Calibration and optimization of a biplane fluoroscopy system for quantifying foot and ankle biomechanics

Eric Thorhauer

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington
2020

Reading Committee:

William R. Ledoux
Ashley Emery
Duane Storti

Program Authorized to Offer Degree:
Mechanical Engineering

University of Washington

Abstract

Calibration and optimization of a biplane fluoroscopy system for quantifying foot and ankle biomechanics

Eric Thorhauer

Chair of the Supervisory Committee:
William Ledoux
Mechanical Engineering and Orthopaedics & Sports Medicine

Joint pain or loss of function manifests in subtle changes in kinematics and the biomechanical behavior of the articulating surfaces, the ligaments, and tendons acting as soft tissue constraints. Understanding the relationships between anatomic morphology and healthy joint function, and how conservative and surgical treatments affect symptomatic behavior in vivo, will inform clinicians' practice. Clinically motivated research studies benefit greatly from the increased spatial and temporal resolution afforded by biplanar fluoroscopic systems, but the kinematic data generated from these studies come at high computational cost and radiation exposure to test subjects and operators. Additionally, there is no off-the-shelf software or hardware solution available currently for foot and ankle gait biplanar fluoroscopic systems. The imaging chains and software processing pipelines are modified and repurposed versions of stock fluoroscopic equipment. By altering this hardware chain, image quality and automated processing, and thus the accuracy and throughput of the system, are diminished. Therefore, the goal of this thesis is to dissect the custom hardware and software pipelines in our laboratory in order to understand, characterize, and attempt to mitigate sources of error while balancing kinematic accuracy, subject safety, and processing throughput.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. INTRODUCTION

1.1 WHY STUDY FOOT BIOMECHANICS?

Mobility and ambulation are fundamental aspects of the human experience that are directly correlated with our quality of life and self-fulfillment. Ignoring modern sedentary tendencies, the feet have historically been our primary interfaces to Earth: allowing us to walk, run, and jump through our constantly evolving reality. To meet the myriad kinetic demands of these tasks, the human foot and ankle anatomy has evolved as a complex structure of 30 bones in a variety of shapes and sizes, linked by over 20 articulating, highly congruent, and narrow joints responsible for transmitting forces through the body. Forces from daily activities vary in frequency, magnitude, mechanism, and effect. Biomechanics applies multidisciplinary engineering principles to studying the body's structure and function in the context of these loading scenarios, while organized research has led to advances in occupational safety, the treatment and rehabilitation of trauma or disease, and the design of prosthetic limbs, joint replacements, and assistive robotics; all to preserve, restore, or improve human function.

Among the principal goals of the Department of Veterans Affairs Puget Sound Healthcare System (VAPSHCS) Center for Limb Loss and MoBility (CLiMB) is the study of human foot and ankle kinematics as it relates to the prevention and treatment of various disorders affecting veteran servicemembers. Multifaceted orthopedic issues like flatfoot disorders, ankle instability, osteoarthritis, joint replacement, and diabetic complications are a small sample of the numerous issues that afflict veterans' health [1-5]. CLiMB combines the expertise of biomechanical engineers from various subdisciplines with clinicians, like orthopedic surgeons, prosthetists, and physical therapists, to pioneer treatment options using state-of-the-art technology. Accurate bio-kinematic data such as joint range of motion or laxity during dynamic tasks of daily living are essential for quantitatively assessing the issues that affect patients and informing best practices for clinicians. Healthcare decisions and practices informed by more reliable biomechanical studies could accelerate data-driven medicine and increase the overall quality of life. Since the foot plays such an essential role in connecting us to our world, we should strive to understand its

form and function as best as possible to gain a fuller appreciation of our evolution and human experience.

1.2 METHODS FOR ASSESSING JOINT KINEMATICS

Since Eadweard Muybridge's pioneering motion studies of human walking using film photography [6], biomechanics researchers have targeted virtually all regions of the body using imaging technology to quantify joint motion. Numerous tools for measuring human kinematics have been developed over the last century and improvements to their resolving power, both in time and space, have allowed researchers to ask more interesting and in-depth biomechanical questions (Table 1.1). Each tool has inherent strengths and weaknesses that prevent its universal application to every research study. Special considerations to the anatomy and research question of interest are required, as these decisions must balance accuracy, risks to test subjects and administrators, and the expenses associated with data collection and processing. Poor data, whether arising from temporal and spatial sampling limitations, noise corruption, or lack of biofidelity or realism of the model or task performed, limit our ability to distill decisive and meaningful clinical conclusions. Risks to test subjects and research staff include injury, exposure to chemicals, radiation, or other environmental factors that elicit ethical and safety considerations. Lastly, complex multidimensional biomechanical analyses usually require expensive equipment, teams of technical experts, willing research participants, and vast sums of time, money, and effort to generate an ostensibly small nugget of data. Given these daunting start-up costs, executing studies of in vivo bone motion is difficult and reliable, high-resolution datasets describing most tasks are rare. For any research study, a practical balance is sought between the capabilities and operational burdens of the data acquisition equipment and methods and the resources available to answer the question of interest.

Table 1.1: Common modalities used for assessing joint biomechanics.

Modality	Application	Advantage	Disadvantage	Spatial Resolution	Temporal Resolution
Motion capture with optical markers [7, 8]	Gait analysis, motion segment kinematics	Established, trusted standard for motion analysis	Bones move relative to overlying skin and soft tissue causing errors	~1mm	100-1000Hz
Video analysis [9]	Sport competitions, outdoor	Low-cost	Limited to simple body models	>1mm	<100Hz
Inertial measurement units [10, 11]	Sport competitions, outdoor	Low-cost	Sensor drift, indirect inferring bone motion	>1mm	~100Hz
MRI [12-14]	Soft tissue deformation, bone tracking, tissue properties	Excellent soft-tissue contrast	Smaller capture volumes limit experimental tasks	<1mm	<10Hz
Computed tomography [15]	Bone kinematics, deformable tissue shape modeling (e.g. heart)	Direct assessment of bone motion, volumetric data	High radiation, limited temporal resolution, small capture volume	<1mm	~1Hz
Standard fluoroscopy [16, 17]	Bone, radiopaque marker kinematics during slow or quasi-static tasks	Direct assessment of bone motion	Medium temporal resolution, reduced X-ray source power lending to poor image quality	<1mm	7.5-30Hz pulsed mode
High-speed X-ray [18-20]	Bone, radiopaque marker tracking during functional, dynamic tasks	Direct assessment of bone motion	Ionizing radiation, large footprint, expense and complexity.	<1mm	>100Hz

Acquisition of accurate sub-millimeter and sub-degree in vivo bone kinematic measurements is challenging but essential to understanding joint function. While a variety of sensors and methodologies have been combined into innovative techniques, many of these modalities are suboptimal with specific consideration to foot and ankle bone motion quantification (Table 1.1). Consider optical motion capture technology which uses retro-reflective markers attached to the skin above bony landmarks observed with an array of cameras. It is the established, widespread clinical standard for assessing human motion [7, 8]. These systems are relatively easy to use, present a very low risk to test subjects and operators, and have firmly established data acquisition and processing pipelines. However, skin-mounted markers are free to move relative to the underlying bones of interest, and at least three markers are required for each bone or motion segment. While many groups have investigated solutions to skin-motion artifact error and others have developed complex, multi-segment foot models [21-23], these technical issues continue to inhibit optical motion capture from being used to investigate complex, multi-bone structures like the foot. Additionally, the expected magnitude of changes or differences in joint kinematics for the orthopedic questions of interest to CLiMB is usually smaller than what may be currently and reliably resolved with optical motion capture. Even with large sample sizes of test subjects, it

may be difficult to detect a clinically meaningful difference in treatments or outcomes with the wrong instrument for quantifying joint motion; one that introduces more measurement variance. Rather than relying on skin-based markers to infer kinematics, the most accurate way to observe the bony motion of a complex structure, like a foot, is to peer inside the body and directly observe the bones with medical imaging. Magnetic resonance imaging (MRI) has been used to quantify bone kinematics [12-14], but is limited to minimally-loaded, spatially, and temporally constrained studies due to the sampling restrictions and geometric confines of the magnet bore and receiver coils. Computed tomography (CT), a three-dimensional (3D) X-ray-based scanning method, allows accurate volumetric imaging of bones [15] but suffers from limited temporal resolutions and the same capture volume limitations as MRI. By eliminating a spatial dimension, two-dimensional imaging methods like fluoroscopy [16, 17] and dedicated high-speed X-ray systems [18-20] based on radiographic projections provide an excellent basis for localizing bones inside the body because they possess favorable spatial and temporal resolution characteristics and larger capture volumes compared to MRI and CT (Table 1.1).

1.3 THE USES OF X-RAY-BASED BIOMECHANICAL ANALYSES

Immediately after their discovery by Wilhelm Roentgen 125 years ago, X-rays were used medically for imaging bones. By exploiting the relative differences in X-ray attenuation between tissues, bones can be directly visualized and localized in a non-contacting manner. Kinematic assessment via direct observation offers a substantial improvement over inferring bone positions from skin-mounted surface markers. Unfortunately, by moving up the electromagnetic spectrum from infra-red light (motion capture) to using X-rays, subjects and operators are exposed to ionizing radiation and special care is required in study design and execution. For biplane systems, the number of X-ray sources, and hence the radiation exposure, is effectively doubled. However, the introduction of a second imaging plane (hence, biplanar imaging) allows for accurate triangulation of objects from the stereo image pairs by reducing the out-of-plane tracking errors associated with single-plane systems. Both single-plane (one source and detector) and biplane X-ray systems evolved as interoperative tools for guiding surgeons' tools during complex procedures. Biplane systems have also been used in Roentgen stereophotogrammetric analysis (RSA) [24-27], which was originally developed in the 1970s by Selvik and colleagues

using plain film X-rays. RSA tracks the motion of radiopaque markers to infer migrations of joint replacement components and has been applied to hips and knees [28-32]. Improvements to computer and imaging technology have modernized RSA marker tracking into a totally digital process with high levels of precision [33]. In addition to surgical and clinical applications, biplanar imaging systems have been utilized to quantify in vivo and in vitro joint kinematics in animals [34-38] and humans. X-ray based kinematic assessments of human joint motion have included the shoulder, wrist, cervical spine, jaw [38], lumbar spine, hip, knee, ankle, and foot [16, 17, 39, 40]. In addition to resolving six-degree of freedom bone kinematics, biplane systems have also been used in conjunction with other imaging modalities, like MRI, to quantify cartilage contact, ligament elongation, soft tissue deformation, suture motion, and ligament graft motion [20, 41-45]. In summary, high-speed biplanar X-ray systems are powerful tools for biomechanists looking to assess in vivo functional joint behavior during realistic, functional tasks. These kinds of data offer unprecedented insight into the composition and function of the human body and provide invaluable feedback that improves the understanding of and planning of clinicians. The challenges associated with the development and operation of biplane systems, however, pale in comparison to these potential discoveries.

Chapter 2. BIPLANAR FLUOROSCOPY AT CLIMB

This chapter provides an overview of the basics of acquiring and processing fluoroscopic images for foot and ankle bone tracking during gait. It includes an overview of the image formation process, hardware and bone tracking software descriptions, the challenges specific to imaging foot and ankle anatomy, and some of the issues with CLiMB's current biplane system implementation that largely motivated this thesis work.

2.1 X-RAY IMAGING BASICS

X-rays' interactions with matter are a function of beam energy and material properties like atomic density. X-rays interacting with matter will either be 1) transmitted (while partially attenuated), 2) absorbed (completely attenuated), or 3) cause a subatomic scatter interaction. X-rays that are too weak to pass through are absorbed by the tissue as radiation. Those that are transmitted through the tissue and reach the detector have an attenuated intensity as a result of passing through the medium. This relative attenuation behavior can be approximated using Beer-Lambert's Law and forms the basis for X-ray image formation:

$$\ln\left(\frac{I}{I_o}\right) = -\mu\, x \tag{2.1}$$

This equation describes the X-ray intensity I passing through a given material in terms of incident intensity I_o, thickness of material x, and X-ray attenuation coefficient μ. The attenuation coefficient (μ) is a function of the X-ray potential energy (source voltage) and the atomic density of the material. Denser tissues, like bone, have higher attenuation constants than tissues like fat, muscle, or skin at equivalent X-ray voltage potentials. Higher-energy X-rays more easily traverse a given material with less attenuation and have a greater propensity for scattering. X-rays that are scattered may reach the image detector on an altered trajectory and contribute to image noise. By landing in an errant location on the detector, the attenuation signal (image) of the X-ray through the subject is corrupted. The simple model of attenuation in Equation 2.1 does not account for this scattering behavior, the true polychromatic nature of real X-ray sources, or the beam-hardening phenomenon that causes preferential attenuation of the weaker energies in these spectra. For the sake of mathematical simplicity, this thesis will accept the deficiencies of this simple attenuation model.

During walking the human foot moves rapidly, and motion blur degrades the accuracy of kinematic tracking. Although fluoroscopy systems allow for pulsed X-ray imaging, which greatly reduces the radiation dose, the acquisition rates are limited to 30 Hz for standard "C-arm" fluoroscopy units, and up to 60 Hz for custom systems. A sampling rate of at least 100 Hz and a camera exposure time no more than 1 millisecond has been suggested for the best images with minimal blur during gait [39]. To achieve these rapid sequences of short X-ray pulses, cine-angiography X-ray sources have been employed [46-48]. The short (~1 millisecond) X-ray pulses from these sources are synchronized to the framerate of the high-speed cameras. These systems produce images with the least amount of motion blur and X-ray scatter; however, they can cost an order of magnitude more than C-arm fluoroscopy units and have a large footprint. Smaller C-arm units have compact X-ray sources, designed for use in operating room situations where framerates faster than the 30 Hz refresh rates of closed-circuit television are not needed. These fluoroscopes are commonly capable of pulsed imaging as well, although at a much lower frame rate of 7.5 or 15 frames per second during surgical applications. This necessitates the use of *continuous* fluoroscopy mode when studying dynamic activities, whereby the source constantly emits X-rays while the high-speed camera records images at an independent frame rate. This frame rate needs to be fast enough to freeze the dynamic motion of the subject and slow enough to use an appropriate exposure time and produce a reasonable volume of image data for each trial.

X-ray source tubes are controlled with two primary parameters: source potential energy (peak kilovolts, kV_p) and current (milliamperes, mA). The potential of the beam governs the contrast between soft and hard tissues in the images. X-rays that are too weak fail to pass through bones onto the detector and are absorbed by soft tissues as radiation, while X-rays with too high of a potential may saturate the detector and are more prone to scatter interactions that contribute to image noise. Increased beam current (mA) equates to more X-rays emitted from the source and incident on the subject and detector. This increases radiation exposure, but higher photon fluence also means more signal to be detected and decreased image noise. Responsible use of X-ray equipment will balance these radiographic parameters to yield the best data (image) at the lowest risk (radiation dose) for subjects and operators.

X-ray detectors collect, amplify, and convert the attenuated X-ray signal into visible light for visualization. These detectors include image intensifier technology that has been used for decades, as well as newer flat-panel X-ray detectors. Flat-panel detectors are extremely sensitive at detecting X-rays, so a lower dose is required to make an exposure. However, current technology limits the temporal sampling rate of flat-panel detectors to around 60 Hz, and this makes them suboptimal for imaging high-speed activities like gait without substantial motion blur. As a result, biomechanical analyses of functional tasks like jumping, walking, and running require a different approach that pairs a high-speed video camera to the older image intensifier technology. Image intensifiers have a curved input phosphor screen that receives X-rays which are converted into electrons, that are then accelerated towards the output phosphor screen of the evacuated image intensifier body. At the output phosphor, the electrons are converted back into visual light (wavelength, $\lambda = 550$ nanometers) that is emitted out of the back of the unit. This visible light is then focused onto a camera sensor for digital recording via a series of lenses. Image intensifiers and the chain of lenses introduce significant spatial distortions and intensity variations that must be addressed before image data can be used in a quantitative manner (Chapter 3). The response time for phosphor-based image intensifier screens is on the order of 1 millisecond [49]. Paired with a high-speed camera capable of acquiring high-resolution grayscale images at up to 1000 frames per second, these custom image intensifier chains are capable of freezing dynamic bone motion during fast maneuvers.

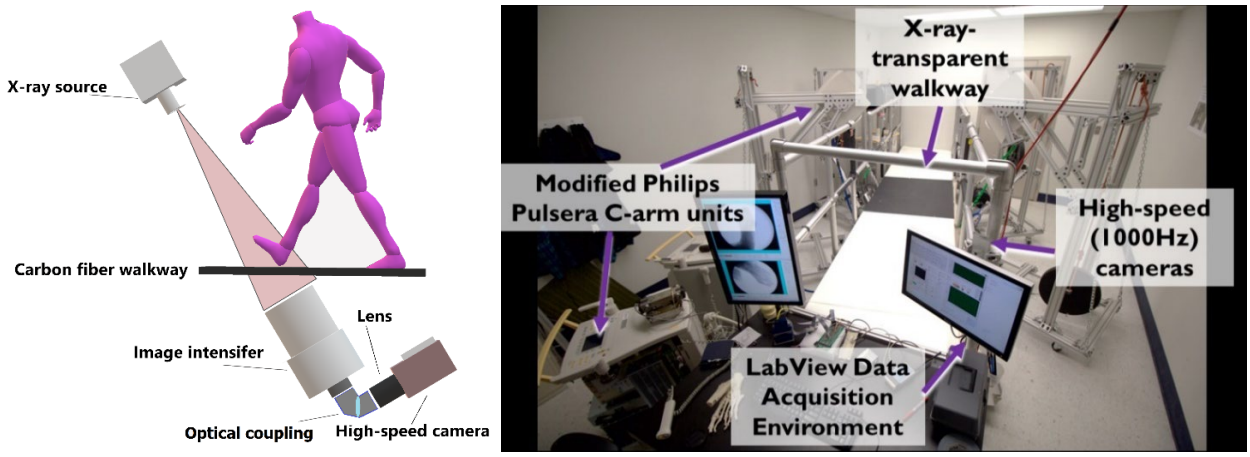

Figure 2.1. Biplane system hardware overview (left) and photo of the CLiMB Biplane Fluoroscopy lab (right). Modified C-arm fluoroscopy units are aimed through a carbon floorplate that allows for direct imaging of the feet during gait.

2.2 X-RAY BONE TRACKING OVERVIEW

Fully descriptive bone kinematics cannot be derived directly from the raw X-ray images, even after distortions and nonlinearities have been accounted for. Rather, computer vision photogrammetric methods are used to virtually model the biplane camera system in silico. Object positions and orientations are estimated by triangulation in the synchronized stereo X-ray image sequences. Accurately determining kinematic poses of objects requires 1) a calibrated mathematical model of the relevant aspects of the X-ray imaging chain, 2) methods for recovering the object pose, and 3) a software pipeline for efficiently processing this high volume of data.

To be sufficiently accurate, the mathematical model of the imaging chain requires knowledge of the camera parameters that project an object from 3D world space to the 2D image space (Chapter 5) and should account for the optical distortions of the camera lenses and X-ray image intensifier (Chapter 3) for each arm of the biplanar system. These model parameters are derived in a calibration procedure (Chapter 5) by localizing radiographic marker control points (Chapter 4) using a custom calibration object and procedure (Chapter 5) that includes a non-linear optimization routine that minimizes geometric error. After building this calibrated stereo camera

model for the biplane imaging system, the distortion-corrected images can be used to estimate the six-degree-of-freedom kinematics of objects in the laboratory coordinate system via stereo triangulation.

There are two main methods for recovering object poses from stereo X-ray images: marker-based and model-based. The marker-based method requires rigidly embedding at least three radiopaque markers to the object of interest. The locations of these markers with respect to the object's coordinate system are obtained using a coordinate measurement machine or computed tomography (CT) scan [39, 50]. Markers are localized on the calibrated stereo fluoroscopic images and triangulated into the three-dimensional world space (laboratory coordinate system). The transformation between the markers in the reference coordinate system (bone or anatomical) and the biplane (world or laboratory) coordinate system gives the pose of the object. This method was pioneered by Selvik, et al. for quantifying joint kinematics and implant migration in static X-ray images [24-26, 28, 29, 32], and has been the gold standard for cadaver-based validation studies quantifying the performance of biplanar systems used in dynamic biomechanical and kinematic analyses [19, 39]. Localization accuracy of these markers on noisy fluoroscopic images directly affects the accuracy of the final estimated kinematics and the validity of this "gold standard". Surgically embedded markers have also been used for in vivo biplane studies of knee bone kinematics [51]. The minimum of three markers per bone make it practically infeasible in the foot and ankle (3 beads x 30 bones = 90 total markers), in addition to the invasiveness it presents to test subjects.

For studies with real, living subjects, the suitable alternative to marker-based tracking is a model-based tracking approach that amounts to a 3D-2D image registration problem. Three-dimensional models of the anatomy of interest (bones) are segmented from subject-specific volumetric CT scans (or taken from a statistical shape database) and digitally reconstructed radiographs (DRR) are created in the biplane software. DRRs are virtual X-ray images made by projecting the volumetric CT bone data onto the fluoroscopic image planes using the calibrated camera model and a simple model of X-ray attenuation physics. The six-degree-of-freedom poses of bones in the virtual software environment are iteratively adjusted in an optimization loop until the DRR projections optimally match the pair of acquired fluoroscope images. Optimal matching is

determined using metrics that calculate the similarity between the DRR projection and the real fluoroscope images. Chapter 6 describes and investigates the performance of various image similarity metrics commonly used for 3D-2D, DRR-based fluoroscope registration. Building a cost function for the optimizer that is robust to image noise, false minima in global 3D registration, and adaptable to different anatomy of interest is important for a smooth optimizer response that quickly and accurately yield the biomechanical variable of interest. This process of 1) setting the 3D bone pose in lab space, 2) generating a DRR at that position, and 3) comparing it to the underlying fluoroscope image is performed for each bone of interest and repeated for every image frame to recover 3D joint kinematics. The optimization of a single bone in a single frame requires the generation of hundreds of DRR images as the bone pose is iteratively micro-adjusted, and a single biplane fluoroscope image sequence can easily contain between 100 and 1000 frames of data that require tracking. As a result of this high computational volume, the process of generating DRR images is commonly performed using a parallel computing architecture like an array of dedicated computational computers connected via message passing interface (MPI) or high-end graphics processing units (GPU) using the Compute Unified Device Architecture (CUDA). The ray-casting approach of forming simulated X-rays is a perfect example of the kind of "embarrassingly parallel" interpolation/sampling problem graphics processing units excel at [52].

Biplane X-ray studies generate an immense amount of data at a high computational expense. Efforts are made, wherever possible in this work, to choose or design algorithms that automate or parallelize data processing streams. Additionally, to reduce the volume of data storage required for large, high-resolution medical imaging files, the post-processing pipeline will avoid writing temporary files for each intermediate processing step with as many being handled on-the-fly by the GPU processor as possible.

