3D reconstruction of the proximal femur with low–dose digital stereoradiography. *Comput. Aided Surg.*, 9(3):51–57.

[87] Lee, M. K., Lee, S. H., Kim, A., Youn, I., Lee, T. S., Hur, N., and Choi, K. (2008). The study of femoral 3D reconstruction process based on anatomical parameters using a numerical method. *JBSE*, 3(3):443–451.

[88] Lester, H. and Arridge, S. R. (1999). A survey of hierarchical non–linear medical image registration. *Pattern recognition*, 32(1):129–149.

[89] Li, X., Yang, J., and Zhu, Y. (2006). Digitally reconstructed radiograph generation by an adaptive Monte Carlo method. *Physics in Medicine and Biology*, 51(11):2745.

[90] Lightweight Java Gaming Library (2015). Version 2.9.

[91] Livyatan, H., Yaniv, Z., and Joskowicz, L. (2003). Gradient–based 2D/3D rigid registration of fluoroscopic X–ray to CT. *IEEE Trans. Med. Imag.*, 22(11):1395–1406.

[92] Ljung, P., Krüger, J., Groller, E., Hadwiger, M., Hansen, C. D., and Ynnerman, A. (2016). State of the art in transfer functions for direct volume rendering. In *Computer Graphics Forum*, volume 35, pages 669–691. Wiley Online Library.

[93] Lodox (Pty) Ltd. (2022). Lodox Systems: Lodox Critical Imaging Technology.

[94] Lorusso, A., Eggert, D. W., and Fisher, R. B. (1995). *A comparison of four algorithms for estimating 3D rigid transformations*. University of Edinburgh, Department of Artificial Intelligence.

[95] Lüthi, M., Gerig, T., Jud, C., and Vetter, T. (2017). Gaussian process morphable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1860–1873.

[96] Ma, J., Wang, A., Lin, F., Wesarg, S., and Erdt, M. (2017). Nonlinear statistical shape modeling for ankle bone segmentation using a novel kernelized robust PCA. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 136–143. Springer.

[97] Ma, J., Wang, A., Lin, F., Wesarg, S., and Erdt, M. (2019). A novel robust kernel principal component analysis for nonlinear statistical shape modeling from erroneous data. *Computerized Medical Imaging and Graphics*, 77:101638.

[98] Macovski, A. (1983). *Medical imaging systems*. Prentice Hall.

[99] Mahfouz, M. R., Hoff, W. A., Komistek, R. D., and Dennis, D. A. (2003). A robust method for registration of three–dimensional knee implant models to two–dimensional fluoroscopy images. *IEEE Trans Med Imag*, 22(12):1561–1574.

[100] Malzbender, T. (1993). Fourier volume rendering. *ACM Transactions on Graphics (TOG)*, 12(3):233–250.

[101] Markelj, P., Tomaževič, D., Likar, B., and Pernuš, F. (2012). A review of 3D/2D registration methods for image–guided interventions. *Med. Image Anal.*, 16(3):642–661.

[102] Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). PET–CT image registration in the chest using free–form deformations. *IEEE Trans Med Imag*, 22(1):120–128.

[103] Matthews, I. and Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164.

[104] Max, N. (1995). Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108.

[105] McCollough, C. H., Primak, A. N., Braun, N., Kofler, J., Yu, L., and Christner, J. (2009). Strategies for reducing radiation dose in CT. *Radiologic Clinics of North America*, 47(1):27–40.

[106] McLaughlin, R. A., John, H., Hawkes, D. J., J.Alison, N., Byrne, J. V., and Tim, C. (2002). A comparison of 2D–3D intensity–based registration and feature–based registration for neurointerventions. In *Lecture Notes in Computer Science*, pages 517–524.

[107] Miao, S., Wang, Z. J., and Liao, R. (2016). A CNN regression approach for real–time 2D/3D registration. *IEEE Trans Med Imag*, 35(5):1352–1363.