2.3 CLIMB'S BIPLANE SYSTEM

The biplanar fluoroscopy system at the VA Puget Sound Healthcare System is based on the design pioneered by the Steadman Phillipon Research group [53]. A pair of Philips BV Pulsera "C-arm" fluoroscopes were disarticulated to allow custom configuration of the X-ray source and detectors within the laboratory space. These C-arms have a 1440 Watt rhenium-tungsten X-ray

source, cesium-iodide X-ray image intensifiers with imaging diameter modes of 31-, 23-, and 17-centimeter and ~2 millisecond response times. The system is positioned about a radiotransparent walkway that permits imaging through the transverse plane of the foot at an oblique angle during gait. This view is unique to the CLiMB biplane system and provides the best compromise to visualizing each of the foot bones over the gait cycle. The system is controlled by a custom LabView control system [54] that synchronizes fluoroscope firing, camera triggering from optical beams, and ground reaction impulse data collection. High-speed cameras (Vision Research Phantom v5.2) are optically coupled to the output phosphor of the X-ray image intensifiers. These cameras can acquire 896 x 1152 pixel, 12-bit grayscale images at 1000 Hz with exposure times as short as 1997 microseconds at full resolution.

Bone kinematics are determined via model-based tracking CT-based DRRs in custom software called Digital Radiograph Registration Accelerated by CUDA Operations (DRRACO) [55]. DRRACO is an evolving, custom software solution to the bone tracking problem that runs on GPU hardware for efficient DRR generation. Prior to usage in DRRACO, fluoroscopic images are processed for spatial and illumination distortions as described by Iaquinto [18, 19]. DRRs are generated using a ray-casting algorithm that samples voxel intensities of CT scans. Each voxel intensity value is proportional to the average X-ray attenuation of its finite sub-volume during CT scanning, expressed in “Hounsfield units” (HU). A cost function based on image similarity drives the optimal registration of the 3D CT model’s DRR projections to the pair of fluoroscope images. The similarity is expressed as a scalar distance between each pair of DRR and fluoroscope images, derived from the normalized cross-correlation (NCC) of the intensities and gradients.

2.4 CURRENT CHALLENGES

Dynamic biplane X-ray imaging of human joints has many technical challenges that have prevented its widespread use. Getting the high-quality X-ray images of all the bones of interest, over the entire activity cycle, with ethical radiation levels is challenging for any anatomy. In addition to these general difficulties, there are specific considerations to imaging dynamic foot and ankle motion, which are described below. Furthermore, both hardware and software

limitations currently prevent the CLiMB biplane system from attaining its full potential as a research tool.

2.4.1 *Foot and ankle imaging challenges*

The anatomy of the foot and ankle (Figure 2.2) presents a few challenges in acquiring useful images for each bone of the foot during gait in both camera views simultaneously. The hindfoot is the most radiodense portion of the foot where a large portion of incident X-rays are attenuated or scattered. The soft tissues of the plantarflexor muscles and heel pad, and complex, highly congruent articulations like the ankle and subtalar joints, lead to images that are high in X-ray mottle and low in contrast. Midfoot bones like the cuneiforms lack distinct osseous landmarks that aid in accurate bone pose determination. The midfoot bones are relatively small on the images, resulting in fewer pixels to use for the 3D-2D registration task. Additionally, joint spacing in this region is typically sub-millimeter. Dynamic calculations of joint spacing as bone-to-bone distances, which are important for studying arthritis progression and surgical corrections [20, 45, 56], are more sensitive to the compound errors from tracking two or more bones. In the forefoot, the metatarsals and phalanges are longer bones with rounded ends that make resolving long-axis rotation less accurate. The soft tissues surrounding the forefoot bones are noticeably thinner, leading to overexposure of the X-ray image if details in the hindfoot are required in the same frame.

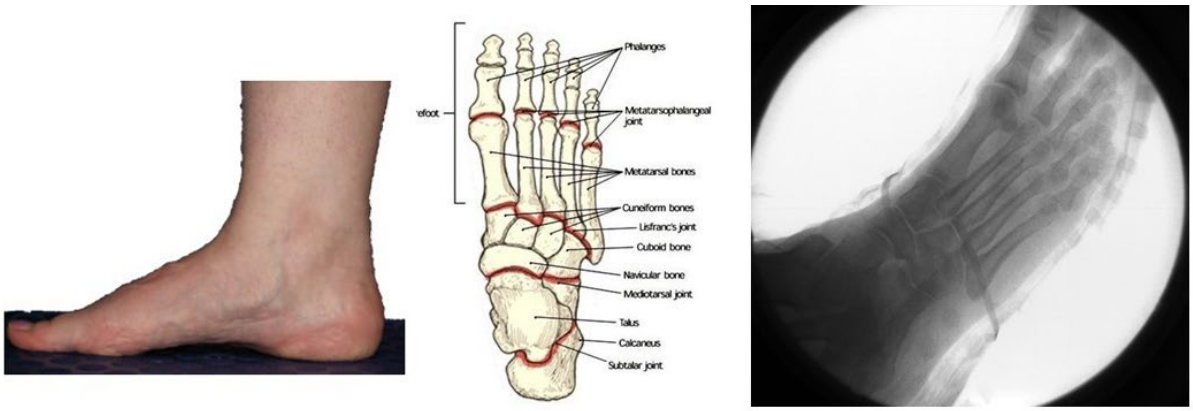

Figure 2.2: The human foot and ankle (left) complex consists of 30 tightly organized bones of varying shapes and sizes (center) that may be viewed under fluoroscopy (right). The fluoroscopy system projects all of the bones' features and any other materials like soft tissues or footwear onto the image plane. The great challenge comes in optimizing the X-ray camera parameters and viewpoints to maximize bone feature content in the image for kinematic tracking.

2.4.2 *Hardware limitations*

Modified, disarticulated C-arm fluoroscopy units are much cheaper than dedicated high-speed X-ray imaging equipment but suffer hardware limitations that limit their effectiveness. Firstly, the X-ray generator of the BV Pulsera is limited to a maximum of 1440 Watts and about 20 mA of tube current. The X-ray source parameters of electrical current (mA) and voltage potential (kV) are not independently adjustable. Rather, they are coupled in the Philips control software as a series of predefined curves that limit imaging flexibility but prevent damage to the X-ray sources. The standard source-to-detector distance of the BV Pulsera is 1 meter. In order to accommodate human test subjects traversing the biplane system walkway, the disarticulated source-to-detector distance is typically between 1.3 and 1.5 meters. Since X-ray intensity diminishes with distance following the inverse-square law, the beam intensity is about half of the nominal amount ($1.41^{-2} = 0.500$). This further reduces image quality and brightness. The lower signal intensity requires: 1) increasing the X-ray source power (which is usually operating at its maximum), 2) increasing camera exposure time (introducing motion blur), 3) increasing camera gain (introducing noise), and 4) opening the lens aperture to allow more light onto the sensor

(introducing a defocusing blur). Lastly, by replacing the standard charge-coupled device (CCD) camera equipment with high-speed camera chains, we can no longer access the advanced image processing functions Philips includes in the software loop. All the exposure parameters, brightness corrections, contrast enhancements, and denoising must be determined and applied manually to achieve image quality remotely close to the stock, intact fluoroscopy units. The main advantage we have in a research setting as opposed to an operating room setting is that these images are not required in real-time. Therefore, we can acquire our images at the best compromise of X-ray source and camera settings and apply these corrections in a series of post-acquisition processing steps that are outlined in Chapter 3. A primary aim of this thesis is to reverse-engineer the pertinent components of the default image signal processing chain and implement them in software optimized for CLiMB's unique combination of hardware and research aims.

2.4.3 *Software limitations*

Stability issues in the behavior of a few software components of the existing biplane system software processing pipeline motivated another large part of this thesis. Camera calibration at CLiMB has been performed using a single image frame of a cube calibration object. A template-matching localization method was used to identify its 15 control points in the noisy image. The camera's projection was modeled using the Direct Linear Transformation method, which has widespread use in optical motion capture systems used in biomechanics research [57-60]. For reasons described in subsequent chapters, this combination of calibration object and algorithm produced unstable camera parameters. These errors directly affect the accuracy of bone tracking in the system. Characterizing and mitigating these camera calibration errors is a core aim of this thesis.

Another instability issue manifested in the model-based bone tracking software. Bones in relatively well-exposed images tracked poorly as a symptom of suboptimal 3D-2D registration performance. This issue was most notable at the periphery of the X-ray portals. Due to the limited diameter of the standard C-arm image intensifiers (31 cm), the typical adult foot will not entirely fit within the capture volume of the biplane system. Tracking of clinically important

regions of the foot such as the ankle or toes relies on accurate tracking at the edges of the capture volume. The spatial distortions, intensity roll-offs, and defocusing blurs that are strongest at the periphery of the fluoroscope portal introduce noise and biases into intensity values of the image. Errors in image intensity values will affect the comparison and registration algorithms during bone tracking. The normalized cross-correlation score between the DRR and the fluoroscopic image (or an edge-enhanced version of both) is the metric used as the primary driver for the 3D-2D registration optimization routine. While these metrics have been used in other DRR-based biplane tracking systems [39, 61, 62], its current implementation in DRRACO behaved poorly, seldomly exceeding a value of 0.5 (on a scale of 0 to 1). Such behaviors hint at underlying issues of signal noise and inadequately utilizing equipment bandwidth. A more robust and stable image similarity metric would allow for faster convergence of the 3D-2D registration optimization during bone tracking. This would increase data processing throughput and reduce errors in kinematic data.

2.5 IMMEDIATE NEEDS OF THE LABORATORY AND AIMS OF THIS THESIS

Biplane imaging systems are expensive, present myriad technical challenges, and produce enormous volumes of data. They are also among the most powerful tools in a biomechanist's arsenal for in vivo human motion quantification. While dedicated, high-speed imaging hardware is preferred for the best image data, a lot is still possible with the older, lower-cost fluoroscopy-based designs, such as CLiMB's biplane system. Although these setups have all the technical limitations of fluoroscopic systems listed earlier, they are the first option for newly minted biomechanics laboratories interested in bone tracking. Regardless of hardware, the keys to success with biplane systems are: 1) optimizing the system for the imaging problem of interest, and then validating its accuracy with an acceptable standard, 2) utilizing radiation in a manner that is responsible for both research study participants and equipment operators, and 3) maximizing processing throughput via automation and careful organization of the data. The signals associated with biplane X-ray imaging are non-linear, noisy, and require careful and extensive processing to be made useful. Additionally, this custom, unconventional assembly of hardware creates many additional sources of signal degradation. The primary aim of this thesis is to gain insight into the sources of error for this biplane system experimentally and increase the

accuracy of the virtual biplane imaging chain model wherever possible. The secondary aim is to establish the experimental protocols for acquiring and processing foot and ankle image data that minimize radiation risk and automate the data workflow with as little user interaction as possible. Biplane systems generate extremely large volumes of data with a high computational burden. Steps are taken, whenever possible, to develop fully automated algorithms or ones that exploit parallel processing technology maximize throughput. The last aim is to establish a series of benchmarks for quantifying and comparing the performance of the biplane system to other systems and future upgrades and modifications to CLiMB's hardware and software. Biplane radiography is a rapidly evolving tool in the biomechanics realm. As new algorithms or hardware undergo development, having an established protocol for quantifying system accuracy and performance is important. This also allows for direct, objective comparison between systems or versions of systems. Even seemingly basic imaging parameters, such as camera positioning or exposure settings may have substantial effects on system accuracy. Establishing a formal pipeline for data processing and measurable benchmarks for quantifying system performance ensures confident conclusions can be drawn from the data and solidifies biplane fluoroscopy as a viable biomechanics research tool.

Chapter 3. FLUOROSCOPIC IMAGE PROCESSING

3.1 INTRODUCTION

Raw fluoroscope images are inherently noisy signals with large spatial and intensity distortions. The fluoroscopic imaging chain is a cascade of quantum-limited interactions that severely degrade the input signal (X-ray source) that itself is governed by a random (Poisson) process. This chapter outlines major sources of these signal corruptions and the preprocessing steps required to make use of data acquired from this custom combination of camera and image intensifier hardware. The unknowns resulting from modifications to the native Philips BV Pulsera fluoroscope hardware configuration outlined in Chapter 2 motivate the experiments of this chapter. The results of these experiments provide important benchmarks for comparison to other biplanar X-ray analysis systems and to inform best practices for in vivo studies.

3.2 IMAGING CHAIN LINEARITY

The Philips fluoroscopy X-ray source control software confounds the X-ray source potential voltage (kV) and source current (mA) parameters into pre-defined curves (Figure 3.1). While these are in place to prevent a user from damaging the X-ray source, they hinder flexibility in setting the imaging parameters. A high enough source current is one that provides a sufficient number of photons for a useable image while balancing the increases in radiation received by subjects and operators. Additionally, the radiographic contrast between low- and high-attenuating tissues is modulated by the tube potential voltage (kV). Since the Philips onboard image exposure control and processing hardware is not available in our modified C-arm, all imaging parameters must be set manually and matched appropriately between the fluoroscope and separate high-speed camera. Without communication between these two systems, we cannot assume linearity between fluoroscope parameters and the resulting grayscale camera images.

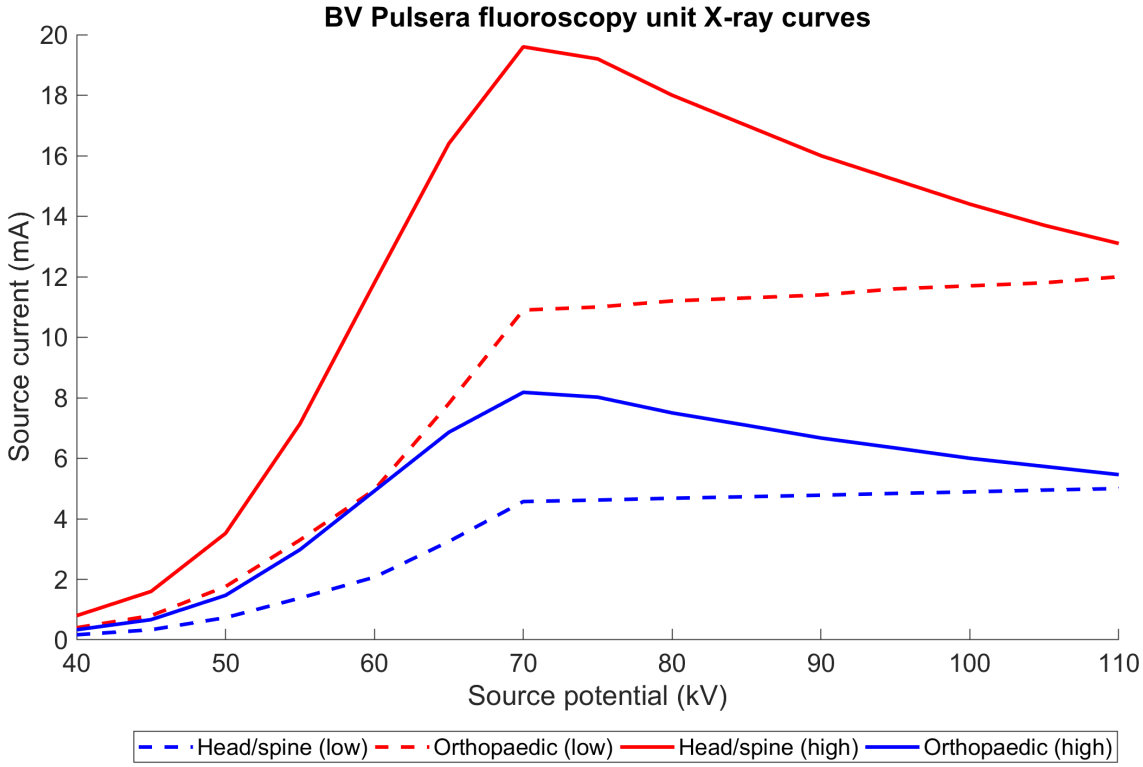

Figure 3.1. Philips manufacturer's fluoroscope tube parameter curves tube potential (kV) vs. tube current (mA) for two pre-defined curves: "Orthopaedic" and "Head/Spine" modes. Within each mode is a "high" and "low" power output mode. The maximum tube output of the BV Pulsera source is under 20 mA ("Head/Spine (high)" mode). The tube voltage is coupled to the tube current, meaning to obtain maximum source current for the best image quality, a narrow band of tube potentials (68 - 72 kV) is available. Additionally, decreases in tube current would require either opening the lens aperture (f-stop), which may decrease focus; increasing camera gain, which may increase noise; or the exposure time, which may introduce motion blur.

Therefore, the first investigation of the biplane system was to verify the linearity of the output (pixel intensities) as a function of input variation (voltage or current). This verification was performed by imaging an X-ray electron density phantom (Gammex 471, Middleton, WI) with a

sweep of X-ray source parameters (kV and mA) using constant camera parameters that are used for typical human gait imaging (exposure time of 1997 microseconds, 896x1152 resolution). Regions of interest were defined inside the temporal-average of 4000-image sequences of cylindrical phantom plugs that represent various real-world tissues (Figure 3.2). For each tissue plug, pixel intensities inside a region of interest in the center of the plug were averaged to obtain the change in recorded fluoroscopic signal as a function of input voltage and current.

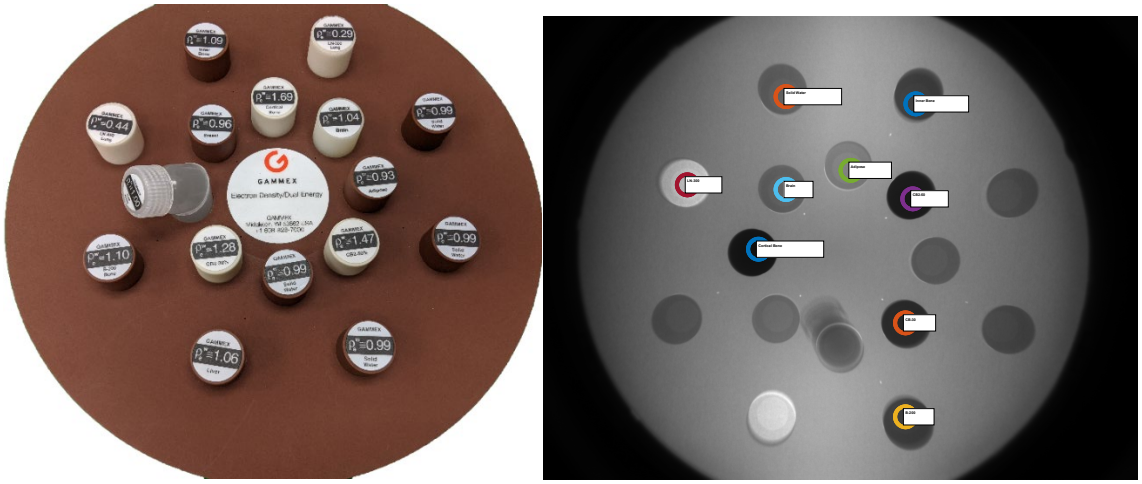

Figure 3.2. Fluoroscopic image (right) of the Gammex 471 Electron Density phantom (left). Mean camera pixel intensities were calculated within the circular regions of interest for each simulated tissue plug.

The camera intensity linearity was determined for a subset of the tissue phantom plugs via linear regression of the intensity values normalized to the maximum possible 12-bit intensity value. (Figure 3.3). The standard operating range of the system is between 60 and 70 kV for in vivo joint imaging, as these portions of the source curves provide the maximum current (19.2 mA). Within this interval of tube voltages, the acquired images increased linearly in brightness (Figure 3.3). While the deviation between plugs increases with kV, this behavior is expected given that the attenuation coefficient (μ) of a material is coupled to the voltage of the incident beam. The within-tissue curves are linear, for example, a linear regression of the Solid Water data yields $R^2 = 0.9397$. Low attenuators, like the lung, exhibit more linear behavior over variation in source voltage compared to high attenuators like cortical bone (Figure 3.3).

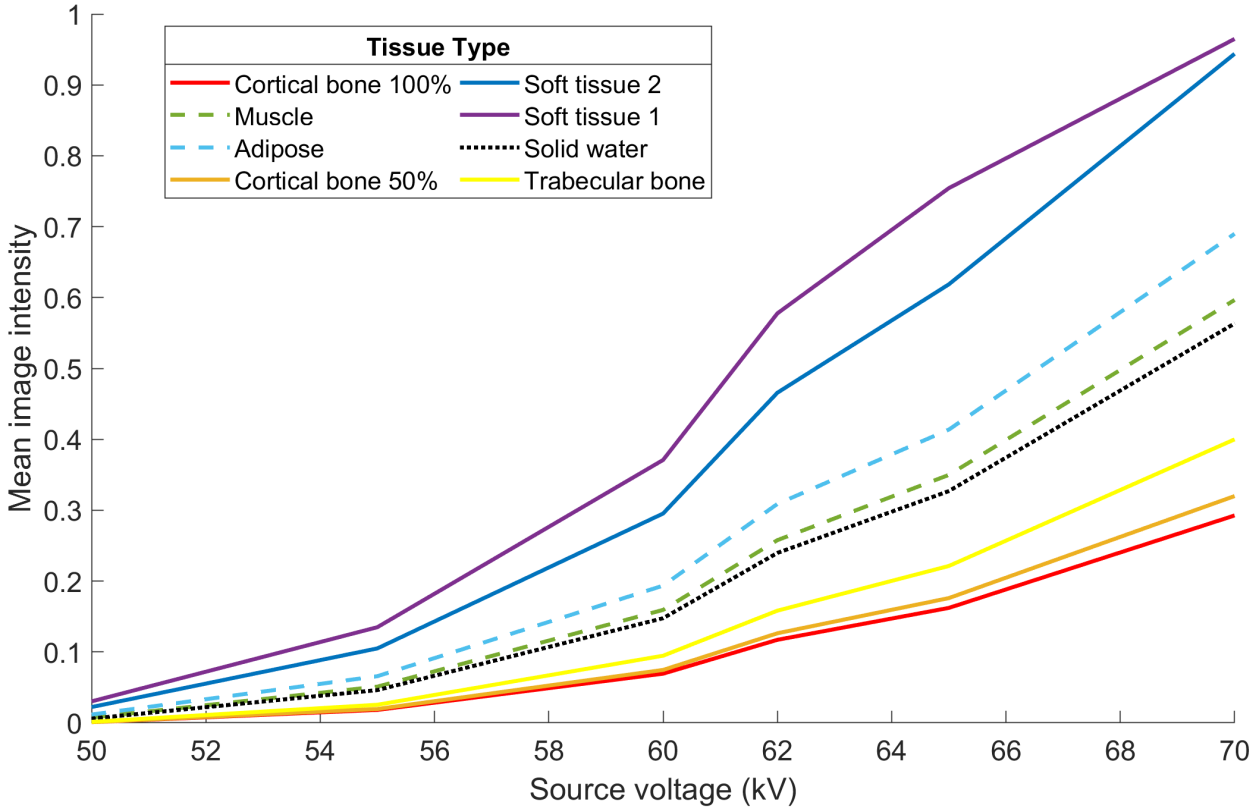

Figure 3.3. Camera intensity linearity as a function of X-ray source voltage for various simulated tissues based on Gammex 471 X-ray phantom data. Materials that attenuate X-rays more, like bone have increased non-linear intensity responses due to the stronger dependency of the attenuation coefficient (μ) on the source voltage (kV). The reduced peaks and greater curvatures of the red, orange, and yellow lines of various bone densities are indicative of this.

3.3 IMAGE INTENSIFIER BRIGHTNESS NON-UNIFORMITY

The X-ray image intensifier introduces a significant and stationary bias of intensity roll-off with increasing radial distance from the center of the detector. An example of this vignetting, or roll-off, is depicted as a surface profile (Figure 3.4). The edges of the X-ray portal have lower pixel intensities. To obtain a uniform exposure across the image, a “flatfield” or “brightfield” correction is performed. Using the same X-ray source parameters (kV and mA) as a data collection trial, a series of images are acquired of the empty capture volume. These images are

averaged across time to reduce high-frequency random noise and obtain a better approximation of the low-frequency detector bias. Remaining high-frequency spatial noise in the time-averaged image may be further reduced by convolving the signal with a local spatial averaging (mean or median) or local maximum kernel. This smoothed surface of pixel intensities represents an approximation of the detector output bias. Using a maximum kernel sets the surface level to an approximation of maximum X-ray fluence in that region of the detector. Revisiting the left side of Equation 2.1, this maximum fluence image (I_o) is used to normalize subsequent trial X-ray images (I) that will have lower intensity values because of an attenuating subject in the field of view. This scaling process effectively converts raw 12-bit camera sensor data into normalized intensities between 0 (full attenuation) to 1 (maximum fluence) as double- or single-precision floating-point data. Correcting the intensity biases improves DRR-fluoroscopic image registration and improves the tracking of objects at the periphery of the image portal. Normalizing to a bounded scale of floating-point numbers gives intuitive meaning to camera intensity values and facilitates setting image processing parameters across different acquisition sessions while reducing the effects of computational integer roundoff errors. The sliding convolution kernel approach to approximating the signal flatfield bias attempts to reduce the high-frequency noise that was not removed from the temporal averaging of multiple flatfield exposures. The smoothness of the flatfield surface is a function of kernel size that must be determined empirically. The flatfield surface could alternatively be estimated by fitting a polynomial or spline surface to the 2D intensity data. However, with an adequately large kernel size, the sliding-maximum method effectively captures local, low-frequency intensity response while suppressing outlier pixels that may cause ill-fitting surface models. In the case of a flatfield image that is corrupted or contains some hardware that cannot be removed from the image (e.g., a strain gauge embedded in the X-ray transparent walkway), a global surface model may be required if the outlier pixel regions cannot be in-painted. The image non-uniformity will vary as a function of source voltage too since the absorption phenomena of the cesium-iodide image intensifier screen is itself a function of the incident energy level. This means that a separate flatfield image, with a different roll-off characteristic, should be collected for each tube potential (kV) setting used during data acquisition.

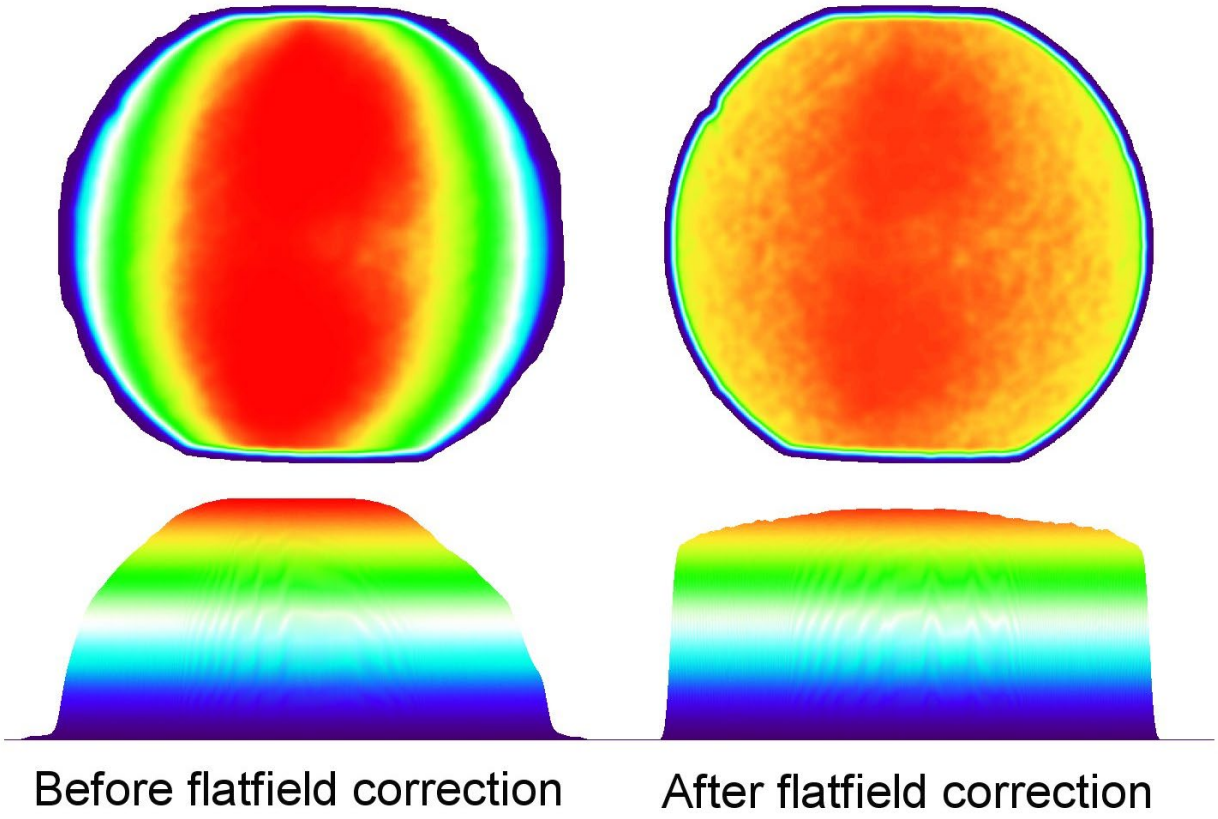

Figure 3.4. Image intensity before (left) and after (right) a flatfield correction has been performed, shown as heat-map images (top) and surface renderings (bottom). Note the strong intensity roll-offs at the edges of the image prior to correction. Flatfield correction is not perfect, but significantly reduces the low-frequency, spatially varying intensity bias inherent in the image intensifier.

3.4 SPATIAL RESOLUTION AND FOCUSING

The image intensifier output screen is coupled to the high-speed camera with a telephoto lens. The lens must be manually focused by imaging a line pair object (Figure 3.5) placed on the image intensifier face. The line pair object contains alternating pairs of attenuating metal bars and air gaps at increasing spatial frequencies. The magnitude of the intensity differences between adjacent light and dark pairs is the signal modulation. It decreases with narrower spacing (higher spatial frequency of line pairs per millimeter) until the bands appear as a homogenous gray

region of zero modulation. The lenses of the biplane system are focused until the maximum number of line pairs per millimeter is resolved. In the most frequently used configuration, the pixel sizes of the fluoroscopic images are equal to about 1/3 millimeter (1000x1000 effective image matrix and 30 cm image intensifier diameter). The maximum intensity modulation frequency that can be resolved is a single pair of pixels, one dark and one bright, for a total distance of 2/3 millimeters per cycle. Therefore, taking the inverse of this value, the theoretical resolving limit of this system in the normal image intensifier magnification mode and at full camera resolution is 1.5 cycles per millimeter. This gives us a target value to check if the system has been adequately focused.

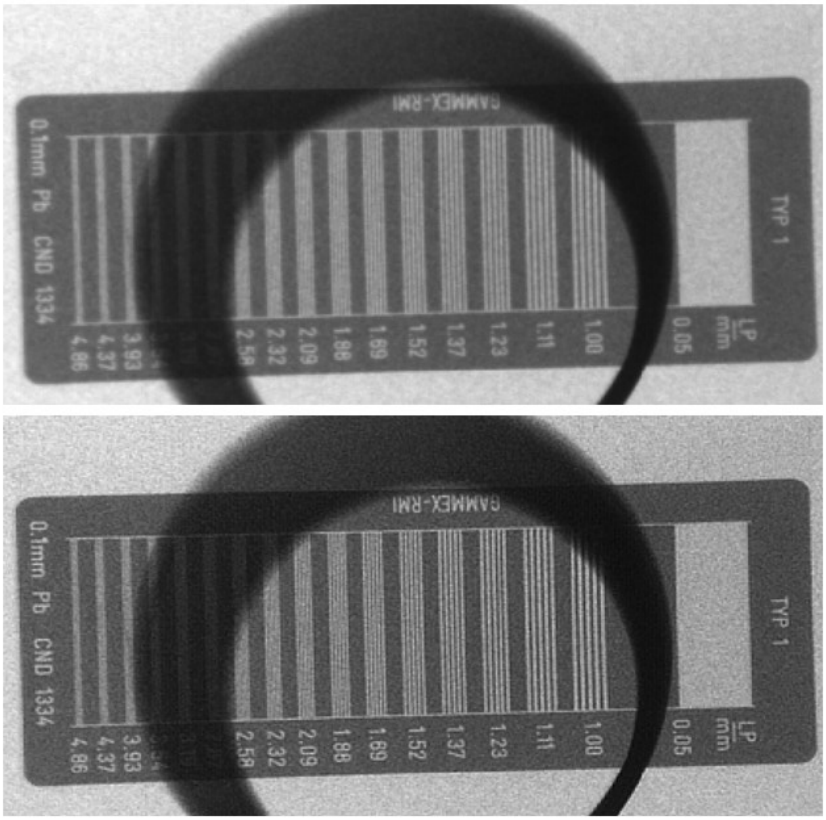

Figure 3.5. Images of a line pair object mounted normal to the image intensifier surface but offset 15 cm using a plastic pipe (the black circular object) to confirm focus at the level of the walkway. The top image is a raw fluoroscope image using a focused lens with a large aperture (f-2.8). The modulation between line pairs decreases with increasing line pair density, as shown by the increasingly homogeneous gray line pair regions on the left half of the object. The bottom image is the top image after post-acquisition deconvolution with the Lucy-Richardson algorithm

and an experimental estimate of the optical chain point spread function to recover the blur introduced by the large aperture (f-2.8). (3.4.3).

3.4.1 *Modulation transfer function*

The modulation transfer function (MTF) is an objective metric for comparing the spatial resolving power of a portion of any optical chain. It is defined as the modulation magnitude as a function of spatial frequency [63]. In the line pair object, this equates to the ratios of 1) intensity contrast between air and metal lines and 2) the spacing between the lines. The final MTF of the complete imaging chain is the product of the MTFs of its constituents; spatial resolving power is limited by the worst-performing component in the optical chain. Measuring modulation in noisy images using a line pair object yields poor estimates of the system MTF. An alternative approach is to calculate the MTF from the edge spread function [64, 65]. The edge spread function is obtained by imaging a sharp, high contrast edge and taking the intensity profile of the resulting image. A block of aluminum was machined to produce a straight edge and was imaged flat against the image intensifier face at a slight (~5°) angle relative to the pixel grid to prevent aliasing of the MTF calculation. The edge spread profile was determined from the Gaussian-smoothed intensity profile of the block edge (Figure 3.6). The edge spread function was differentiated and normalized to form the line spread function. Lastly, a Fourier transformation of the normalized line spread function yields the MTF. From an example modulation transfer function of the CLiMB biplane system (Figure 3.7), it can be seen that the modulation magnitude approaches the noise floor of the system around the predicted maximum resolution (1.5 cycles/millimeter), and so the system is in maximum focus (blue curve). Curves are also shown for the system with the lens out of focus (red) and the in-focus images with an image detail-enhancement performed (green). The focusing of biplane fluoroscopy systems is typically performed by manually adjusting the focusing ring during X-ray exposure of a line pair object. Future efforts will include integrating feedback from real-time calculation of the MTF of the biplane system while imaging the edge spread object to drive the focusing motors of the custom camera lens. These curves provide insight into the contrast of the system for resolving objects of different sizes. For example, the modulation magnitude is reduced to half for a sinusoidal profile 2 mm long (0.5 cycles/millimeter). These data provide useful lower bounds for choosing radiopaque marker sizes that will provide adequate radiographic contrast and reliable localization

on noisy fluoroscopic images (Chapter 4). The MTF curves are an objective metric for comparing the performance of different optical systems, confirming lens focus, and observing changes in signal contrast to image processing or the presence of noise.

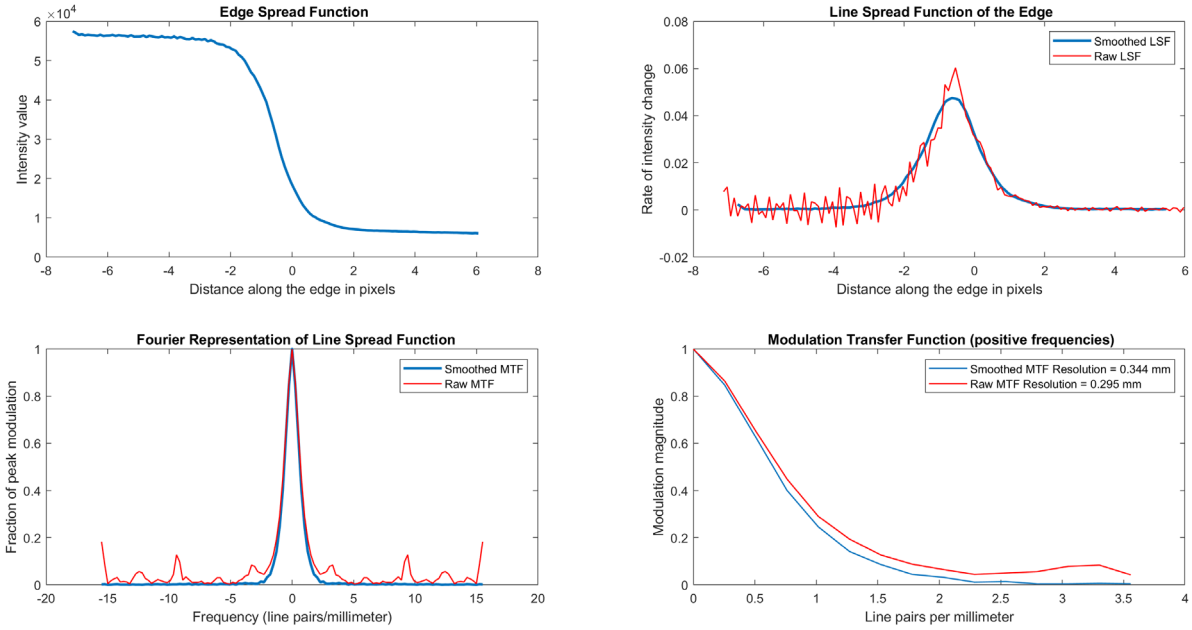

Figure 3.6. The edge spread function of an attenuating metal bar is the intensity profile of the resulting image as a function of the distance across the edge (top left). Taking the derivative of this signal produces the line spread function (top right). The Fourier transform of the line spread function gives the modulation response in the (spatial) frequency domain. Normalizing this Fourier representation and taking the positive frequencies produces the final MTF plot (bottom right) that depicts the amount of modulation the system can detect as a function of spatial frequency. These curves asymptotically decay to the system resolving limit, around 1.5 line pairs per millimeter.

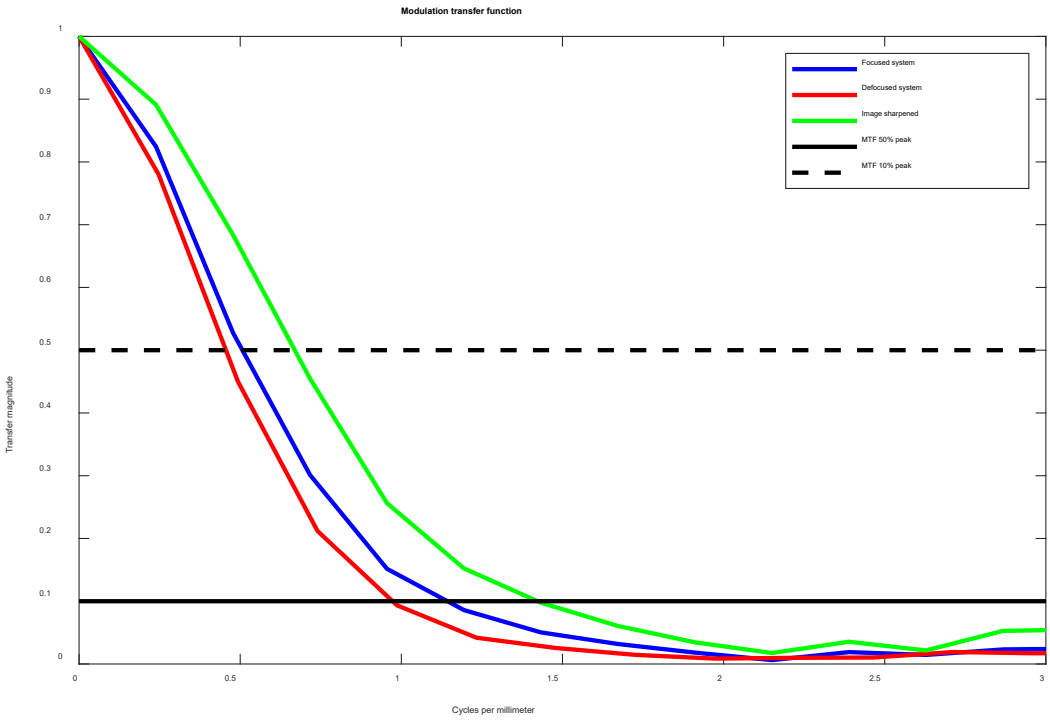

Figure 3.7. Example modulation transfer functions obtained from measuring the edge spread function of the biplane system in three conditions: a focused system (blue), a defocused lens (red), and images enhanced with a sharpening filter (green). The system tapers to zero-magnitude as the spatial frequency approaches the theoretical limit of 1.5 cycles per millimeter. The 50% modulation level (dashed line) intersects the MTF curve at around 0.5 cycles per millimeter.

3.4.2 *Point spread function*

The modulation transfer function is a powerful metric to quantify the blur or loss of resolving power in the biplane system images. This end-effect blur is the combination of the individual blurs and degradations introduced by various components of the image formation process. Examples include veiling glare, lens imperfections, lens aperture defocusing, optical misalignments, and poor focusing. These blurs can each be described by their own point spread function (PSF) and lumped into a single parameter of the optical system, which describes how the input signal of a point source (smaller than the size of a pixel) is “spread out” over the image pixels (Figure 3.8). The lumped PSF was experimentally determined by imaging a 5 mm thick lead bar with a series of holes smaller than the pixel size (0.1 to 0.3 mm) placed on the face of

the image intensifier. These holes are reasonable approximations to point sources since they are smaller than the effective pixel sizes (0.327 mm) and X-ray focal spots (0.6 mm) of the system. Hole quality was inspected with a microscope, and due to manufacturing errors, only the 0.3 mm holes were deemed acceptably circular. The lead portion of the bar blocks most incoming X-rays, but some X-rays pass through the small hole and their signal is spread out beyond the pixel under the hole to neighboring pixels. The PSF kernel is calculated from 10,000 static images that are time-averaged to reduce Poisson noise, cropped, and normalized to the peak magnitude and background (darkfield) signal.

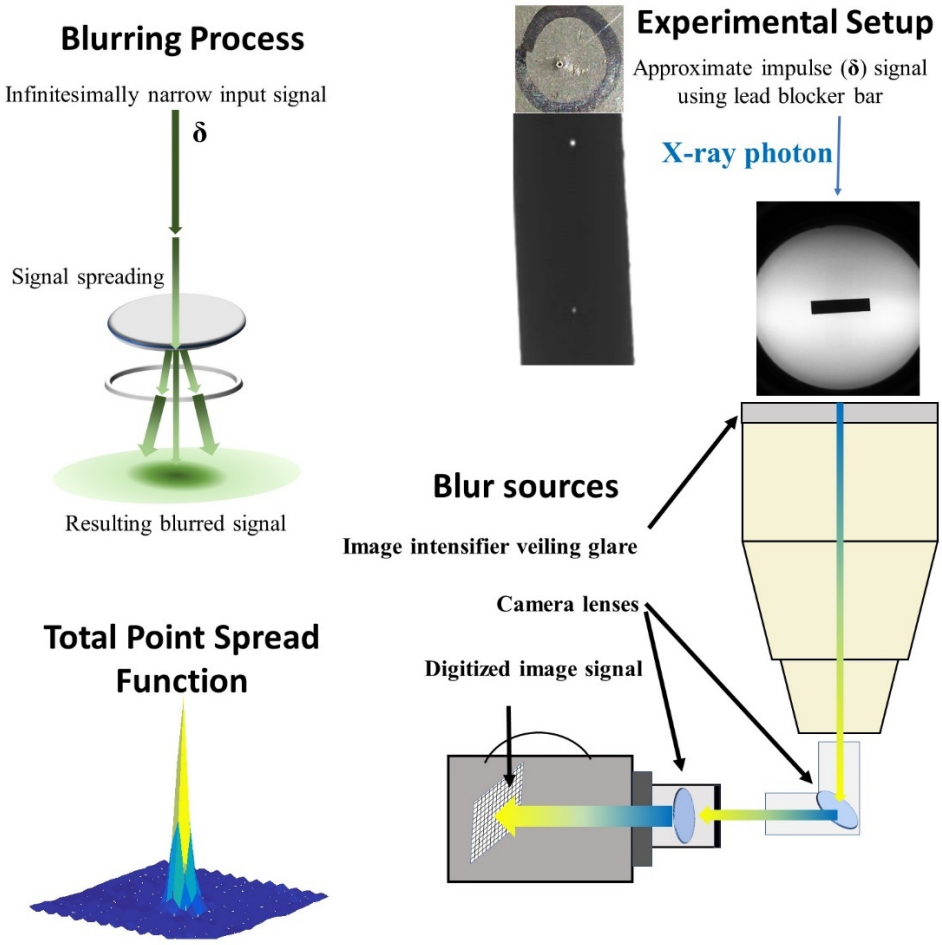

Figure 3.8. The blurring process degrades an impulse input function of an infinitesimally narrow photon into a blurred, quantized sensor recording. This blurring process is a combination of the individual blurs imparted on the image signal by diffraction in the lenses and veiling glare in the image intensifier phosphor screens. The blurring of the biplane imaging chain is described mathematically by its point spread function, which may be empirically determined by imaging a lead blocker bar with a microscopic hole.