[108] Mishra, A., Mondal, P., and Banerjee, S. (2014). 2D/3D non–rigid image registration by an efficient demons approach. In *2014 IEEE 27th International Symposium on Computer–Based Medical Systems*.

[109] Mitton, D., Deschênes, S., Laporte, S., Godbout, B., Bertrand, S., de Guise, J. A., and Skalli, W. (2006). 3D reconstruction of the pelvis from bi–planar radiography. *Comput. Methods Biomech. Biomed. Engin.*, 9(1):1–5.

[110] Mu, Z. (2016). A fast DRR generation scheme for 3D–2D image registration based on the block projection method. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 169–177.

[111] Munbodh, R., Chen, Z., Jaffray, D. A., Moseley, D. J., Knisely, J. P., and Duncan, J. S. (2008). Automated 2D–3D registration of portal images and CT data using line–segment enhancement. *Medical Physics*, 35(10):4352–4361.

[112] Mutsvangwa, T., Burdin, V., Schwartz, C., and Roux, C. (2015). An automated statistical shape model developmental pipeline: application to the human scapula and humerus. *IEEE Trans. Biomed. Eng.*, 62(4):1098–1107.

[113] Mutsvangwa, T., Wasswa, W., Burdin, V., Borotikar, B., and Douglas, T. (2017). Interactive patient–specific 3D approximation of scapula bone shape from 2D X–ray images using landmark–constrained statistical shape model fitting. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pages 1816–1819.

[114] Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275.

[115] Neogi, T., Bowes, M. A., Niu, J., Souza, K. M., Vincent, G. R., Goggins, J., Zhang, Y., and Felson, D. T. (2013). Magnetic resonance imaging–based three–dimensional bone shape of the knee predicts onset of knee osteoarthritis: data from the osteoarthritis initiative. *Arthritis & Rheumatism*, 65(8):2048–2058.

[116] Niri, R., Gutierrez, E., Douzi, H., Lucas, Y., Treuillet, S., Castañeda, B., and Hernandez, I. (2021). Multi-view data augmentation to improve wound segmentation on 3d surface model by deep learning. *IEEE Access*, 9:157628–157638.

[117] Ntasis, E., Cai, W., Sakas, G., and Nikita, K. S. (1999). Real time digital reconstructed radiograph (DRR) rendering in frequency domain. In *BMES/EMBS Conference, 1999. Proceedings of the First Joint*, volume 2, page 1041. IEEE.

[118] Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644.

[119] Odersky, M. (2014). The Scala language specification version 2.11. In *Programming Methods Laboratory, EPFL*. Citeseer.

[120] Organization, W. H. (2003). *Prevention and management of osteoporosis: report of a WHO scientific group*. Number 921. World Health Organization.

[121] Penney, G. P., Weese, J., Little, J. A., Desmedt, P., Hill, D. L., and Hawkes, D. J. (1998). A comparison of similarity measures for use in 2D–3D medical image registration. *IEEE Trans Med Imag*, 17(4):586–595.

[122] Pizer, S. M., Fletcher, P. T., Joshi, S., Thall, A., Chen, J. Z., Fridman, Y., et al. (2003). Deformable M–reps for 3D medical image segmentation. *International Journal of Computer Vision*, 55(2):85–106.

[123] Pizer, S. M., Jung, S., Goswami, D., Vicory, J., Zhao, X., Chaudhuri, R., et al. (2013). Nested sphere statistics of skeletal models. In *Innovations for Shape Analysis*, pages 93–115.

[124] Pluim, J. P., Maintz, J. A., and Viergever, M. A. (2003). Mutual–information–based registration of medical images: a survey. *IEEE Trans Med Imag*, 22(8):986–1004.

[125] Pomero, V., Mitton, D., Laporte, S., de Guise, J. A., and Skalli, W. (2004). Fast accurate stereoradiographic 3D–reconstruction of the spine using a combined geometric and statistic model. *Clin. Biomech.*, 19(3):240–247.

[126] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1996). *Numerical recipes in C*, volume 2. Cambridge University Press.