3.4.3 *Blur restoration via deconvolution*

Image blurring can be mathematically described as the convolution of an image, **I**(x,y), with a blurring kernel, **h(x,y)**:

$$\mathbf{B}(\mathrm{x},\mathrm{y}) = \boldsymbol{h}(x,y) * \boldsymbol{I}(x,y) \tag{3.1}$$

where **B**(x,y) is the blurred image and "$*$" is the convolutional operator. Starting with the blurred images acquired with the system, **B**(x,y), and assuming an accurate model of **h**(x,y) is known, the ideal deblurred image can be recovered by solving for **I**(x,y). This inverse problem, "deconvolution", is solved in the Fourier domain, where convolution is equivalent to multiplication. Deconvolution can amplify high-frequency noise without appropriate windowing in the Fourier domain or knowledge of the noise power spectrum. The Lucy-Richardson deconvolution algorithm iteratively deblurs images using an initial guess of the blur kernel and is optimized for Poisson noise [66]. The empirically derived PSFs were applied to raw fluoroscopic images to recover blur degradation (Figure 3.9). This method effectively compensates for blur in the images due to misfocusing, veiling glare, and large-aperture defocusing. This is particularly useful for recovering aperture defocusing for wide-open apertures like f-1.8 that maximize the amount of light reaching the camera sensor but impart a blur due to the diffraction limit of the optics. Utilizing large apertures reduces the amount of X-ray radiation (light) needed to expose a subject. PSFs of real optical systems are spatially varying within the image domain, and a separate point spread function exists for each combination of lens blurring and aperture settings. As such, future work will explore the implementation of a spatially varying point spread deconvolution scheme based on a uniform, sparse grid of PSF signatures collected as part of the standard calibration process. The removal of spatially correlated PSF biases in the images is a potentially much more powerful tool than the simple sharpening that results from a single, global PSF kernel deconvolution. Since the raw blurred images are effectively low-passed, it may be preferable, in certain poor contrast situations, to have the data of small markers spread out over multiple neighboring pixels like a Gaussian. Biases in the initial PSF kernel estimate or running too many iterations of the filter will introduce significant deviations of the image intensity values, and risk making the images too noisy for bone tracking. With the achieved detail enhancement (Figure 3.9), these deconvolved images have immediate utility for human operators of the DRRACO software visualizing bones while making initial pose estimates manually and for edge/gradient-based image similarity metrics. With the capability of recovering aperture-induced blurs from the images, we can now open the lenses to the largest apertures, allowing maximum light to reach the camera sensor. This helps to overcome some of the X-ray source power and detector sensitivity limitations. In a pulsed, or high-power X-ray imaging setup,

shorter X-ray pulses (less light) could be used in conjunction with maximum apertures to minimize subjects' radiation exposure.

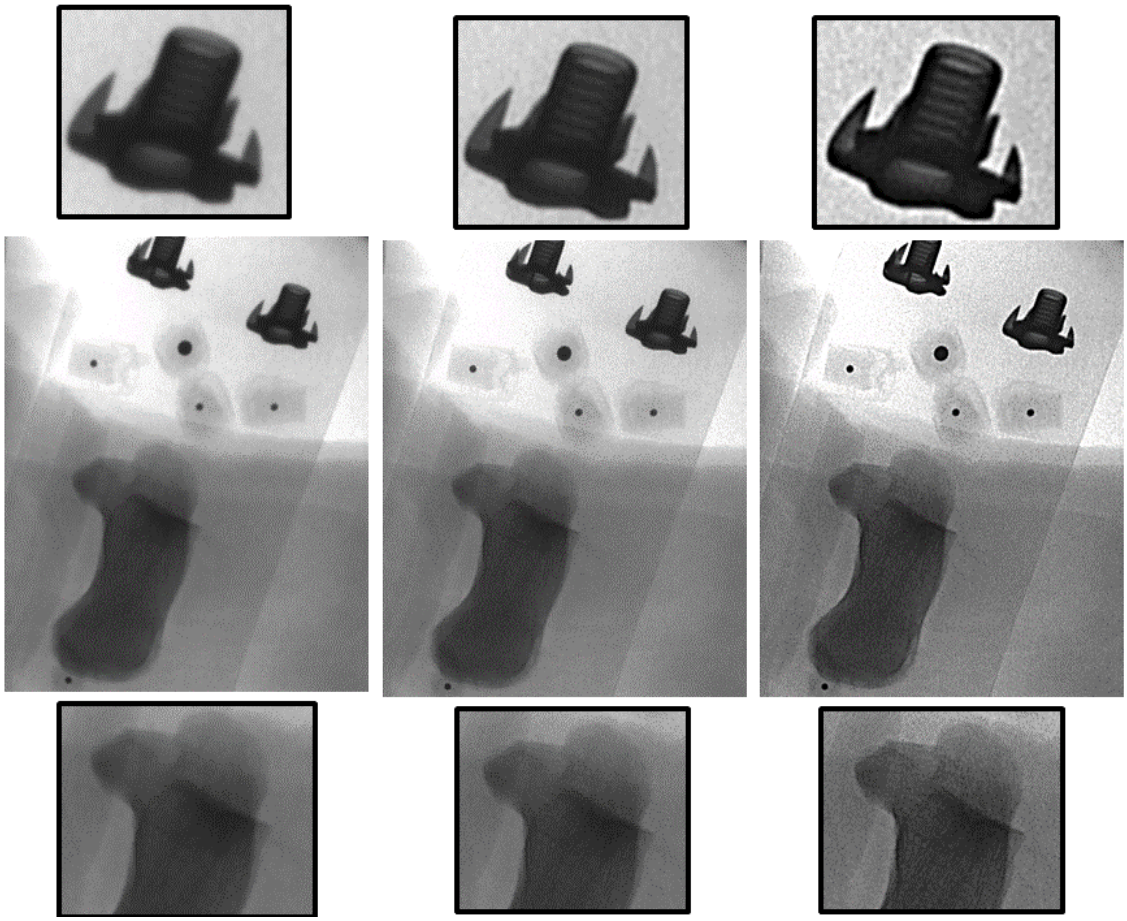

Figure 3.9. Example raw (left), deconvolved, single iteration (center), and deconvolved (five iterations) images of a bone phantom processed with the Lucy-Richardson deconvolution algorithm in MATLAB. Insets of a hardware fastener and calcaneus are shown for detail. Note the increased definition in the fastener threads and the bone edges for the deblurred images. The background of the most processed (right) image is also the noisiest.

3.5 SPATIAL DISTORTION CORRECTION

3.5.1 *Introduction*

The X-ray image intensifier introduces significant spatial distortions to the fluoroscopic images. The curved surface of the input phosphor introduces a radial pincushion distortion pattern, and interference from external electromagnetic fields skews the paths of electrons traveling inside the image intensifier from the input phosphor to the output phosphor. An "S"-shaped distortion results that is a function of the image intensifier's orientation relative to external fields, such as the geomagnetic field [64, 67] (Figure 3.10). Methods for correcting this distortion utilize images of regular grids of control points to generate a spatial warping map from each pixel in the raw images to the ideal, undistorted pixel locations in the corrected images [68-75]. These distortion correction algorithms may be grouped into the broad categories of either local or global models. Local models divide the image into smaller patches and fit piecewise functions of low order (n = 2) between neighboring control points [71] or model the warping of the image as a thin plate spline surface [76]. They have the advantage of accurately characterizing local distortions and produce a lower minimum model error. However, the accuracy of these local reconstructions is heavily dependent on accurate localization of control points on noisy fluoroscopic images. Global correction models are based on polynomial models of the distortion [70, 77, 78] and are more robust to errors in the localization of the grid control points. However, polynomial models of distortion must be of a sufficiently high enough order to characterize higher-frequency local distortions while avoiding overfitting that may introduce errors into the corrected image. Each type of model can account for both the radial and sigmoidal distortions in fluoroscopic images.

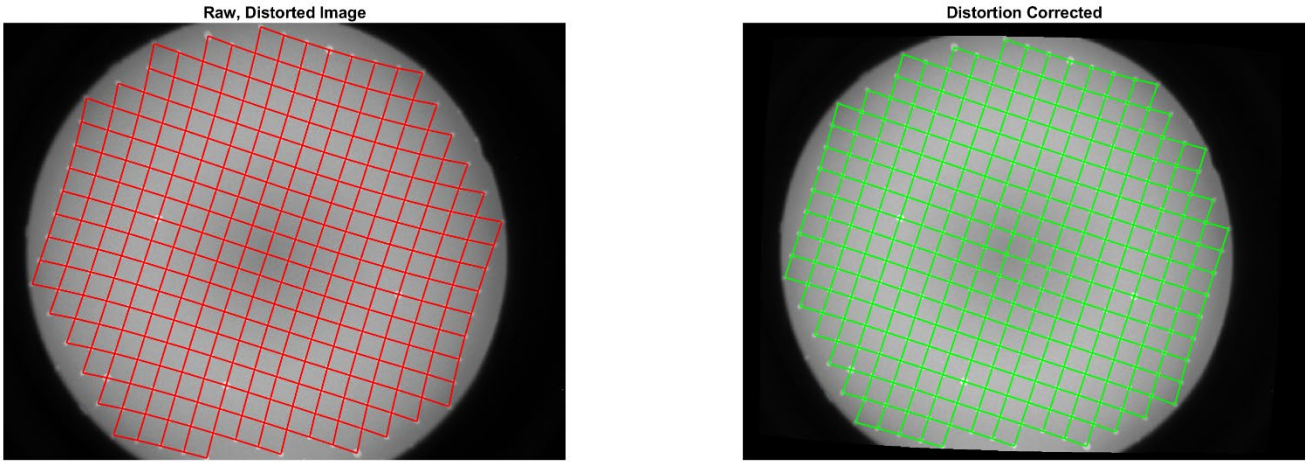

Figure 3.10. Image intensifier images prior to (left/red) and after (right/green) a distortion correction procedure. Note the spatially varying, swirling, "S"-shaped and pincushion distortions in the raw image. The corrected image has been returned to a nearly perfect rectilinear grid.

In interoperative fluoroscopic imaging, the presence of objects or equipment that produce radiofrequency interference, and the frequent repositioning of the system introduce a wide range of distortion profiles. For biomechanics research, biplane systems are typically stationary during data collection, and so a single distortion correction mapping may be applied for all fluoroscopic images acquired for a given biplane system configuration. Additionally, the distortion patterns during experimental trials are not expected to have significant local distortion artifacts since there are typically not surgical tools or other items typically present in the shot. Considering this along with the expectation of very noisy fluoroscopic images, a global distortion model that is less sensitive to localization errors of the control points may provide a better spatial distortion correction mapping with less aliasing or artifact.

The current CLiMB practice for performing distortion correction utilizes a thin-plate spline algorithm based on a single image of the calibration grid. Iaquinto, et al. [19] reported a reduction in the root-mean-square error from 6.56 mm to 0.17 mm (0.51 pixels) after correcting images with this method. In this study, the distortion-correction grid was also used as the test object for quantifying distortion correction residuals. Global polynomial corrections for image intensifier distortions were first proposed by Fahrig [77] and Gronenschild [70]. In biplane systems, Guan used a 9^{th}-order polynomial [79] and Lichiti used a 5^{th}-order model for local

image distortions as part of a larger calibration model. Mery described both a 3rd-order polynomial model inspired by Brack [80] and Jaeger [81] and a combination of a hyperbolic surface and sinusoidal distortion model for image intensifier modeling [82]. Using a reference object separate from the calibration grid, Mery reported submillimeter 3D reconstruction residuals for multiple correction models and determined that cubic polynomial models obtained similar results to the hyperbolic model without the added complexity. Kedgley et al. experimented with global polynomial functions of different orders (1st, 2nd, and 3rd) for the x- and y-directions of the distortion grid [78] and compared the results using inter-marker distances of a known bead lattice object. Instead of choosing a single polynomial order for both spatial directions, fitting polynomials of differing orders along the x- and y- directions prevents overfitting of the model along one of the coordinate directions. In addition to inter-marker distances, Lichiti also calculated relative angles between marker vectors as a means of assessing kinematic accuracy from the images [83]. These studies provide evidence that global correction methods may outperform the local, thin-plate spline algorithm used by CLiMB's current software implementation [19]. The aim of this experiment was to use the lab's existing calibration grid object to explore both local and global correction methods, and to compare their ability to accurately reconstruct marker positions which are fundamental components used in biplane studies. These data will directly inform the best practices for correcting image intensifier spatial distortions from CLiMB's current biplane system.

3.5.2 *Methods*

An aluminum plate with a uniform grid (15 mm spacing) of holes [19] was fixed flat on the image intensifier face and imaged for 1000 frames using the typical foot and ankle imaging settings. The mean of these image frames was taken to reduce random pixel noise and provide better delineation of the hole-plate boundaries. To automatically detect the grid holes, the raw image was processed with a morphological "opening" operation (dilation followed by erosion) on the grid image using a structuring element larger than the expected holes. This result was then subtracted from the raw image. That resulting image was filtered to remove small speckle-noise regions and the intensity-weighted centroids of the remaining hole signatures taken as the centers of the grid control points. This typically results in about 250 control points per image with the

current grid, depending on X-ray source settings, camera focusing, and image quality. Distortion correction was performed with three families of algorithms. First, two local models: the thin-plate spline method [18, 19, 76] currently used by CLiMB and a MATLAB Image Processing Toolbox Local Weighted Mean (piecewise 2nd-order functions) routine. Second, two global polynomial models (a 3rd-order in both grid directions, and a 4th-order in both grid directions) that are the "Poly3" and "Poly4" variants of the fitgeotrans.m routine in the MATLAB Image Processing Toolbox were tested. To build polynomial models larger than 4th-order, and mixed polynomial models where x- and y-directions have different orders, custom MATLAB code was written to determine the polynomial coefficients via least-squares regression of the normalized control point locations and the ideal, undistorted grid locations multiplied by a Vandermonde matrix in a manner described by Mery [82]:

An n-th order polynomial of two variables, *x,* and *y*, may be written as:

$$P(x,y) = \sum_{j=0}^{n_y} \sum_{i=0}^{n_x} c_{ij} x^i y^j \tag{3.2}$$

Which may be organized in matrix form as a Vandermonde system, **Vc**=**d**:

$$\begin{bmatrix} x_0^0 y_0^0 & x_0^1 y_0^0 & \dots & x_0^n y_0^0 & x_0^0 y_0^1 & x_0^1 y_0^1 & \dots & x_0^n y_0^1 & \dots & x_0^0 y_0^n & x_0^1 y_0^n & \dots & x_0^n y_0^n \\ x_1^0 y_1^0 & x_1^1 y_1^0 & \dots & x_1^n y_1^0 & x_1^0 y_1^1 & x_1^1 y_1^1 & \dots & x_1^n y_1^1 & \dots & x_1^0 y_1^n & x_1^1 y_1^n & \dots & x_1^n y_1^n \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_m^0 y_m^0 & x_m^1 y_m^0 & \dots & x_m^n y_m^0 & x_m^0 y_m^1 & x_m^1 y_m^1 & \dots & x_m^n y_m^1 & \dots & x_m^0 y_m^n & x_m^1 y_m^n & \dots & x_m^n y_m^n \end{bmatrix} \begin{bmatrix} c_{0,0} \\ c_{1,0} \\ \vdots \\ c_{n,0} \\ c_{0,1} \\ c_{1,1} \\ \vdots \\ c_{n,1} \\ \vdots \\ c_{0,n} \\ c_{1,n} \\ c_{n,n} \end{bmatrix} = \begin{bmatrix} u_0 & \dots & u_m \\ v_0 & \dots & v_m \end{bmatrix} \tag{3.3}$$

Where **V** is the Vandermonde matrix for the system of equations based on the expected grid node locations, **c** is the vector of polynomial coefficients we wish to solve for, and **d** is the array of *m*-distorted grid node coordinates (*u*,*v*). If the number of grid nodes (typically, $m \geq 250$) exceeds the number of coefficients (e.g.: n = 22 generates 231 terms), the system is over-determined and the coefficients, **c,** may be solved for via least-squares regression. Solving Vandermonde systems (Equation 3.3) requires inverting the large matrix **V**; an operation that may be plagued with numerical instabilities. Adding too many polynomial terms to the model without enough control

points can make the matrix ill-conditioned, with no useable solution obtainable. Since the image matrix is on the order of 1000^2 pixels, control point location coordinates vary typically between 100 and 1000 pixel units in the image. Raising these numbers to large powers also leads to numerical instabilities, particularly for the higher-order polynomials. To improve the stability of solving for the polynomial coefficients, grid points are first normalized about their centroid as described by Zisserman [84]. The centroid of the points is subtracted from their locations, and the distances of each point to the new origin are scaled so the root-mean-square distances are equal to $\sqrt{2}$. In another attempt to mitigate errors arising from ill-conditioned matrices, the least-squares system of polynomial coefficients is solved using QR factorization (MATLAB *mldivide.m* routine). This approach allows for the exploration of higher-order global models that may better-capture local distortion phenomena while exploiting the robustness of a global model to small localization errors.

Polynomial distortion corrections were generated for order combinations (image *x*-direction, *y*-direction) ranging from (2,2) up to (20,20) to assess the stability of solving for the coefficients with the roughly 250 data points of the current distortion grid design. The rank and condition numbers for the Vandermonde matrices were calculated for each permutation of polynomial models. For each model type, the Euclidean distances between the ideal grid points and the detected grid points warped into the corrected positions were calculated to assess model fit. To quantify linear accuracy, the inter-marker distances of bead pairs from a separately machined lattice of radiographic markers were reconstructed and compared to their nominal distances (20, 40, 80, 100 mm), which are comparable to typical distances between markers embedded in human foot bones. The marker lattice was imaged at multiple positions and orientations while lying flat on the image intensifier face to fully sample the image intensifier input space while reducing errors from magnification or out-of-plane localization. Geometric accuracy was also quantified by calculating the error between vectors of bead pairs that formed 90° angles in the lattice as a metric for the rectilinearity of the corrected images.

3.5.3 *Results*

The data preconditioning steps led to a substantial reduction in the condition numbers of the Vandermonde matrices (Figure 3.11, left axis). The benefits of normalization are obvious as polynomial orders increase in the x- and y-direction, where the matrices become grossly more ill-conditioned. For a matrix, **A**, with condition number cond(**A**) = x^y, about *y* digits of precision are expected to be lost from the matrix inversion. The preconditioning step limits numerical precision loss to about 6 digits for polynomials up to order n = 15. The extent to which these large-rank matrices introduce errors into the distortion correction map is be assessed by observing the residual fit of the models (Figure 3.11, right axis). Poorly conditioned Vandermonde matrices produce suboptimal polynomial coefficient estimates with the higher residual model errors. The mean grid model residual (Figure 3.11, right axis) is less than half of a millimeter for the normalized data (blue curve, right axis) compared to using the raw pixel locations (red curve). The model fit residual statistics for properly normalized systems were determined (Table 3.1). The custom-written code for solving the Vandermonde systems of arbitrary polynomial order combinations generated the exact same results as their MATLAB Image Processing Toolbox equivalents in the 3rd and 4th -order models. The Local Weighted Mean algorithm and global polynomial models between 9th and 15th -order had similar model residual errors. The root-mean-square and mean residuals were comparable between these algorithms, however, the polynomial models exhibited lower minimum, maximum, and standard deviations in error (Table 3.1). Compared to the other local and global algorithms, the previously implemented Thin Plate Spline algorithm had higher root-mean-square and mean residual errors (Table 3.1) that were at more than twice the effective sensor pixel size (0.327 mm). The best-performing global models and the Local Weighted Mean had mean residuals below the Nyquist frequency of the imaging system (0.327/2 = 0.163 mm).

All algorithms achieved a root-mean-square inter-marker reconstruction error equal to or better than 1 mm (Table 3.2). The local weighted mean and most of the global polynomial models achieved similar accuracies, slightly worse than 0.1 mm (Table 3.2). Inter-marker distance errors were grouped by spatial location on the image intensifier face (Figure 3.12) and calculated as the mean and standard deviations over all locations (Figure 3.13). Polynomials of order greater than n = 4 and lower than n = 16 had similar accuracies and precisions (Figure 3.12), with the 5th, 6th,

9^{th}, 10^{th}, and 15^{th}-order variants yielding similar root-mean-square inter-marker distances (Table 3.2). The highest order models (9^{th}, 10^{th}, 15^{th}) exhibited decreasing grid residuals with increasing model order (Table 3.1), as might be expected, but models higher than 6^{th}-order did not generate lower inter-marker distance errors. The thin-plate spline algorithm tended to uniformly shrink all distances between markers across the image (Figure 3.12 and Figure 3.13), while polynomial models of order 5 and larger and the Local Weighted Mean exhibited a slight overestimation of marker distances. Mixed-order polynomials of insufficient power introduced local artifacts in the bead lattice distance reconstructions, as illustrated by the clusters of similar color (Figure 3.12 and Figure 3.14). Additionally, for these algorithms the orthogonality of the bead lattice was skewed, introducing angle measurement errors exceeding 1° (Figure 3.14) for the simplest global models. Thin Plate Spline angle errors exhibited very low mean, but higher standard deviation compared to Local Weighted Mean and the highest-order polynomial models (Table 3.3).