[127] QRM GmbH (2022). Bone Density Calibration Phantoms - A PTW Company - QRM.

[128] Rechenberg, I. (1970). *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*.

[129] Reyneke, C., Thusini, X., Douglas, T., Vetter, T., and Mutsvangwa, T. (2018). Construction and validation of image-based statistical shape and intensity models of bone. In *2018 3rd Biennial South African Biomedical Engineering Conference (SAIBMEC)*, pages 1–4. IEEE.

[130] Reyneke, C. J. F., Lüthi, M., Burdin, V., Douglas, T. S., Vetter, T., and Mutsvangwa, T. E. M. (2019). Review of 2-D/3-D Reconstruction Using Statistical Shape and Intensity Models and X-Ray Image Synthesis: Toward a Unified Framework. *IEEE Reviews in Biomedical Engineering*, 12:269–286.

[131] Roberts, M., Pacheco, E., Mohankumar, R., Cootes, T., and Adams, J. (2010). Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semi–automatic segmentation. *Osteoporosis International*, 21(12):2037–2046.

[132] Rueckert, D., Frangi, A., and Schnabel, J. (2001). Automatic construction of 3D statistical deformation models using non–rigid registration. In *Medical Image Computing and Computer–Assisted Intervention–MICCAI 2001*, pages 77–84. Springer.

[133] Russakoff, D. B., Rohlfing, T., Rueckert, D., Shahidi, R., Kim, D., and Maurer, C. R. (2003). Fast calculation of digitally reconstructed radiographs using light fields. In *Proceedings of SPIE*, volume 5032, pages 684–695.

[134] Sabella, P. (1988). A rendering algorithm for visualizing 3d scalar fields. In *Proceedings of the 15th annual Conference on Computer Graphics and Interactive Techniques*, pages 51–58.

[135] Sadowsky, O., Chintalapani, G., and Taylor, R. H. (2007). Deformable 2D–3D registration of the pelvis with a limited field of view using shape statistics. In *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pages 519–526. Springer.

[136] Sarkalkan, N., Nazli, S., Harrie, W., and Zadpoor, A. A. (2014). Statistical shape and appearance models of bones. *Bone*, 60:129–140.

[137] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675.

[138] Schönborn, S., Egger, B., Morel-Forster, A., and Vetter, T. (2017). Markov chain monte carlo for automated face image analysis. *International Journal of Computer Vision*, 123(2):160–183.

[139] Schroeder, W., Martin, K. M., and Lorensen, W. E. (1998). *The Visualization Toolkit An Oject-Oriented Approach to 3D graphics*. Prentice-Hall, Inc.

[140] Segal, M. and Akeley, K. (1999). The OpenGL graphics system: A specification (version 1.1).

[141] Seoud, L., Cheriet, F., Labelle, H., and Dansereau, J. (2011). A novel method for the 3D reconstruction of scoliotic ribs from frontal and lateral radiographs. *IEEE Trans. Biomed. Eng.*, 58(5):1135–1146.

[142] Sharma, G., Saevarsson, S., Amiri, S., Montgomery, S., Ramm, H., Lichti, D., Lieck, R., Zachow, S., and Anglin, C. (2012). Radiological method for measuring patellofemoral tracking and tibiofemoral kinematics before and after total knee replacement. *Bone and Joint Research*, 1(10):263–271.

[143] Siddon, R. L. (1985). Fast calculation of the exact radiological path for a three–dimensional CT array. *Medical Physics*, 12(2):252–255.

[144] Škerl, D., Tomaževič, D., Likar, B., and Pernuš, F. (2006). Evaluation of similarity measures for reconstruction–based registration in image–guided radiotherapy and surgery. *International Journal of Radiation Oncology Biology Physics*, 65(3):943–953.

[145] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Trans Med Imag*, 32(7):1153–1190.

[146] Spoerk, J., Bergmann, H., Wanschitz, F., Dong, S., and Birkfellner, W. (2007). Fast DRR splat rendering using common consumer graphics hardware. *Medical Physics*, 34(11):4302–4308.

[147] Staub, D. and Murphy, M. J. (2013). A digitally reconstructed radiograph algorithm calculated from first principles. *Medical Physics*, 40(1).

[148] Steger, S., Kirschner, M., and Wesarg, S. (2012). Articulated atlas for segmentation of the skeleton from head & neck CT datasets. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1256–1259. IEEE.

[149] Stegmann, M. B. (2000). *Active appearance models*. PhD thesis, Technical University of Denmark, 2000. 125, 129.

[150] Steininger, P., Philipp, S., Fritscher, K. D., Gregor, K., Benedikt, S., Markus, H., Karsten, S., and Rainer, S. (2008). Comparison of different metrics for Appearance–Model–Based 2D/3D–registration with X–ray images. In *Informatik Aktuell*, pages 122–126.

[151] Steinke, F. and Schölkopf, B. (2008). Kernels, regularization and differential equations. *Pattern Recognition*, 41(11):3271–3286.

[152] Styner, M. A., Rajamani, K. T., Nolte, L. P., Zsemlye, G., Székely, G., Taylor, C. J., and Davies, R. H. (2003). Evaluation of 3D correspondence methods for model building. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 63–75. Springer.

[153] Sun, W., Niessen, W. J., van Stralen, M., and Klein, S. (2013). Simultaneous multiresolution strategies for nonrigid image registration. *IEEE Trans. Image Process.*, 22(12):4905–4917.

[154] Taghizadeh, E., Terrier, A., Becce, F., Farron, A., and Büchler, P. (2019). Automated CT bone segmentation using statistical shape modelling and local template matching. *Computer Methods in Biomechanics and Biomedical Engineering*, 22(16):1303–1310.

[155] Taha, A. A. and Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):29.

[156] Tang, T. S., MacIntyre, N., Gill, H., Fellows, R., Hill, N., Wilson, D., and Ellis, R. E. (2004). Hardware–assisted 2D/3D intensity–based registration for assessing patellar tracking. In *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pages 1095–1096. Springer.

[157] Tang, T. S. Y. and Ellis, R. E. (2005). 2D/3D deformable registration using a hybrid atlas. *Med. Image Comput. Comput. Assist. Interv.*, 8(Pt 2):223–230.

[158] Thermo Fisher (2022). Amira software - life science research. Version 5.4.3.

[159] UNIBAS-GRAVIS (2018). SCALISMO software - scalable image analysis and shape modelling. Version 0.17.

[160] Väänänen, S. P., Grassi, L., Flivik, G., Jurvelin, J. S., and Isaksson, H. (2015). Generation of 3D shape, density, cortical thickness and finite element mesh of proximal femur from a DXA image. *Medical Image Analysis*, 24(1):125–134.

[161] Väänänen, S. P., Isaksson, H., Julkunen, P., Sirola, J., Kröger, H., and Jurvelin, J. S. (2011). Assessment of the 3-D shape and mechanics of the proximal femur using a shape template and a bone mineral density image. *Biomechanics and Modeling in Mechanobiology*, 10(4):529–538.

[162] Väänänen, S. P., Jurvelin, J. S., and Isaksson, H. (2012). Estimation of 3D shape, internal density and mechanics of proximal femur by combining bone mineral density images with shape and density templates. *Biomechanics and Modeling in Mechanobiology*, 11(6):791–800.

[163] Valenti, M. (2016). *Statistical shape models based 2D/3D registration methods for knee orthopaedic surgery*. PhD thesis, Technical University of Denmark, 2000. 125, 129.

[164] Varnavas, A., Carrell, T., and Penney, G. (2015). Fully automated 2D–3D registration and verification. *Medical Image Analysis*, 26(1):108–119.

[165] Vicente, S., Kolmogorov, V., and Rother, C. (2008). Graph cut based image segmentation with connectivity priors. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

[166] Von Berg, J., Dworzak, J., Klinder, T., Manke, D., Kreth, A., Lamecker, H., Zachow, S., and Lorenz, C. (2011). Temporal subtraction of chest radiographs compensating pose differences. In *Medical Imaging 2011: Image Processing*, volume 7962, page 79620U. International Society for Optics and Photonics.