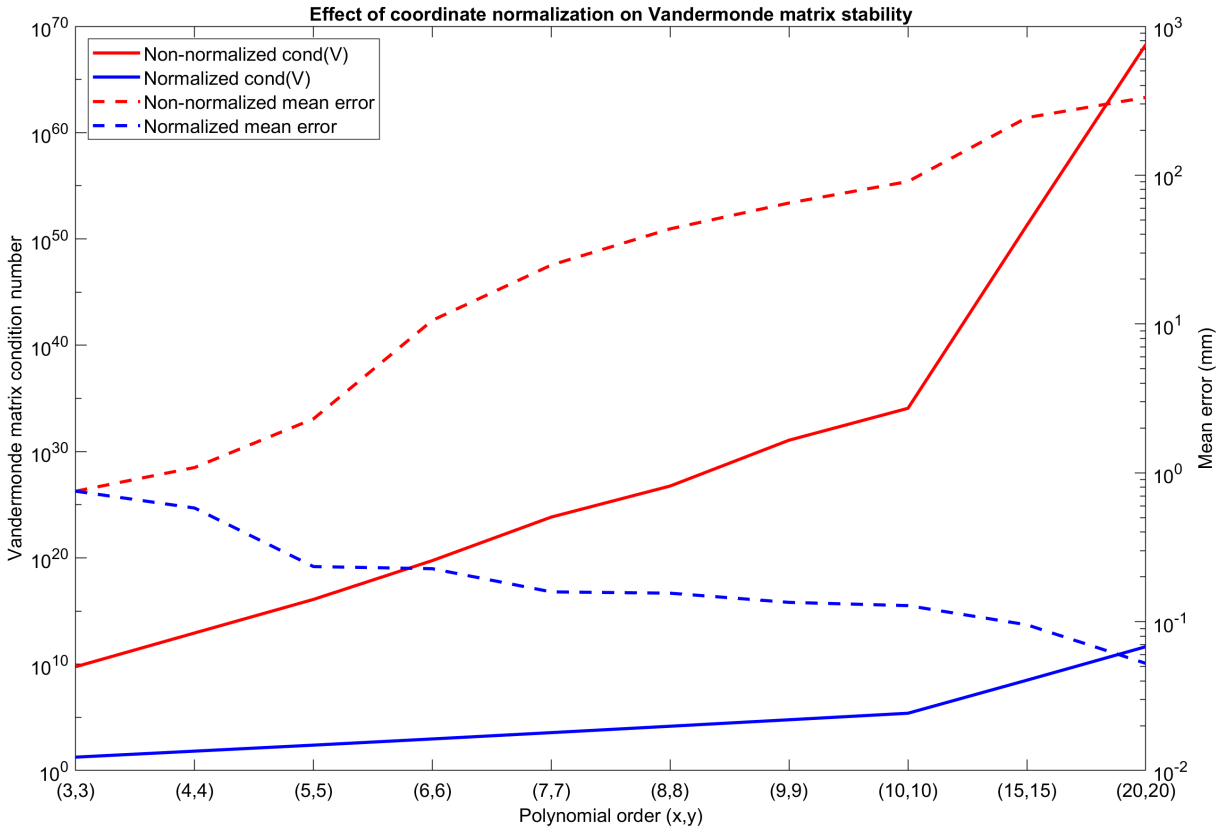

Figure 3.11. The performance and stability of higher-order Vandermonde systems for the global polynomial distortion correction model with (solid) and without (dashed) normalization of the input data. Without normalization, the condition numbers (red lines/ left vertical axis) of the Vandermonde matrices rapidly increase with increasing polynomial order (horizontal axis).

Table 3.1. Distortion correction model fit statistics of grid control point residual errors.

Correction algorithm		Model fit residuals (mm)				
		Mean	Root-mean-square	Standard deviation	Maximum	Minimum
Local	Thin plate spline	0.740	0.810	0.330	1.665	0.067
	Local weighted mean	0.135	0.158	0.083	0.444	0.007
Global polynomial order (x,y)	MATLAB Poly3	0.752	0.851	0.398	2.591	0.029
	MATLAB Poly4	0.580	0.684	0.363	2.198	0.004
	(2,2)	4.409	5.017	2.397	11.622	0.135
	(2,3)	2.806	3.468	2.042	12.858	0.138
	(3,3)	0.752	0.851	0.398	2.591	0.029
	(3,2)	2.638	3.200	1.814	11.850	0.185
	(2,4)	2.798	3.464	2.046	12.959	0.161
	(3,4)	0.693	0.796	0.392	2.329	0.041
	(4,4)	0.580	0.684	0.363	2.198	0.004
	(4,3)	0.599	0.702	0.367	2.462	0.014
	(4,2)	2.599	3.172	1.822	11.711	0.078
	(3,5)	0.586	0.691	0.367	2.744	0.035
	(4,5)	0.396	0.494	0.296	2.412	0.016
	(5,5)	0.234	0.271	0.135	0.896	0.003
	(5,4)	0.361	0.422	0.219	2.111	0.039
	(5,3)	0.428	0.506	0.270	2.800	0.017
	(5,6)	0.229	0.263	0.129	0.816	0.022
	(6,5)	0.227	0.261	0.129	0.735	0.010
	(4,6)	0.393	0.490	0.294	2.421	0.030
	(6,4)	0.352	0.412	0.215	1.906	0.048
	(3,6)	0.584	0.690	0.367	2.708	0.021
	(6,3)	0.426	0.502	0.266	2.639	0.039
	(6,6)	0.227	0.261	0.129	0.739	0.003
	(7,7)	0.159	0.180	0.085	0.446	0.005
	(8,8)	0.155	0.175	0.082	0.414	0.013
	(9,9)	0.135	0.151	0.069	0.374	0.005
	(10,10)	0.128	0.145	0.069	0.344	0.006
	(15,15)	0.095	0.112	0.060	0.307	0.001
	(20,20)	0.052	0.068	0.044	0.183	0.000

Table 3.2. Distortion correction model lattice inter-marker distance reconstruction performance.

Correction algorithm		Lattice inter-marker distance errors (mm)		
		Root-mean-square	Mean	Standard deviation
Local	Thin plate spline	0.222	-0.181	0.127
	Local weighted mean	0.116	0.092	0.070
Global polynomial order (x,y)	MATLAB Poly3	0.214	0.108	0.185
	MATLAB Poly4	0.203	0.108	0.172
	(2,2)	1.000	-0.331	0.943
	(2,3)	0.831	-0.247	0.793
	(3,3)	0.214	0.108	0.185
	(3,2)	0.514	0.050	0.512
	(2,4)	0.829	-0.246	0.792
	(3,4)	0.207	0.112	0.174
	(4,4)	0.203	0.108	0.172
	(4,3)	0.203	0.107	0.172
	(4,2)	0.509	0.048	0.507
	(3,5)	0.188	0.118	0.146
	(4,5)	0.176	0.110	0.138
	(5,5)	0.114	0.081	0.080
	(5,4)	0.122	0.081	0.091
	(5,3)	0.121	0.080	0.091
	(5,6)	0.113	0.081	0.079
	(6,5)	0.112	0.080	0.078
	(4,6)	0.175	0.109	0.137
	(6,4)	0.121	0.080	0.091
	(3,6)	0.187	0.117	0.146
	(6,3)	0.120	0.079	0.091
	(6,6)	0.112	0.080	0.078
	(7,7)	0.118	0.094	0.071
	(8,8)	0.117	0.094	0.070
	(9,9)	0.119	0.095	0.072
	(10,10)	0.119	0.095	0.072
	(15,15)	0.122	0.094	0.077
	(20,20)	0.407	0.058	0.403

Figure 3.12. Inter-marker distance errors (millimeters) of the bead lattice object using various global and local distortion correction schemes (columns). Rows are marker pairs grouped by location in the image plane, hence the clusters of similar over-estimation (green) or underestimation (brown). These clusters represent local biases in the corrected images introduced by an insufficient level of complexity of the polynomial model. Stable distortion corrections with small errors are possible with using polynomials as high as 15th-order due to the normalization preconditioning procedure.

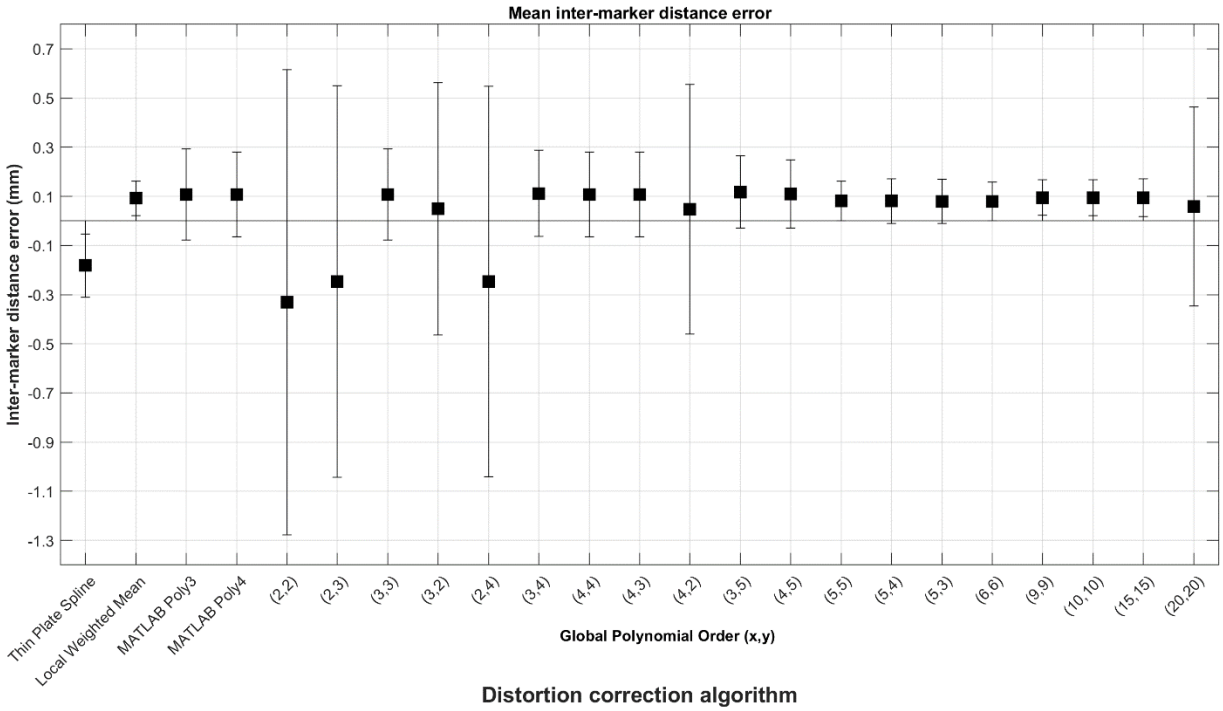

Figure 3.13. Mean and standard deviations of inter-marker distance errors (millimeters) of the bead lattice object using various global and local distortion correction schemes. Local weighted mean and global polynomials of order $5 > n < 20$ produced the smallest errors. The thin-plate spline algorithm tended to underestimate inter-marker distances.

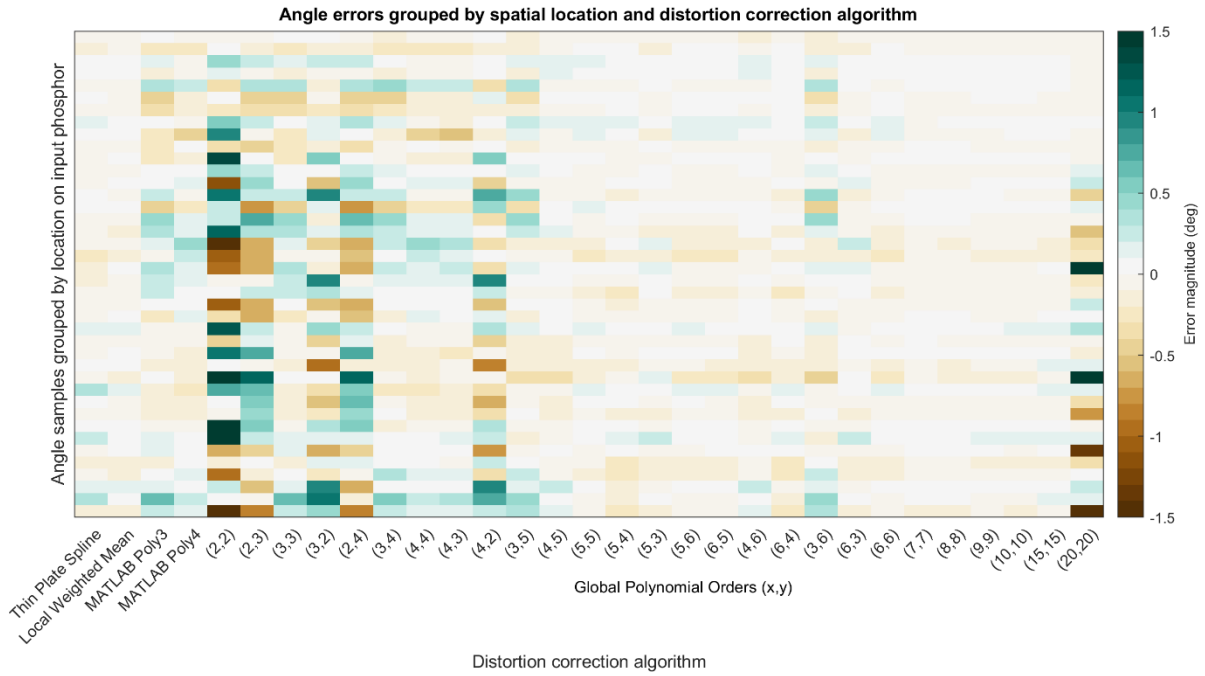

Figure 3.14. Inter-marker angle errors (degrees) of the bead lattice object using various global and local distortion correction schemes (columns). Marker pairs (y-axis) are sorted by location in the image plane, yielding clusters of similar overestimated (green) or underestimated (brown) 90° angle pairs. Clusters represent homogeneous local biases in the corrected images caused by an insufficient level of complexity of the polynomial model. Overfitting errors appear for the highest-order polynomial (20,20).

Table 3.3. Distortion correction model lattice inter-marker angle reconstruction performance

Correction algorithm		**Lattice inter-marker angle errors (degrees)**		
		Root-mean-square	Mean	Standard deviation
Local	Thin plate spline	-0.003	-0.003	0.114
	Local weighted mean	-0.014	-0.014	0.078
Global polynomial order (x,y)	MATLAB Poly3	0.018	0.018	0.235
	MATLAB Poly4	-0.004	-0.004	0.170
	(2,2)	0.114	0.109	0.972
	(2,3)	0.006	0.005	0.481
	(3,3)	0.018	0.018	0.235
	(3,2)	0.027	0.026	0.448
	(2,4)	0.005	0.004	0.463
	(3,4)	0.022	0.021	0.224
	(4,4)	-0.004	-0.004	0.170
	(4,3)	-0.002	-0.002	0.169
	(4,2)	0.008	0.007	0.421
	(3,5)	0.015	0.015	0.212
	(4,5)	-0.015	-0.015	0.115
	(5,5)	-0.022	-0.022	0.093
	(5,4)	-0.031	-0.031	0.100
	(5,3)	-0.014	-0.014	0.090
	(5,6)	-0.022	-0.022	0.095
	(6,5)	-0.022	-0.022	0.093
	(4,6)	-0.016	-0.016	0.117
	(6,4)	-0.032	-0.032	0.100
	(3,6)	0.014	0.014	0.215
	(6,3)	-0.016	-0.016	0.094
	(6,6)	-0.021	-0.021	0.092
	(7,7)	-0.009	-0.009	0.050
	(8,8)	-0.016	-0.016	0.057
	(9,9)	-0.018	-0.018	0.062
	(10,10)	-0.018	-0.018	0.066
	(15,15)	-0.003	-0.003	0.084
	(20,20)	0.016	0.013	0.740

3.5.4 *Discussion*

This study quantified the performance of X-ray image intensifier distortion correction algorithms using metrics of inter-marker distances and angles on experimental data obtained under typical in vivo imaging settings. Local and global correction families had both mediocre-performing models, like thin-plate spline and 3rd-order polynomial, and better-performing models like local weighted mean and 5th-order polynomials and higher. Pre-conditioning of the input coordinates by normalization is essential to solving the high-order polynomial systems (5th and greater) that accurately account for image intensifier distortion patterns. Unwarping polynomials up to order 15 in each direction were able to stably recover distortions over the image domain. The highest-order model tested (20,20) led to overfitting and unstable image reconstructions with the current number of grid control points available. The MATLAB Local Weighted Mean algorithm had comparable errors to the best global polynomial models, like 5th-order and higher systems. The mixed-order global polynomial models like (2,3), (3,2), (4,5) typically exhibited higher errors than their matched-order counterparts (3,3), (5,5).

The results for this study differ from Kedgley [85], who reported the best performance from a mixed (3,2) polynomial model based on root-mean-square errors of an array of 4 test markers. A possible source for the difference is the number of control points used to form the distortion model. Kedgley used 131 points [78], the present study typically had around 250 reliably detected points, and Pandy reported the results of using a 9th-order polynomial model based on 289 steel beads [79]. With a higher density of grid points (n > 350), presumably the 15th- and 20th- order models would fit better. Mery reported a 3D reconstruction error of 0.15 +/- 0.79 mm using a cubic polynomial model based on a 70-point calibration object, which falls between the RMS and mean lattice errors for the 3rd-order model. Englander used 4500 control points to obtain a mean error of 0.031 mm in a biplane X-ray system [48]. Future designs of the physical distortion grid should explore balancing the density of the grid control points with the expected level of noise that is a function of marker size. Smaller markers provide a higher number of total control points, but each has a smaller signature and thus more uncertainty.

The current recommendation for distortion correction procedures of this biplane system, based on the experimental data is to use the global polynomial model of order 8 or higher. This final

selection of model order could be approached from an optimization standpoint that seeks to find the distortion with the lowest residual, or better yet the lowest error of an independent test object like the bead lattice. While the Local Weighted Mean algorithm has similar, if not slightly better performance compared to the 8^{th}-, 9^{th}-, or 10^{th}-order global models, the global models have lower maximum, minimum, and variance in the residuals and inter-marker errors. Further, the global polynomial form uses the coefficients in the Vandermonde matrix to directly map image pixel coordinates to the correct location. The ability to express both the forward- and back-projection of a point between distorted and undistorted image space as a closed-form matrix multiplication is computationally advantageous compared to the Local Weighted Mean method, which requires interpolating both image location and intensity in the undistorted image using the local piecewise basis functions [82, 86].

The distortion correction results may be further improved with modifications to both the calibration grid and unwarping algorithm. The current distortion correction plate design could be redesigned to have X-ray attenuating grid control points (metallic beads) embedded in a radiotransparent plastic. Additional submillimeter holes for quantifying the spatially varying point spread functions of the image intensifier (Section 3.4.2) could be included in such a design. Holdsworth, et al. describe a distortion correction panel consisting of machined channels in a grid pattern that are filled and evacuated with an X-ray attenuating fluid for use in an operating room environment [74]. This design allows for permanent physical installation of the correction plate on image intensifiers that are difficult to access, as in the case of placing them under a walkway for gait analysis and facilitates rapid and very repeatable grid placement during a data collection session.

Lastly, a few small modifications could also be made to the correction algorithms. The Local Weighted Mean's local basis functions could increase in complexity from the piecewise 2^{nd}-order models to a 3^{rd}-order local patch scheme. However, increasing the local polynomial order too high will to over-fitting artifacts in the resulting grid and cause issues in rectifying the boundary conditions between neighboring grid cells. Presently, the Local Weighting function is purely based on the distance to neighboring grid nodes. In the future, the confidence in control point localization based on inter-frame variances or marker signature size could be added as

weighting terms to the neighboring nodes. The global polynomials could also be expressed in an alternative to the power basis, like the Bernstein or Chebyshev polynomial basis. Bernstein polynomials are orthogonal and defined on closed intervals; a prerequisite met by the normalization procedure. Vandermonde systems expressed in Bernstein form may exhibit even more favorable condition numbers and stability and Bernstein polynomials are easy to interpolate between different orders [87]. The differences in the reconstruction errors between global models (Figures 3.12 and 3.14) might vary more smoothly with increasing polynomial order if expressed in a Bernstein basis, or very high-order systems, like the 20^{th}-order, may achieve a stable and meaningful fit with the existing array of calibration points. Expression of the correcting polynomials in alternative bases will be investigated in the future.

The image intensifiers introduce substantial spatial errors in the image data that may be corrected with experimental data. This investigation quantified the performance of numerous correction algorithms in the context of CLiMB's intended use of the biplane system. By achieving an adequate spatial distortion correction with low residual errors that is robust to image noise, the corrected images may now be quantitatively used for photogrammetric analysis in conjunction with the camera projection models described in Chapter 5 to perform marker-based tracking (Chapter 4) or the model-based tracking procedure (Chapter 6).

Chapter 4. RADIOPAQUE MARKER LOCALIZATION

4.1 INTRODUCTION

Localization algorithms determine marker centroids on a fluoroscopic image from its X-ray projection intensity profile. Surface renderings of example marker intensity profiles were generated (Figure 4.1). A core component of biplane X-ray data processing, markers may be used to determine bone poses, estimate soft tissue deformation fields, or act as control points for calibration objects. A marker localization algorithm should be robust to quantum noise corrupting the pixel intensities and low signal-to-noise ratio, provide repeatable estimates of centroid location, and be computationally efficient.

This chapter describes a selection of marker centroid algorithms that were tested on experimental marker images of a test object to quantify the stability and accuracy of centroid determination in the presence of noise. The results are given and are used to inform the best marker localization algorithm for the CLiMB biplane fluoroscopy system's current intended use.

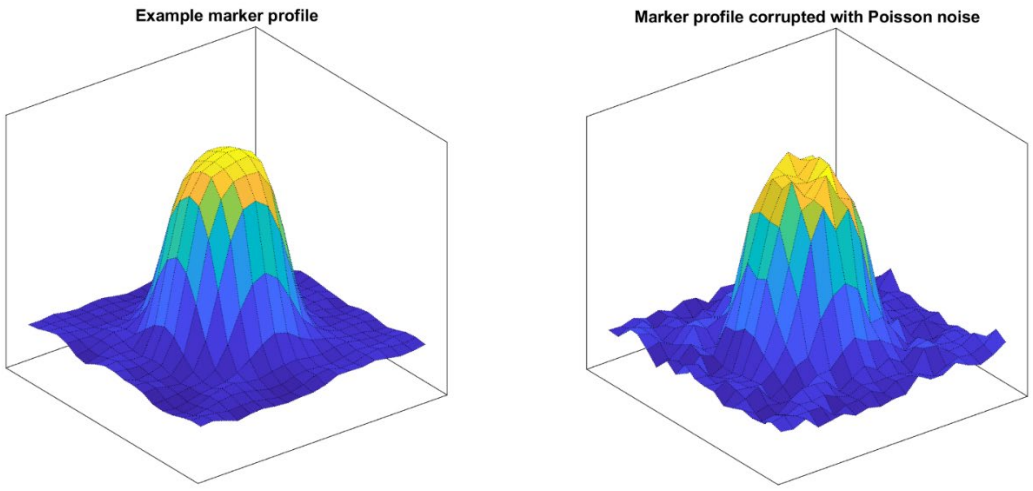

Figure 4.1. Surface renderings of a typical, clean X-ray marker projection profile (left) and the same marker corrupted with Poisson noise (right). In this example of a noise-corrupted marker image, while the general marker shape remains, high-frequency pixel intensity fluctuations amplify background noise and the center of the marker no longer corresponds to the peak of the signal intensity.

4.2 LOCALIZATION ALGORITHMS

Initial guesses of centroid locations are supplied from manual user input or estimated from automatic image processing like intensity thresholding or a feature detector. The following algorithms are used to determine marker centroid location $[\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}]$ in pixels:

4.2.1 *Intensity-weighted center of gravity*

$$\hat{\boldsymbol{x}} = \frac{\sum x * I_{xy}}{\sum I_{xy}} \qquad \hat{\boldsymbol{y}} = \frac{\sum y * I_{xy}}{\sum I_{xy}} \tag{4.1}$$

Where **x** and **y** are the pixel locations of intensities $\mathbf{I_{xy}}$. This algorithm is computationally inexpensive and assumes that the peak of the marker intensity profile corresponds to the center of the marker projection. However, given the high level of noise in the fluoroscopic images (Figure 4.1, right), the random variations in pixel intensities will violate this assumption. This algorithm works best for low-noise, high-contrast images of larger markers that contain more pixel samples.

4.2.2 *Gaussian-weighted center of gravity*

$$\hat{\boldsymbol{x}} = \frac{\sum x * W_{xy} * I_{xy}}{\sum W_{xy} * I_{xy}} \qquad \hat{\boldsymbol{y}} = \frac{\sum y * W_{xy} * I_{xy}}{\sum W_{xy} * I_{xy}} \tag{4.2}$$

Where $\mathbf{W_{xy}}$ is a distance-weighted 2D Gaussian weighting matrix of widths (σ_x, σ_y), centered at the initial guess location (x_i, y_i):

$$\boldsymbol{W}_{xy} = e^{-\left(\frac{(x-\hat{x})^2}{\sigma_x^2} + \frac{(y-\hat{y})^2}{\sigma_y^2}\right)} \tag{4.3}$$

The surface renderings of Gaussians are reasonable approximations to radiographic projections of a metal sphere (Figure 4.2, top row). Gaussian width (σ) is based on the expected marker size in units of pixels. The advantage of this algorithm over 4.2.1 is the distance-weighting that de-emphasizes pixels far from the initial centroid guess that may be corrupted by noise.

4.2.3 *Iterative Gaussian model*

This algorithm utilizes Equations 4.2 and 4.3, but the weighting matrix, $\mathbf{W}_{xy}$, is updated iteratively with each new estimate of the centroid location $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. Iteration is terminated when the shift in estimated centroid location falls below a pre-defined threshold.

4.2.4 *Iterative Sigmoid model*

Borlin, et al. [33] utilized a rotationally-symmetric sigmoidal function model of projection profiles for tracking radiopaque markers in joint replacement components via Roentgen stereophotogrammetric analysis (RSA). This model is defined by Equation 4.4:

$$\boldsymbol{W}_{xy} = \frac{k}{1+\exp(-\beta(d-\alpha))} \quad , \qquad d = \sqrt{(x-\hat{x})^2 + (y-\hat{y})^2} \tag{4.4}$$

The parameter k is the marker intensity relative to the background. Parameter α is the full-width half-maximum of the marker. β is a parameter for marker contrast that tunes the roundness of the profile. Examples of these models for various values of β were determined (Figure 4.2, bottom row). The sigmoidal shape more-closely approximates the sharp contrast and peak flatness of real marker signatures compared to the Gaussian. This algorithm uses the same iterative method as 4.2.3.

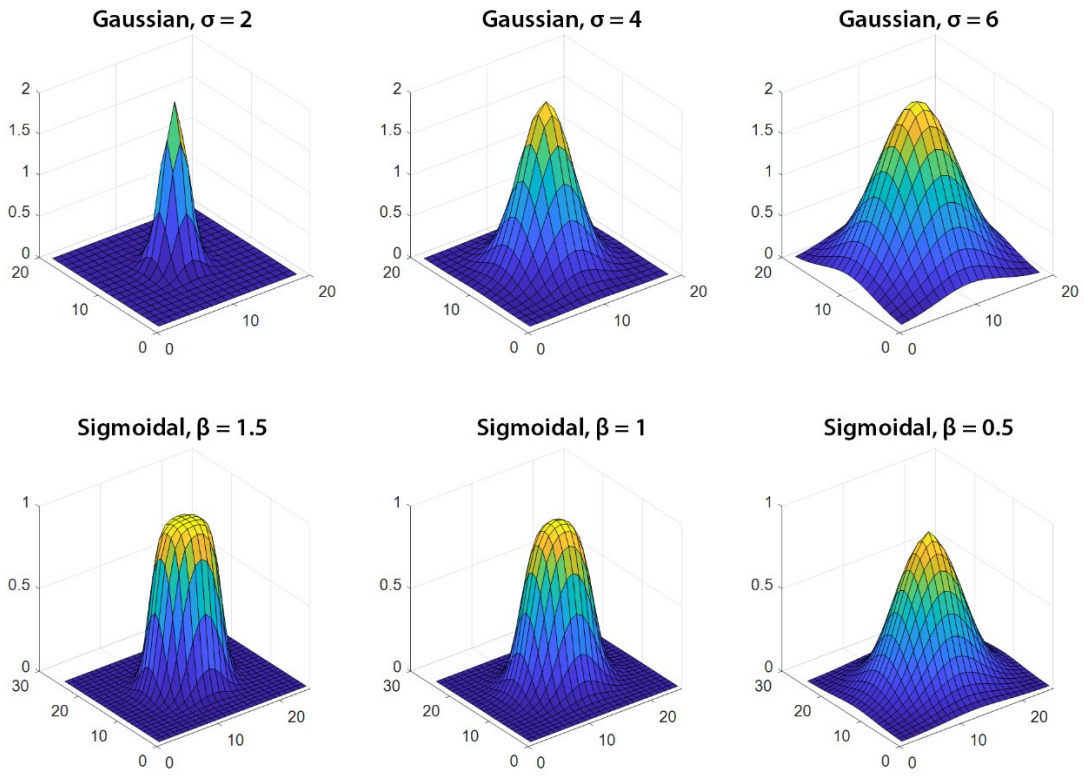

Figure 4.2. Surface renderings of the Gaussian (top row) and Sigmoidal (bottom row) marker models at various values of σ and β, respectively. Note that the Gaussian model with appropriate width has a more pointed peak than the Sigmoidal model and the example marker profiles in Figure 4.1.

4.2.5 *Template-matching Gaussian model*

Template-matching is a procedure that seeks to locate instances of a template within a larger image. As the template slides along each element of the image, the cross-correlation between the two signals is calculated. The location of the peak correlation between a template of a marker projection and the fluoroscopic image is taken as the centroid of the marker. Traditionally, these computations have been performed in the spatial domain via matrix convolution. Guizar-Sicairos et al. have developed an algorithm for performing subpixel registration of a template to an image using Fourier representations of both inputs [88]. By exploiting phase information in the frequency domain, efficient registration is possible in a manner more robust to noise. Templates

of the marker are based on the expected imaged diameter in pixel units used in conjunction with the Gaussian model equation (4.3).

4.2.6 *Template-matching Sigmoid model*

This method utilizes the same cross-correlation template-matching algorithm [88] but uses the Sigmoidal model as the template.

4.3 LOCALIZATION ALGORITHM COMPARISON

The accuracy and sensitivity of these centroid algorithms to initial guess error and the presence of image noise were analyzed on fluoroscopic images of a test object imaged with the parameters typically used for an in vivo data collection.

4.3.1 *Methods*

An array of 21 (5 x 2 mm and 16 x 3 mm) stainless steel beads (tolerance +/- 0.0025 mm) were embedded as a 20 mm grid pattern into an acrylic plate using a computer-controlled mill with an accuracy of 0.002 mm. This lattice of markers (Figure 4.3) was attached flat to the face of the X-ray image intensifier in five different positions spanning the face of the detector. A sequence of 1000 X-ray images of the static object was acquired at each position. Images were processed to correct geometric distortions (Chapter 3), and initial guesses of the marker centroids were determined by using the weighted centroid algorithm (4.2.1) on the temporal average of the 1000 frames of data. Averaging *n* static X-ray images reduces random noise by a factor of $\sqrt{n}$, and as a result bead profiles look more like Figure 4.1 (left). Each algorithm mentioned in Section 4.2 was exercised on the marker beads of each raw frame of the X-ray image sequence using the initial guesses from the temporally averaged image. In total, 105,000 data points (5 positions x 21 beads x 1000 images) were collected for comparison of the six localization algorithms (4.2.1, 4.2.2, 4.2.3, 4.2.4, 4.2.5, 4.2.6).

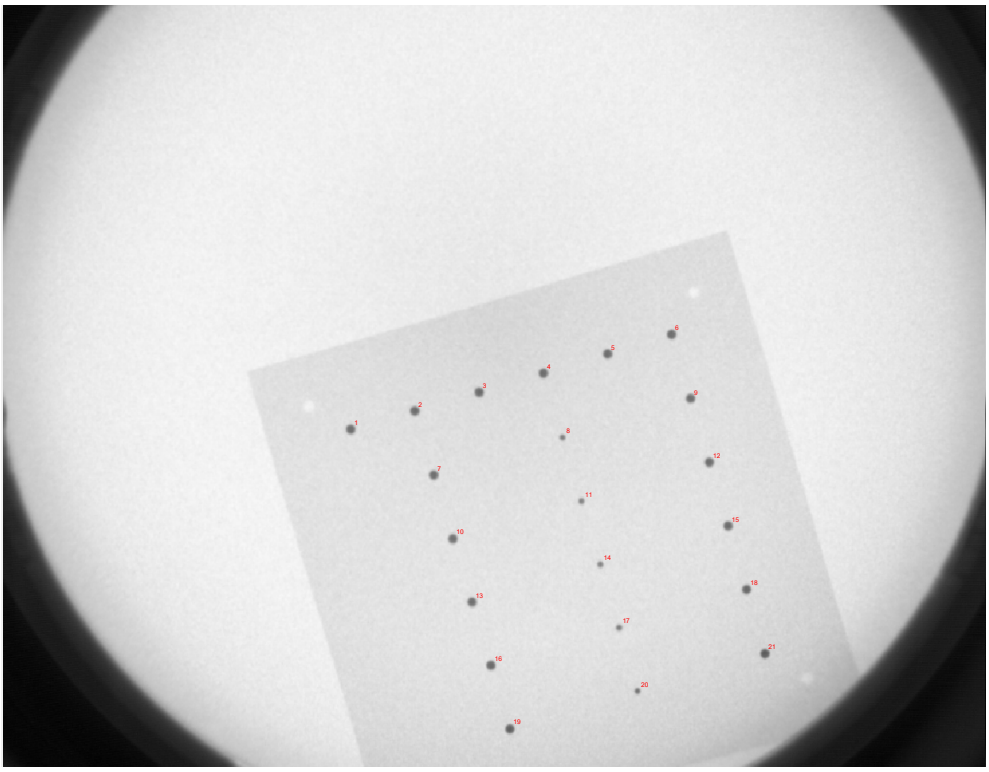

Figure 4.3 Acrylic lattice of stainless-steel beads imaged in one of multiple positions of the X-ray image intensifier face. Outer beads are 3 mm diameter and inner beads are 2 mm diameter. All beads were spaced on a 20 mm regular grid. Bead signatures were between 9 and 12 pixels wide.

Various performance characteristics of each algorithm were quantified using the following experiments:

1.) Robustness to an error in the initial centroid guess was assessed by perturbing the initial guess positions to each algorithm and observing the magnitudes of errors in the estimated centroid location. Perturbations of nine different magnitudes between 0 - 300 % of the bead radius were generated using Monte Carlo sampling (n = 1000, uniform distribution) for each marker. This metric quantifies the ability of the localizer to find the solution with a poor initial guess. Weaker requirements for the initial guess reduce the time burden of an operator manually providing precise coordinate estimates or allows for automated algorithms based on image segmentation and feature detection to perform the task with little or no human input.

2.) Stability of centroid localization was assessed as the variation in centroid tracking over the sequence of static images. Shifts in the centroid estimate across frames are the result of noise in pixel intensities. Image exposure settings during data collection mimicked those utilized for in vivo foot trials to obtain similar noise profiles.

3.) Accuracy of marker localization was quantified by comparing the inter-marker distances between neighboring bead centroids to the nominal manufactured distances. The distances between the centroids of neighboring markers of the lattice grid were calculated and scaled from units of pixels into millimeters. Calculations were limited to the nearest neighbor on the grid to reduce the effects of imperfect distortion correction. Imaging the planar object flat against the detector face minimizes out-of-plane measurement errors in the inter-marker distance calculation.

4.3.2 *Results*

4.3.2.1 Robustness to initial starting guess error

The average errors in final centroid localization for each algorithm are depicted as a function of increasing magnitudes of error in the initial guess (Figure 4.4). Given the 2-3 mm diameter beads and the camera pixel resolution (0.327 millimeters/pixel) used in this experiment, a final error of 3-4.5 pixels means the algorithm converged to a solution within the bead's profile. None of the tested algorithms could repeatably and completely recover the centroid position (zero error) from a perturbed input, and the magnitudes of the output errors generally positively correlated to those of the perturbation magnitude. A notable exception is the template-matching Gaussian method. In this case, the response error was constantly near 1.5 pixels across all perturbation magnitudes. The intensity-weighted algorithm (4.2.1) and the iterative Sigmoidal (4.2.4) algorithms were least sensitive to initial guess errors, with maximum final errors of about 2 pixels. Iterative Gaussian (4.2.3) converged to the best solution for the smallest perturbation magnitude set. Gaussian-weighted center of gravity (4.2.2), iterative Gaussian (4.2.3), and template-matching Sigmoidal (4.2.6) exhibited rapidly increasing error magnitudes with greater error in initial guess. For these three algorithms, initial guesses outside of the bead signature did not converge to inside of the bead radius (3-4.5 pixels).

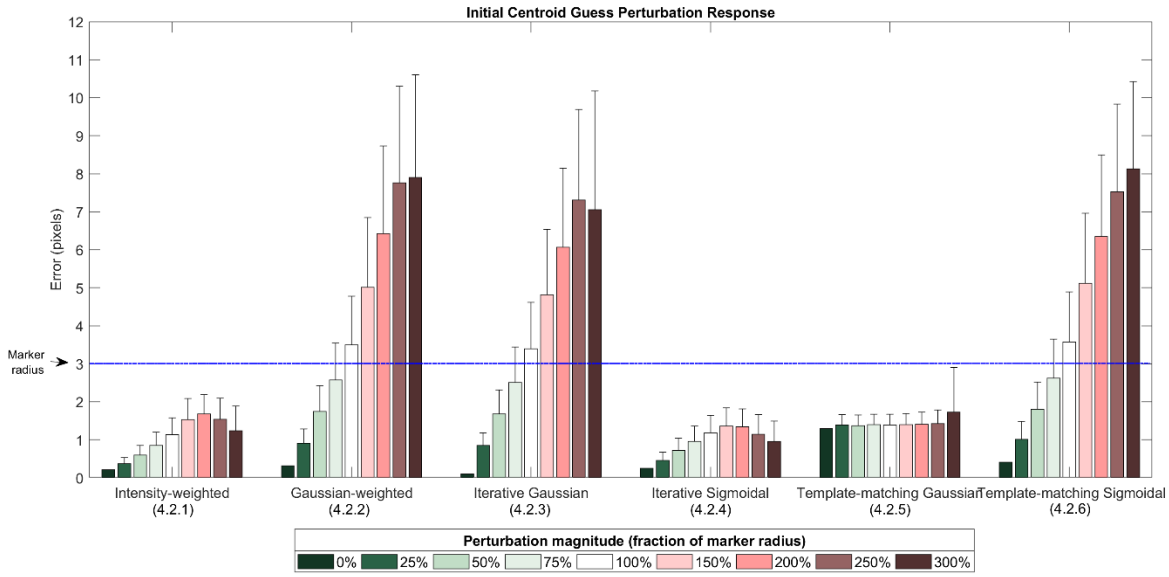

Figure 4.4 Mean final centroid estimation errors for each algorithm as a function of increasing perturbation distance from the best initial guess location. Error bars are the standard deviations across 1000 samples. The equivalent smallest marker radius (3 pixels) is marked with a blue line.

4.3.2.2 Inter-frame centroid stability

Summary statistics of the drifts in static marker centroids across image frames are presented in units of pixels (Table 4.1). The iterative Gaussian (4.2.3) method, with the smallest (subpixel) centroid shifts, was an order of magnitude more stable than the other algorithms. The center-of-gravity algorithms (4.2.1 and 4.2.2) and the iterative Sigmoidal (4.2.4) had similar across-frame stabilities, with average drifts on the order of half of a pixel. The template-matching methods (4.2.5 and 4.2.6) were equally the least-stable algorithms across frames. Mean drifts were larger than one pixel and across-frames variances were also the largest of any method. The template-matching methods were incapable of subpixel stability in the situation of typical fluoroscopic image noise.

Table 4.1. Marker centroid frame-to-frame localization stability statistics of 21 beads imaged in a static position for 1000 frames, in units of pixels.

Algorithm	Mean	Standard deviation	Root-mean-square
Intensity-weighted (4.2.1)	0.607	0.246	0.655
Gaussian-weighted (4.2.2)	0.473	0.217	0.520
Iterative Gaussian (4.2.3)	0.042	0.022	0.048
Iterative Sigmoid (4.2.4)	0.503	0.242	0.558
Template-matching Gaussian (4.2.5)	1.332	0.720	1.514
Template-matching Sigmoid (4.2.6)	1.363	0.819	1.589

4.3.2.1 Inter-marker distances

Errors in the inter-marker distances of neighbors on the bead lattice were calculated for each algorithm (Figure 4.5, Table 4.2).

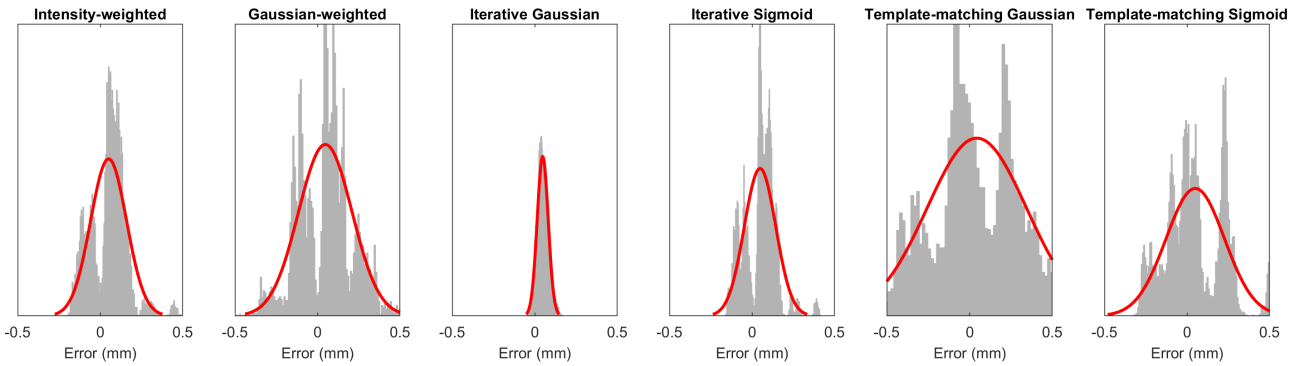

Figure 4.5. Histograms of errors (millimeters) in the inter-marker distances for each algorithm (n=105,000 samples). Fitted normal distributions are overlaid in red. The parameters of these distributions are listed in Table 4.2. The iterative Gaussian method (4.2.3) performed best with a narrow distribution of errors centered close to zero.

Table 4.2. Parameters (μ, mean and σ, standard deviation) of the normal distributions fitted to the inter-marker distance error data (Figure 4.5), expressed in millimeters.

Algorithm	**μ**	**σ**
Intensity-weighted (4.2.1)	0.051	0.109
Gaussian-weighted (4.2.2)	0.047	0.162
Iterative Gaussian (4.2.3)	0.048	0.033
Iterative Sigmoid (4.2.4)	0.048	0.095
Template-matching Gaussian 4.2.5)	0.044	0.301
Template-matching Sigmoid (4.2.6)	0.049	0.177

Comparing the histograms (Figure 4.5, Table 4.2), the iterative Gaussian (4.2.3) error data best-matched a normal distribution, while the other, non-template matching algorithms (4.2.1, 4.2.2, 4.2.4) appear as bimodal distributions of centered at zero error. This dip in the middle of distributions, near-zero error, could be due to the step distance threshold that terminates the iterative process. This effectively establishes a floor for the minimum distance error above zero. The intensity-weighted algorithm (4.2.1) and iterative Sigmoidal (4.2.4) error distributions were skewed slightly towards overestimating the inter-marker distances. As listed in Table 4.1, the iterative Gaussian method (4.2.3) yielded the smallest average distance measurement error (better than 1/20 pixel) and narrow distributions (σ = about 1/10 pixel). The iterative Sigmoid method was comparable in terms of accuracy but worse in precision. These data indicate that Gaussian profiles are a better approximation to the radiographic signatures of the CLiMB's marker beads.

4.3.3 *Discussion*

The ideal localization algorithm is both robust to input error and resolves to the correct solution repeatably, despite noise in the image intensity data. None of the algorithms possess both traits. The intensity-weighted (4.2.1) was the most robust to errors in the initial input guess (Figure 4.4). This makes sense since guess errors cause subtle shifts in the kernel, the values near the center of the kernel are still included in the calculation, while periphery intensity values are added or removed from the calculation. The iterative Gaussian function (4.2.3) converged to the tightest solution of inter-marker distances (Figure 4.5) but was more sensitive to initial guess errors than other algorithms (Figure 4.4). A possible explanation is, with a large enough

perturbation the weighting matrix may pick up outlier values, causing the optimizer to get stuck in a local minimum. Template-matching approaches performed the worst in most tests. They exhibited the worst stability across frames and had the widest distributions of inter-marker errors (σ term in Table 4.2). Perhaps this is because of the limited amount of input data (typically 19x19 pixel patches = 361 data points) available to provide an estimate of the signal frequency content in the Fourier domain, or because the signatures of the initial templates are not refined, as in the iterative methods. As indicated by the narrow distribution of errors in Table 4.2 and Figure 4.5, the iterative Gaussian algorithm performed slightly better than the iterative Sigmoid model, perhaps due do the tapered shape of the Gaussian. The Sigmoid marker signature tracked in plain-film RSA images is not the same as the noisier fluoroscopic counterpart. These results suggest a hierarchical approach to marker tracking on fluoroscopic images. Initial guesses may be provided via manual user input, image segmentation, or feature detectors or seeded using the intensity-based center of gravity (4.2.1). This algorithm yielded similar robustness to initial guess errors as the iterative Sigmoidal model (4.2.4) but at a substantially cheaper computational cost amenable to the practical task of making many initial guesses over a trial of data. Following the initial guess, the refined localization procedure should use the iterative Gaussian approach (4.2.3) that minimal inter-marker errors and robustness to fluoroscopic noise occurring across frames of image data.

Multiple limitations to this series of experiments limit extrapolation of the results beyond the parameters and equipment tested in this study. Marker beads smaller than 2 or larger than 3 millimeters were not tested, but it is assumed the number of available pixels in the radiographic marker shadow and the amount of signal attenuation (a function of bead thickness) that creates marker contrast would influence localization results. Markers were tracked in a simple acrylic lattice that provided ideal conditions not observed during in vivo imaging. Acrylic attenuates X-rays to a similar degree as pure water. The additional contrast loss due to soft tissue attenuation or scatter would certainly increase localization errors. The lattice was only imaged flat to the image intensifier face to minimize the effects of object magnification, scatter artifacts, and the projection-based biases in the marker centroids. However, a better metric for total distortion correction performance of the full biplanar system would be to assess the propagation of image intensifier plane errors to full 3D localization errors in the typical capture volume. The full 3D

reconstruction accuracy of a given marker from the stereo images is a function of the camera calibration produces, models, and noise levels which will be discussed in Chapter 5. Still, these data provide essential feedback on the best algorithms to be used for calibration and marker-based bone tracking, and they also establish a noise floor that is useful for designing future studies that may use radiographic marker displacements as means of strain measurement [41, 42]. The stability, accuracy, and robustness of marker localization in the presence of fluoroscopic image noise is an essential component to the overall virtual biplane software environment. By having an arsenal of tools available for the localization task and understanding their performance tradeoffs, CLiMB researchers can be more confident in the biplane data they generate and process.