[167] Wang, D. and Tanaka, T. (2020). A robust method for kernel principal component analysis. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 294–297. IEEE.

[168] Wang, F., Davis, T., and Vemuri, B. (2002). Real–time DRR generation using cylindrical harmonics. *Medical Image Computing and Computer–Assisted Intervention (MICCAI)*, pages 671–678.

[169] Wang, L., Gao, X., and Fang, Q. (2013). A novel mutual information–based similarity measure for 2D/3D registration in image guided intervention. In *Orange Technologies (ICOT), 2013 International Conference on*, pages 135–138.

[170] Weiskopf, D. (2007). *GPU-based interactive visualization techniques*. Springer.

[171] Westermann, R. and Sevenich, B. (2001). Accelerated volume ray–casting using texture mapping. In *Proceedings of the Conference on Visualization'01*, pages 271–278. IEEE Computer Society.

[172] Westin, S. H., Arvo, J. R., and Torrance, K. E. (1992). Predicting reflectance functions from complex surfaces. *ACM SIGGRAPH Computer Graphics*, 26(2):255–264.

[173] Whitmarsh, T., Humbert, L., Barquero, L. M. D. R., Di Gregorio, S., and Frangi, A. F. (2013). 3D reconstruction of the lumbar vertebrae from anteroposterior and lateral dual–energy X–ray absorptiometry. *Medical Image Analysis*, 17(4):475–487.

[174] Whitmarsh, T., Humbert, L., De Craene, M., Barquero, L. M. D. R., and Frangi, A. F. (2011). Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual–energy X–ray absorptiometry. *IEEE Trans Med Imag*, 30(12):2101–2114.

[175] Whitmarsh, T., Humbert, L., De Craene, M., Luis, M., Fritscher, K., Schubert, R., Eckstein, F., Link, T., and Frangi, A. F. (2010). 3D bone mineral density distribution and shape reconstruction of the proximal femur from a single simulated dxa image: an in vitro study. In *Medical Imaging 2010: Image Processing*, volume 7623, page 76234U. International Society for Optics and Photonics.

[176] Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput. Surv.*, 55(4).

[177] Wu, S., Nakao, M., Tokuno, J., Chen-Yoshikawa, T., and Matsuda, T. (2019). Reconstructing 3d lung shape from a single 2d image during the deaeration deformation process using model-based data augmentation. In *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 1–4.

[178] Wybier, M. and Bossard, P. (2013). Musculoskeletal imaging in progress: the EOS imaging system. *Joint Bone Spine*, 80(3):238–243.

[179] Yao, J. and Taylor, R. H. (2001). Construction and simplification of bone density models. In *Medical Imaging 2001: Image Processing*.

[180] Yao, J. and Taylor, R. H. (2003). Assessing accuracy factors in deformable 2D/3D medical image registration using a statistical pelvis model. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1329–1334. IEEE.

[181] Yu, W., Chu, C., Tannast, M., and Zheng, G. (2016). Fully automatic reconstruction of personalized 3D volumes of the proximal femur from 2D X–ray images. *Int. J. Comput. Assist. Radiol. Surg.*, 11(9):1673–1685.

[182] Yu, W., Tannast, M., and Zheng, G. (2017). Non–rigid free–form 2D–3D registration using a B–spline–based statistical deformation model. *Pattern Recognition*, 63:689–699.

[183] Zhang, B., Sun, S., Sun, J., Chi, Z., and Xi, C. (2010). 3D reconstruction method from biplanar radiography using DLT algorithm: Application to the femur. In *2010 First International Conference on Pervasive Computing, Signal Processing and Applications*, pages 251–254.

[184] Zheng, G. (2006). Reconstruction of patient–specific 3D bone model from biplanar X–ray images and point distribution models. In *Image Processing, 2006 IEEE International Conference on*, pages 1197–1200. IEEE.