Chapter 5. CAMERA CALIBRATION

5.1 OVERVIEW

A *camera* projects objects in the 3D world coordinate frame ($\mathbf{X}_{\text{world}}$) onto a two-dimensional (2D) image frame ($\mathbf{X}_{\text{image}}$) (Figure 5.1). A mathematical mapping between the world and image frames is given by the 3x4 camera projection matrix, $\mathbf{P}_{\text{camera}}$ (Equation 5.1a).

$$\left|\boldsymbol{X}_{\boldsymbol{image}}\right| \;=\; \left|\boldsymbol{P}_{\boldsymbol{camera}}\right| \quad \left|\boldsymbol{X}_{\boldsymbol{world}}\right| \tag{5.1a}$$

This matrix is calculated via a calibration process that utilizes observations of control points ($\mathbf{X}_{\text{image}}$) of a calibration object of known world-coordinate system geometry ($\mathbf{X}_{\text{world}}$). Once the two camera matrices of the biplane system are determined, bones or markers may be localized or reconstructed in lab space. This process is known as "tracking" and uses the observations of the object ($\mathbf{X}_{\text{image}}$) in conjunction with the camera matrices ($\mathbf{P}_{\text{camera}}$) to solve for the position, $\mathbf{X}_{\text{world}}$. Each of these processes is critical to generating accurate data acquired on a biplane fluoroscopic system. Unsurprisingly, both processes are noise-corrupted as all three components of the camera equation (5.1a) contain errors due to things like X-ray physics, optical aberrations, and electronic read-out and quantization noise.

A "pinhole" camera model has traditionally been used to model fluoroscopic imaging chains [80, 89]. Pinhole camera models are simple projection models that do not account for lens distortions, blurring, or other optical aberrations that affect the image formation process. Beyond a simple projection, additional parameters for modeling radial or tangential lens distortions have also been described for visible light photogrammetry [90, 91], although Mery found such models to be insufficient for counteracting X-ray imaging chain distortions [82, 86]. For this reason, these optical and electromagnetic image distortions are accounted for using the methods outlined in Chapter 3 before calculating this simpler pinhole projection matrix. It is assumed that the image pre-processing steps outlined in the previous chapters like distortion correction, flatfield

correction, and deconvolution have been applied to remove those degradations and produce the best approximation to the true projection image.

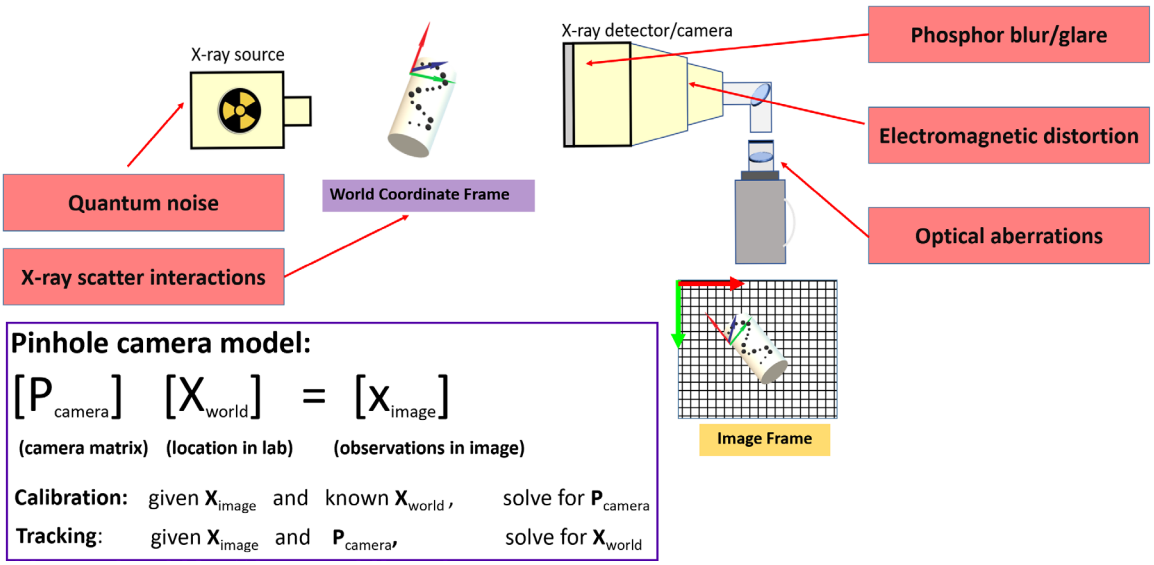

Figure 5.1. Schematic of pinhole camera model of a single arm of the biplane imaging system and the various error sources that challenge the simplicity of that model. The procedure of calibration uses observations of known calibration object to produce a camera matrix. That matrix is used in with image observations to reconstruct 3D positions in the lab.

The 3x4 pinhole camera projection matrix provides the mapping between the image and world, but the elements of this matrix are not the camera parameters. Rather, this matrix may be decomposed into a pair of matrices, **K** and **E,** that represent the physically-meaningful intrinsic and extrinsic camera parameters, respectively (Equation 5.1b) [92]:

$$|\boldsymbol{X_{image}}| = |\mathbf{K}| \quad |\mathbf{E}| \quad |\boldsymbol{X_{world}}| \tag{5.1b}$$

As shown in full form by Equation 5.1c, a world point, $\mathbf{X}_{\text{world}}$, expressed as a 4x1 vector in homogenous coordinates, is mapped using the extrinsic matrix, **E**, and intrinsic camera parameter matrix, **K**, as a homogenous (3x1) vector of the resulting image pixel coordinates, $\mathbf{X}_{\text{image}}$ [92]. The three primary intrinsic camera parameters are the camera lens focal length (***f***)

and the center of the optical axis relative to the image sensor ($\mathbf{c}_u$, $\mathbf{c}_v$). The 3x4 extrinsic camera matrix (**E**) is the row-concatenation of a 3x3 rotation matrix of direction cosines (**R**) and a 3x1 translation column vector (**t**) that, taken together, describe the pose of the camera in the world (lab) coordinate system. The full mathematical camera projection model used to map points between the fluoroscope images and their lab location is:

$$\begin{vmatrix} u \\ v \\ 1 \end{vmatrix} = \begin{vmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{vmatrix} \begin{vmatrix} X \\ Y \\ Z \\ 1 \end{vmatrix} \tag{5.1c}$$

Camera *calibration* is the process of determining the elements that form the matrix $\mathbf{P}_{\text{camera}}$. Images of a calibration object of known geometry are corrected of distortions using the methods described in Chapter 3.5, and the observed X-ray projections of the control points are localized on the images (Chapter 4), yielding observations, $\mathbf{X}_{\text{image,}}$ on the left side of Equation 5.1c. The locations of the control points in the world coordinate system (embedded in the calibration object) are derived using high-resolution computed tomography (CT) volumetric scans and provide the known values $\mathbf{X}_{\text{world}}$ on the right side of Equation 5.1c. In the calibration object, if the fiducial markers are spherical and of adequate size, their 3D centers may be calculated with submillimeter accuracy using intensity-weighted centroiding algorithms.

Traditionally, biplane X-ray camera systems have been calibrated using the direct linear transformation (DLT) methods [18, 19, 39]. DLT has a long history in biomechanical studies using both optical motion capture and static radiographic technologies to quantify various human kinematics [57, 60, 92, 93], so its application to biplanar imaging was a natural evolution. DLT generates the camera matrix relating the world coordinates of the calibration object's markers to their image coordinates via pinhole projection in an algebraic total least-squares sense. The system of DLT equations is solved for parameters that minimize the algebraic residual error without regard to any physical camera constraints or parameters. Direct linear transformation algorithms are easy to implement with modern linear algebra solvers and computer vision software packages. The DLT produces the 3x4 projection matrix ($\mathbf{P}_{\text{camera}}$) with 11 independent parameters (the last matrix element is unity). This matrix is decomposed using Melen's method [94] into the camera extrinsic and intrinsic matrices required to fully describe the virtual camera.

When these camera parameters are known, the X-ray projection process can be simulated for marker points or bones using the DRR algorithm.

While practical and straightforward to implement, the DLT method has some drawbacks that contribute to measurement noise and fundamentally limit the overall metric accuracy of the biplanar imaging system. The DLT algorithm has been shown to be sensitive to extrapolation errors of the calibration in regions of the capture volume that are outside of the calibration object control points' convex hull [57, 58, 95]. Essentially, this means a calibration object, or the samples of its calibration points must be spatially distributed across the entire capture volume. If high accuracy across the entire volume is required, this necessitates a calibration object that is either large enough to span the space while being dimensionally stable, or able to be spatially manipulated throughout the imaging volume. The DLT method is also sensitive to noise in the input data [57-60, 96]. This noise may exist in the image observation locations because of residual errors in the distortion correction or errors in the localization of the marker signature itself. In addition to noisy image coordinates, another source of error comes from the calibration object's fiducial marker positions. These positions may be uncertain due to localization errors or thermal expansion of the calibration object material [82]. Weighting the centroid input data of the DLT method based on reliability or variance of the localizations can improve fitting, but a global minimum of system error in the presence of outliers may not be possible using this algebraic method. Anecdotally, this technique for generating camera matrices has proven unstable in the current implementation of the CLiMB biplane system, causing large variations in the extrinsic camera parameters and focal lengths.

To counter these issues, global optimization methods such as bundle adjustment refinement of camera parameters have been applied to optical and X-ray imaging systems with great effect [82, 83]. These methods utilize a nonlinear method, such as Levenberg–Marquardt, that is driven by a cost function of the total residual 3D reconstruction error of the calibration control points. They also have the advantage of utilizing multiple observations of the control object placed in different spatial locations, to produce a globally optimum calibration, and are robust to frames of missing marker data. The fundamental strength in this method, however, is the focus on optimizing meaningful camera parameters rather than just a linear algebra matrix. By decomposing the X-

ray imaging chain into its core physical camera components, each may be expressed in the optimization function with appropriate weights and constraints. For example, the X-ray source-to-detector distances may be measured a priori in the lab and used to constrain the focal length estimates. This offers increased flexibility over the DLT-based method, which obfuscates these components into a single 12-parameter camera matrix that must be decomposed into the meaningful camera parameters via Melen's decomposition technique [94].

Lastly, by minimizing the geometric error of the system rather than the algebraic error, bundle adjustment should yield camera parameters that produce more uniform error residuals throughout the capture volume. Hartley and Zisserman illustrate the coupling of focal length and reprojection error that exists in the DLT algorithm ([92], see Chapter 7.2) from the Law of Triangles. The effect is a spatial pattern of increased reconstruction error as a function of the distance away from the camera, along the path of the focal length. This is unacceptable for studies requiring the tracking of small bones, like a toe, that may pass through the entire capture volume during a walking trial, and corrupts validation studies relying on inter-marker distances as a metric. Despite these shortcomings, the DLT is used to provide an initial guess of the camera matrix before bundle adjustment, and, as with any optimization scheme, good initial guesses of parameters are essential to converging to a true global optimum. Therefore, while the final reconstructive error of the photogrammetric system is minimized with the latest in nonlinear optimization algorithms, an understanding of the behavior of the DLT model is important to CLiMB for improving the accuracy of the system and establishing baseline limitations.

This chapter explores and quantifies the sources and propagations of error in the lab's current pinhole camera model for reconstructing the spatial location of objects in the virtual biplane software. The following sensitivity analyses are motivated by CLiMB's dissatisfaction with the current calibration pipeline and are parameterized using the statistical variances in the imaging data experimentally determined from the previous chapters. The insights gained from these simulations informed a new calibration object design and calibration procedures, which are outlined in the following sections, and will aid in the design of future research studies.

5.2 SENSITIVITY ANALYSIS

Studying the camera projection equations (5.1a-c) will reveal that errors in the localization of the control points, both in the fluoroscopic images ($\mathbf{X}_{image}$) and the 3D world coordinate system ($\mathbf{X}_{world}$) contribute to errors in the camera matrix. No camera model can perfectly model all aspects of a real X-ray imaging chain. However, for any model, there should exist a set of parameters that provide the best approximation to the real camera system. Deviations in the virtual biplane model from the true camera parameters used at the time of fluoroscopic image acquisition will decrease system accuracy. Errors in the camera model can be quantified using metrics like reprojection error, which is the distance between a world point's projected location onto the image plane using the camera matrix and the localized observation of the corresponding point in the image [92]. Another metric, reconstruction error, is the 3D distance between a known world point and its estimated (reconstructed) location using the corresponding noisy observations in both imperfectly modeled biplane cameras. These metrics will be used to quantify the model's performance in response to different noise levels. The propagation and the extent of these errors in the virtual biplane camera chain is the focus of this analysis.

A multitiered sensitivity analysis of the pinhole camera model for a typical biplane foot and ankle imaging configuration was performed to understand these physical measurement limitations imposed by the hardware and software. The first tier investigated the effects of the centroid localization errors (Chapter 4) of the calibration object on camera parameters derived using the DLT method. This investigation will highlight the limitations of the current calibration object and the DLT algorithm pipeline. The second analysis quantified the uncertainty of reconstructing 3D points as a function of the noise in the camera model and the points' image projections. This analysis will provide a lower bound on the 3D metric reconstruction accuracy possible with the current system. All analyses were conducted in the Dakota software package [97] from Sandia National Labs using biplane imaging geometry that is typical for the foot and ankle gait studies of CLiMB. The current CLiMB calibration object was virtually placed at the center of the capture volume, and the 15 marker coordinates projected onto the idealized image planes as the observations by each camera. This synthesized ideal test projection scenario omitted confounders

like optical distortions, X-ray scatter, and sensor readout noise to focus on the effects of the localization errors.

5.2.1 *Sensitivity of camera matrices due to localization noise*

5.2.1.1 Methods

A centered-parameter sensitivity analysis was performed to quantify the effects of errors localizing the calibration object markers on the resulting virtual camera models via Monte Carlo sampling. The localization error of the markers was simulated by adding increasing amounts of bi-directional spatial perturbation to the initial idealized marker coordinates. For each camera, 5000 Monte Carlo perturbation samples of the marker image coordinates were generated, moving one of the calibration markers at a time. The uniform perturbation distributions were generated using Latin Hypercube Sampling and ranged between 0.01 pixels and 0.25 pixels. These ranges represent some of the best and worst-performing localization algorithms (Table 4.1 in Chapter 4). Errors at or above the sensor spatial Nyquist frequency (0.5 pixels) are assumed to cause unusable camera models and are not considered. For each of the extrinsic (6) and intrinsic camera (3) parameters, the univariate effects of the localization error magnitude were determined in Dakota. Additionally, Sobol indices were calculated in Dakota using polynomial chaos expansion approximation to rank how susceptible each camera parameter is to variance in the model input (localization error) [97, 98].

5.2.1.2 Results

The reconstruction of the cameras with different sets of noisy marker pixel coordinates produced spatially perturbed cameras as expected (Figure 5.2). This manifested as highly variable extrinsic parameters describing the camera's position and orientation in lab space, and the intrinsic parameter of focal length. Focal length estimates varied around 100 millimeters in response to the subpixel localization errors (Figure 5.3, left). Considerable variation in the estimation of the two components of the image center was also observed (Figure 5.3, center and right). Large jumps in extrinsic parameters were also observed (Figure 5.4). Camera orientation, described by the three Euler angles, was less sensitive to localization error, with all deviations not exceeding 3

degrees (Figure 5.4, bottom). However, the estimated camera position in lab space (Figure 5.4, top), like focal length, saw large jumps to compensate for the noisy marker positions. This led to shifts in the virtual X-ray camera positions on the order of centimeters. Reprojection errors of the markers (Figure 5.5) increased nonlinearly in response to the perturbation magnitude, reaching a maximum of about 0.12 pixels across all perturbations. The Sobol indices (Table 5.1) rank the contributions of the various camera parameters in terms of their first-order and the global effects that consider interactions with other parameters. The extrinsic parameters dominate this camera model, with localization variance explaining nearly half of the variance in camera position and a significant portion of the orientation. The parameter with the smallest influence (1.4%, overall) in explaining model output variance was the camera focal length (Table 5.1).

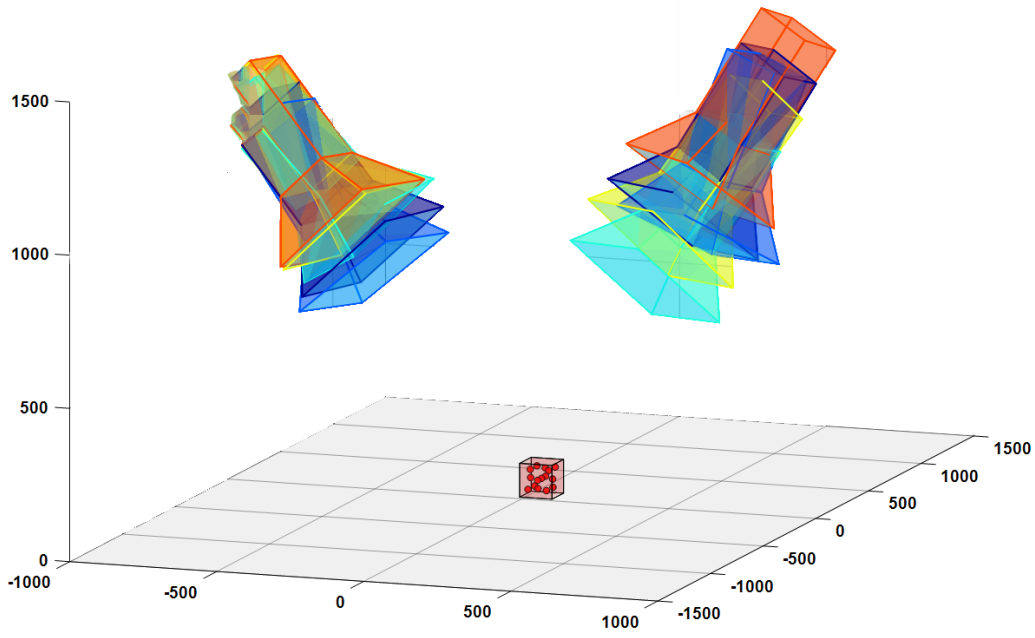

Figure 5.2 Example of five pairs of biplane cameras generated from a single set of noisy fluoroscope calibration images. Errors in the marker localization produce inconsistent camera parameters that undoubtedly affect the projection matrices and, hence, the reconstruction accuracy of the system.

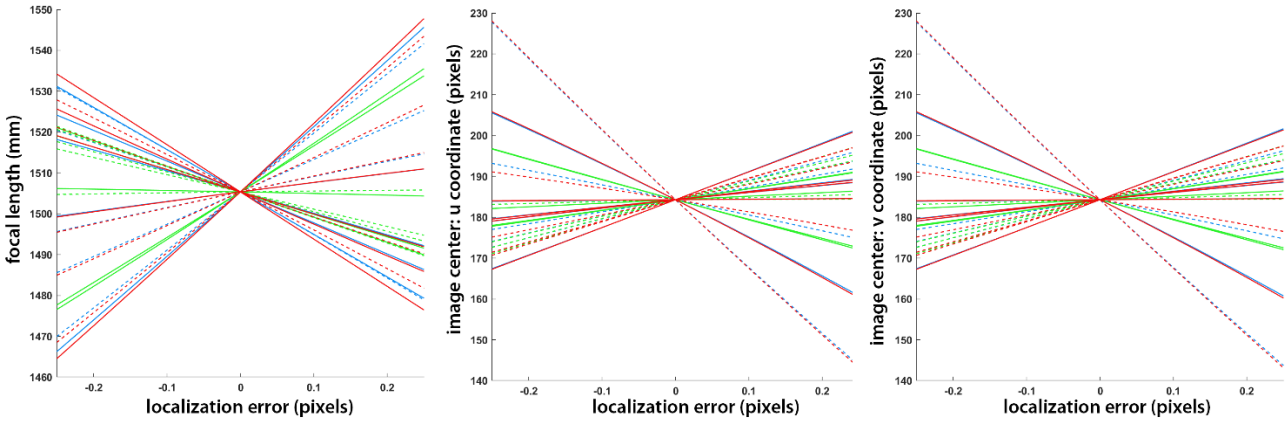

Figure 5.3 Sensitivity results of the camera intrinsic parameters (focal length, image center ***u*** coordinate, and image center ***v*** coordinate) in response to localization errors in the calibration object. Each line trace is a marker in the calibration object. The horizontal axes are the perturbation magnitudes.

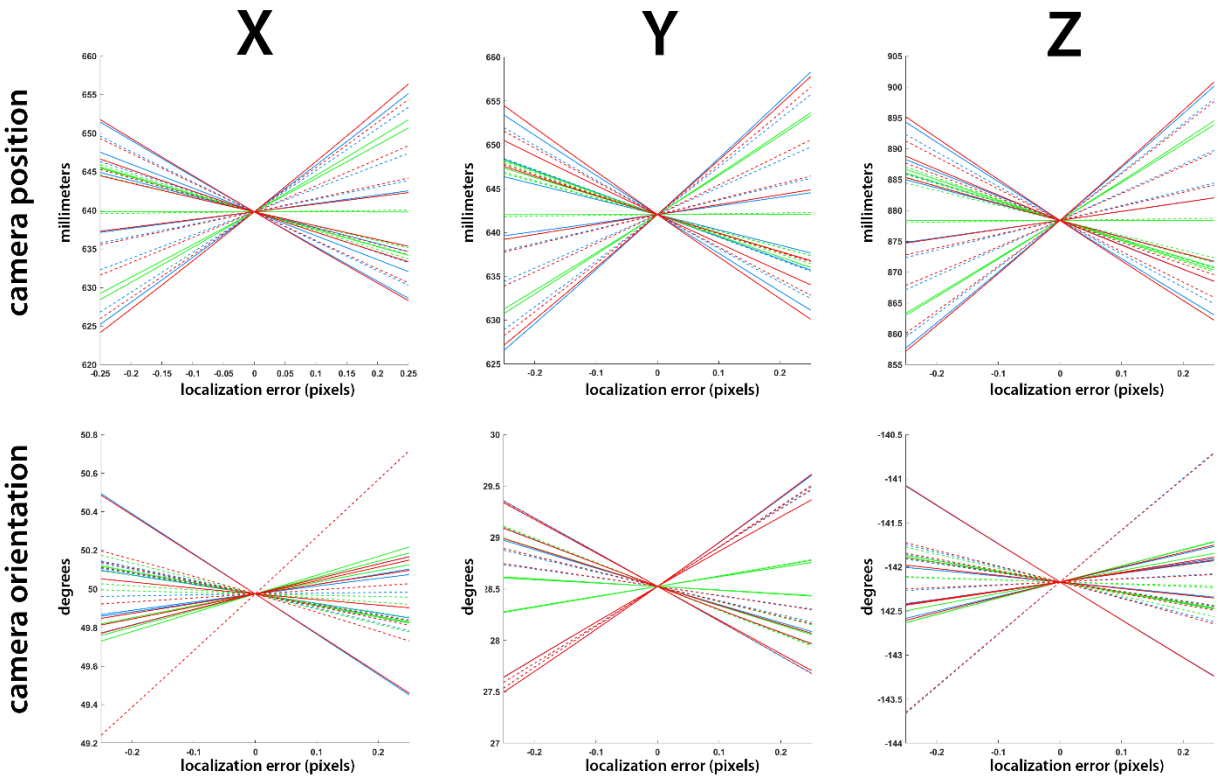

Figure 5.4 Sensitivity results of the camera extrinsic parameters (three positions, three orientations) in response to localization errors in the calibration object. Each line trace is a perturbed marker in the calibration object. The horizontal axes are the perturbation magnitudes.