[185] Zheng, G. (2010a). Effective incorporating spatial information in a mutual information based 3D–2D registration of a CT volume to X–ray images. *Comput. Med. Imaging Graph.*, 34(7):553–562.

[186] Zheng, G. (2010b). Statistical shape model–based reconstruction of a scaled, patient-specific surface model of the pelvis from a single standard AP X–ray radiograph. *Med. Phys.*, 37(4):1424–1439.

[187] Zheng, G. (2011). Personalized X–ray reconstruction of the proximal femur via intensity-based non–rigid 2D–3D registration. In *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pages 598–606. Springer.

[188] Zheng, G. (2013). 3D volumetric intensity reconstruction from 2D X–ray images using partial least squares regression. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 1268–1271. IEEE.

[189] Zheng, G., Nolte, L., and Ferguson, S. J. (2011). Scaled, patient–specific 3D vertebral model reconstruction based on 2D lateral fluoroscopy. *Int. J. Comput. Assist. Radiol. Surg.*, 6(3):351–366.

[190] Zheng, G. and Schumann, S. (2009). 3D reconstruction of a patient–specific surface model of the proximal femur from calibrated X–ray radiographs: A validation study. *Medical Physics*, 36(4):1155–1166.

[191] Zheng, G. and Yu, W. (2017). Statistical shape and deformation models based 2D–3D reconstruction. In *Statistical Shape and Deformation Analysis*, pages 329–349. Elsevier.

[192] Zhu, Z. and Li, G. (2011). Construction of 3D human distal femoral surface models using a 3D statistical deformable model. *J. Biomech.*, 44(13):2362–2368.

[193] Zollei, L., Grimson, E., Norbash, A., and Wells, W. (2001). 2D–3D rigid registration of X–ray fluoroscopy and CT images using mutual information and sparsely sampled histogram estimators. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages 696–703. IEEE.

Appendix A

Target pose and model parameters

Table A.1 Randomly generated sets of pose parameters used as targets for the rigid registration experiment.

Index	Rotation			Translation	
	roll	pitch	yaw	trans-X	trans-Y
1	5.742	14.598	24.179	2.085	2.275
2	-13.945	4.164	30.385	-1.346	-2.951
3	17.419	12.719	12.318	-1.274	-0.825
4	36.316	-7.717	-9.947	-1.176	-3.476
5	10.932	13.662	-9.209	2.596	-0.059
6	-13.606	4.255	-15.853	1.268	0.387
7	11.257	-31.764	5.132	-2.684	0.801
8	0.49	-2.664	-12.509	-0.119	-0.765
9	15.659	1.379	21.68	2.4	2.646
10	-11.079	-18.177	3.703	0.367	-2.034
11	8.952	10.806	24.08	-2.014	-1.175
12	-2.998	27.398	6.446	0.336	-0.761
13	-36.055	5.877	31.607	-1.679	-1.921
14	-6.118	-3.994	-17.596	-0.115	-1.394
15	-7.176	13.455	-5.383	1.681	0.776
16	-20.085	24.827	-20.357	-3.01	1.881
17	-1.401	-3.952	-10.117	-2.296	-0.761
18	-18.619	14.449	12.257	-0.351	-1.186
19	-7.443	21.171	-2.842	1.186	-0.442
20	11.292	-3.964	11.388	0.629	-1.101
21	-22.359	-17.469	33.22	0.536	1.226
22	5.972	-12.211	14.746	-2.001	-1.399
23	-8.337	5.052	-1.61	1.731	1.256
24	-6.796	0.382	-16.519	-1.556	-0.793
25	24.337	5.873	-21.045	0.973	1.402
26	36.949	-18.524	-4.934	0.572	-0.912
27	-15.747	-13.697	4.01	-0.221	-2.535
28	-16.342	2.727	2.239	-2.796	-0.31
29	-25.232	33.386	10.783	1.191	1.322
30	-10.131	-15.849	7.65	0.45	-0.449