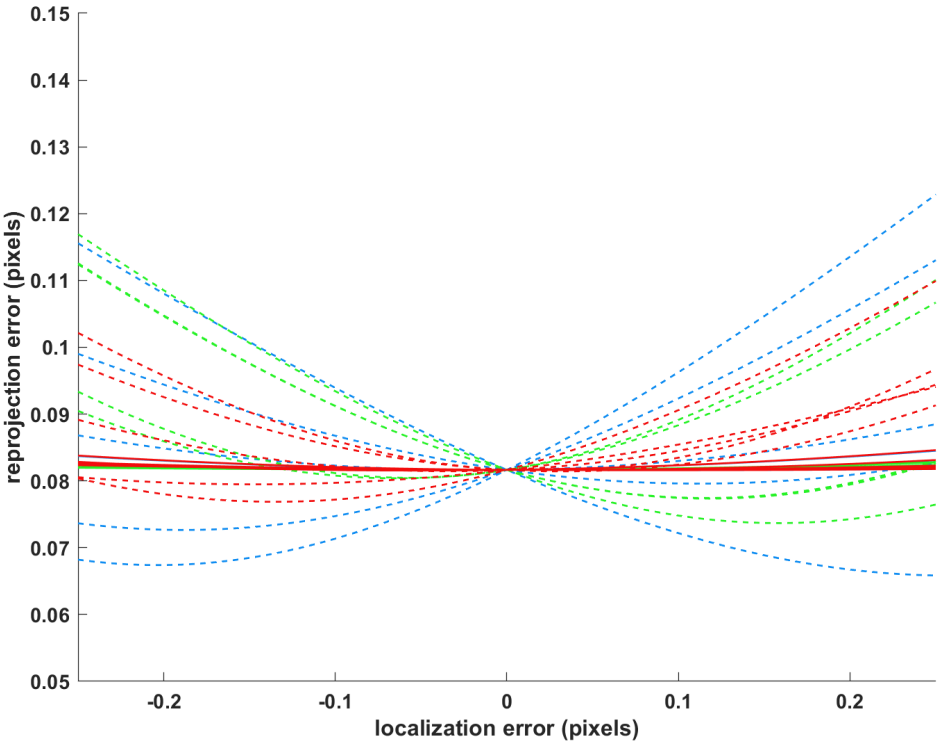

Figure 5.5 Sensitivity results of the camera reprojection error in response to localization errors in the calibration object. Each line trace is a perturbed marker in the calibration object. The horizontal axes are the perturbation magnitudes.

Table 5.1 First-order and total-effects (Sobol indices) of the camera parameters.

	Camera parameter	First-order	Total-effect
extrinsic	**Position: X**	22.3%	49.8%
	Position: Y	11.0%	41.5%
	Position: Z	10.4%	44.4%
	Rotation: X	4.6%	11.5%
	Rotation: Y	4.7%	15.5%
	Rotation: Z	1.8%	11.3%
intrinsic	**image center: u**	0.3%	7.8%
	image center: v	4.1%	8.8%
	focal length	1.8%	1.4%

5.2.1.3 Discussion

As mentioned earlier, the focal length in the camera model is equivalent to the X-ray source-to-detector distance and is essential to generating accurate DRR images with this projection geometry. This parameter is also minimized as an artifact of the DLT algorithm [92]. The preferential treatment in the DLT is reflected in the lowest ranking of the Sobol indices amongst the other camera parameters. However, the significant variations in the focal length, on the order of centimeters, from small, expected magnitudes of marker noise. These results coupled with the highly sensitive extrinsic parameter of camera position noise lead to grossly different source-to-detector geometries. This can lead to suboptimal virtual biplane projection models that will produce DRRs that cannot match the acquired fluoroscopic images, and thus cannot solve the localization problem. The camera orientations were affected less by marker noise than camera positions, most likely because the perturbations were not large enough relative to the long focal lengths used in CLiMB's biplane model. In effect, the subpixel perturbations were not enough to turn the cameras to "look" in another direction. While reprojection error did increase with more perturbation of the marker coordinates, the inflated errors and outliers of the true camera parameters like focal like and position are not reflected in this slowly increasing metric. This suggests reprojection errors cannot be solely relied upon to judge the adequacy of a camera model or parameter fit. Lastly, the sensitivity of the extrinsic parameters will be exacerbated in a biplanar system utilizing two noisy camera matrices. Despite truly being a rigid stereo camera array, the cameras appear to move grossly in the lab space relative to each other depending on the noise in the marker observations. As a result, each separately imperfect camera will contribute to the overall error in the system in reconstructing 3D points or objects. This end-effect is investigated in the next section.

5.2.2 *Propogation of localization uncertainty into 3D reconstruction errors*

Having characterized the sensitivity of the pinhole camera model, as applied in our lab, in response to localization errors, attention is turned towards the errors in the end-product: the 3D reconstruction of a point in lab space. Ideally, the biplane system would have minimal and uniform reconstruction errors through the entire capture volume. This is not the case in reality but is usually ignored. During gait, the foot and ankle bones will move throughout the entire capture volume during a single biplane acquisition trial. It is important to have as close to

uniform reconstruction accuracy as possible, or to at least characterize and understand spatial patterns and fundamental limits to these errors.

5.2.2.1 Methods

For these simulations, the ideal biplane model mimicking the setup for gait acquisition was created again. A lattice of test points was virtually generated, spaced 1 centimeter apart in the intersecting volume of the virtual biplane model. The initial idealized camera matrices and marker projections were again corrupted with increasing levels of marker noise based on the variances observed in the experiments of Chapter 4. For each iteration, the lattice test points were projected onto the image planes defined by the noisy camera matrices and subsequently reconstructed back into a 3D point. The reconstruction errors are the distances between the original lattice point and its calculated position. At each lattice point location, the mean and maximum reconstructed error over all iterations of noisy camera matrices was calculated.

5.2.2.2 Results

Reconstruction error varied as a function of spatial location in the biplane capture volume. Looking at the pattern of error (Figure 5.6), mean reconstruction errors across all perturbations tended to increase as distance away from the X-ray source increased (increasing focal length). This is indicated by the hotspots at the bottom of the cone intersections, and the relatively lower error (lighter color in Figure 5.6) at the top of the intersecting volume. Localization errors (units of pixels) propagated into 3D reconstruction error (units of millimeters, Figure 5.7). Even small localization errors on the order of 0.05 pixels, as we might expect from the best of the localizer algorithms described in Chapter 4, yield nontrivial mean and maximum reconstruction errors, although they do not exceed 1.0 mm (Figure 5.7, top). Larger localization errors of 0.25 and 0.50 pixels easily yield errors in 3D reconstruction above a millimeter.

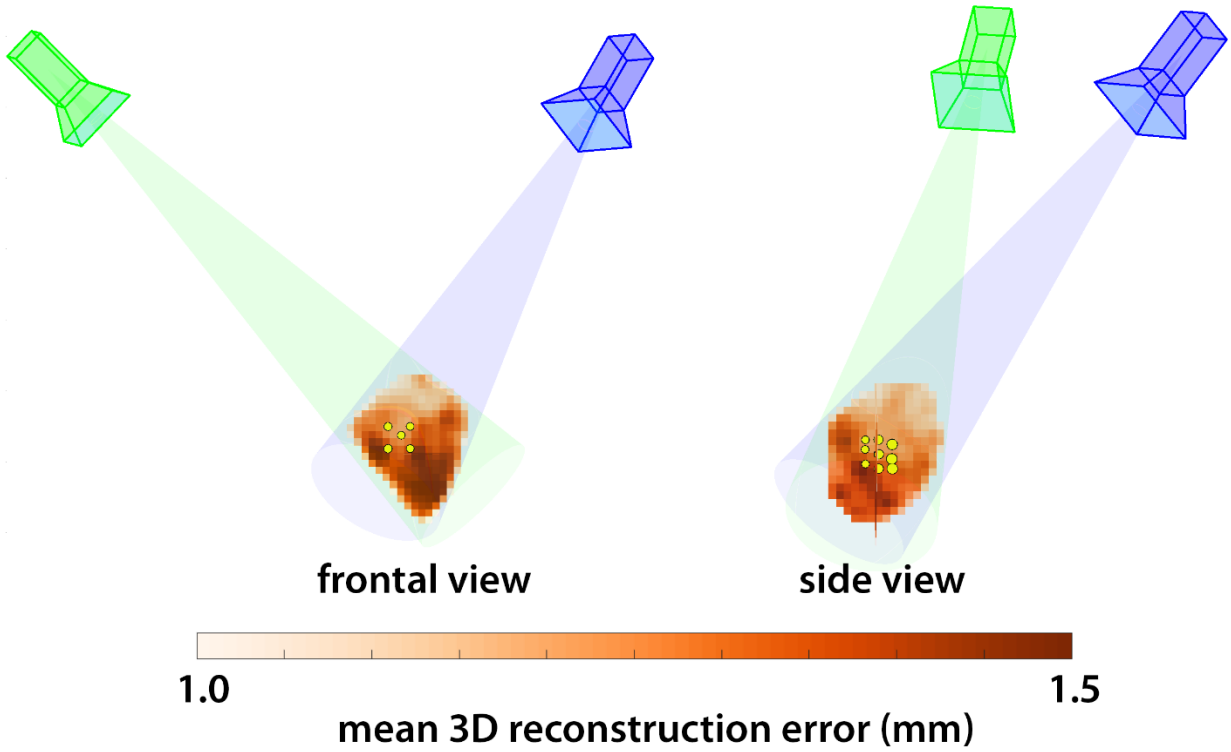

Figure 5.6 The non-uniform spatial distribution of 3D reconstruction errors from the uncertainty quantification analysis of the effects of localization error. The two intersecting cones of the biplane imaging system form the capture volume. The calibration object position is shown with yellow markers. Errors are largest at the bottom of the volume, the points farthest from both X-ray sources.

5.2.2.3 Discussion

This uncertainty analysis quantified the propagation of marker localization errors for CLiMB's standard foot and ankle imaging setup. The outputs of this model indicate significant spatial correlations in the reconstructive accuracy of the current biplane model that result in "dead zones" of increasing error that should be rectified with a bundle adjustment procedure or avoided if metric accuracy is important. The importance of subpixel localization algorithms explored in Chapter 4 is justified by the large shifts in possible 3D accuracy even between the 0.05- and 0.10-pixel error cases (Figure 5.7). Considering that joint spaces in the foot and ankle are typically less than 1.0 millimeters, this stresses the critical importance of these camera models and localization algorithms in producing biomechanical data that can be trusted with confidence.

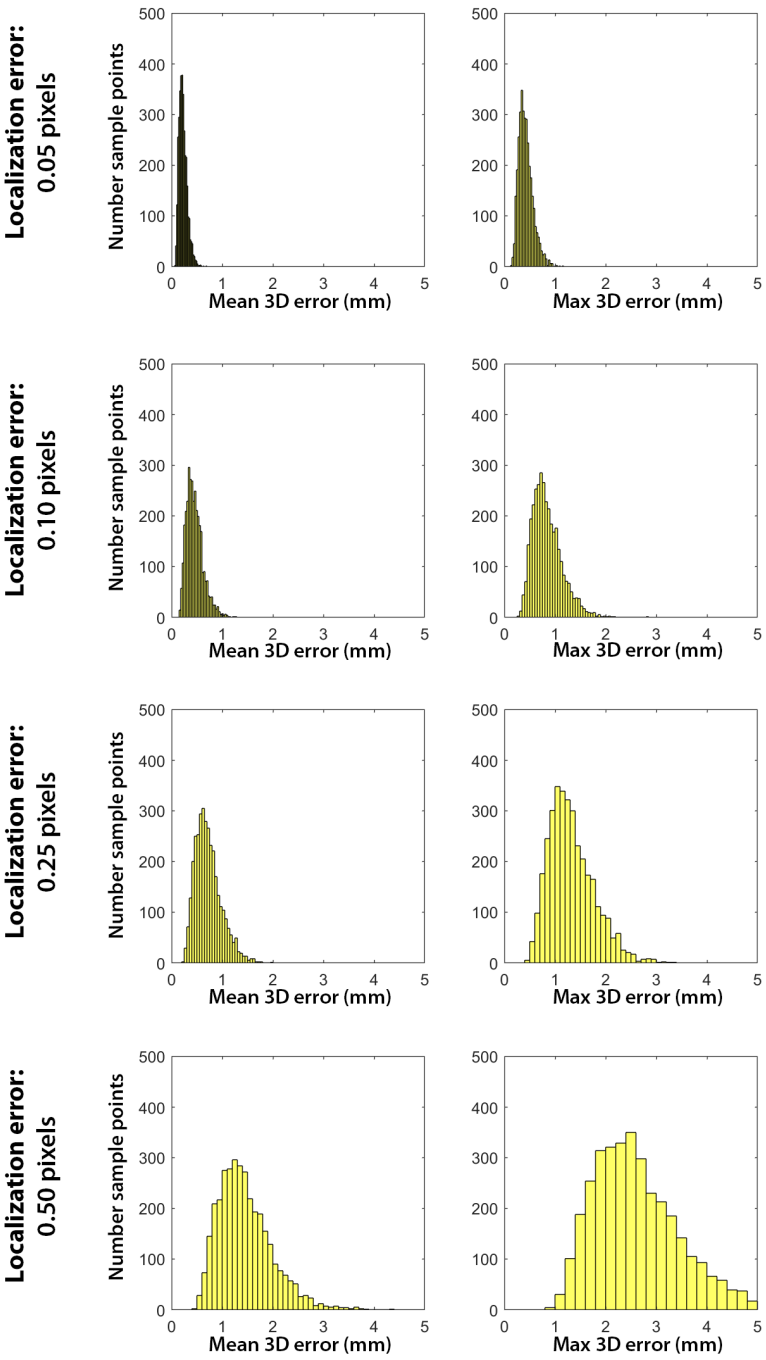

Figure 5.7 Histograms of the mean and maximum 3D reconstruction errors of the lattice test points as a function of various localization error levels.

5.3 CALIBRATION OBJECT

A photogrammetric camera calibration object should consist of reliably identifiable landmarks, with known and stable geometry. These marker centroids should be easy to localize on noisy fluoroscope images with a high level of accuracy to reduce calibration errors, and the correspondences of marker observations between the two fluoroscope views must be determined before the direct linear transformation calculation. Ideally, the calibration object should span the entire intersecting imaging volumes of the biplanar system or be able to undergo spatial manipulation to produce multiple samples of calibration points spread across the lab space. As shown in the previous section, the key to accurate and stable camera matrices is an adequate number of precisely localized fiducial points that can be correctly identified in both simultaneous camera views.

5.3.1 *Issues with the current calibration protocol*

The present calibration object (Figure 5.8, top left) is a solid cube with 15 spherical markers arranged in three offset planar groups of differing marker diameters. Each group of five 1.78, 3.17, or 4.76 mm diameter markers beads were embedded into the 100mm block of R1/BB Butter-Board resin (Golden West Manufacturing, USA) by drilling out 8-, 13- or 18-mm diameter cores to install and localize the marker beads with a probe. The current calibration procedure automatically tries to detect the marker signatures against the image background of the cube material using a template matching routine. The user then manually orients and matches a 3D virtual model of the cube to match the two fluoroscope views to determine the marker correspondences. The projective nature of X-ray images combined this particular marker geometry (three planes) make for a challenging and time-consuming process to obtain a good camera pose estimation without a lot of trial and error since the user cannot rely on typical visual cues, like surface light reflections, utilized by the human eye. Substantial time saving is possible by automating the tasks of calibration marker detection, localization, and labeling. While these are software-based tasks, they are drastically simplified by a discerning design of the calibration object itself.

As quantified by the sensitivity analyses, the algorithms for marker localization and camera matrix determination are also sensitive to errors in marker centroid coordinates. The primary issue with the current calibration cube object is the use of small markers with poor contrast on noisy images (see the lower middle portion of Figure 5.8 for an example). Anecdotally, often two or three markers cannot be reliably localized in the cube image due to the lack of contrast to the radiodense cube medium. This reduces the total number of markers available from both views to solve for the eleven DLT parameters. While the bare minimum number of required markers is six, it has been suggested that the number should be three times the number of parameters in the DLT model [77, 78]. To be explicit, this means a calibration object with two to three dozen separate markers. Substantial improvements are possible with calibration markers that yield a cleaner X-ray signature for localization and correspondence labeling.

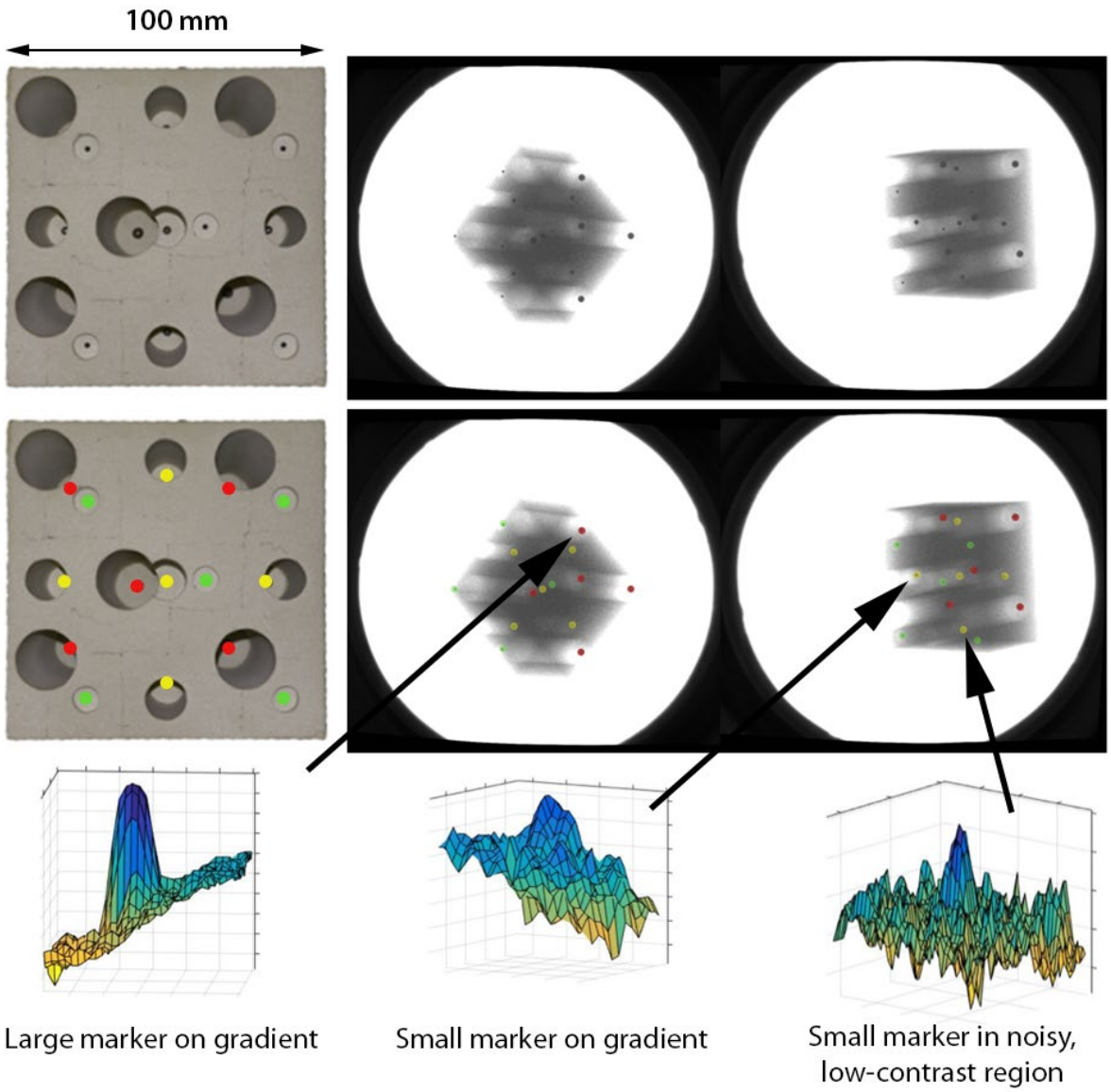

Figure 5.8 Current calibration cube (top left) and example fluoroscopy images (top middle and right). Sample marker signatures are shown in the bottom row from a typical calibration image. Except for the largest five markers, most markers are difficult to reliably localize given the background noise and lack of contrast. Identification of each marker in the fluoroscopic images (middle row) is a nontrivial task that requires estimating the orientation of the cube in both views.

5.3.2 *New calibration object design*

To overcome the limitations imposed by the existing cube, a new calibration object for the biplane system was designed and fabricated. Learning from these previous experiences, key design attributes for the new object included: high marker contrast relative to the background medium, automatic marker segmentation and labeling, a larger calibration region, and automatic pose estimation of the object. The resulting design borrowed heavily from the work of Rougee, et al. [99, 100], who described a helical calibration object for fluoroscope calibration. To accommodate future studies requiring the operation of the X-ray image intensifiers in a variety of magnification modes, two spiral calibration objects were desired; each optimized for the smallest and largest possible biplane capture volumes as determined by the image intensifier magnification levels. Helical patterns of varying bead density and spirality were simulated in DRRACO software at typical biplane imaging setups to simulate marker occlusion situations that would hinder localization and labeling in the software.

The final designs consisted of 23 and 33 stainless steel beads in a unique, alternating pattern of 3.0 mm and 5.0 mm diameters embedded in castable polyurethane foam (Master Dyna-Cast Fast-Cast Urethane, Freeman Manufacturing and Supply Co, USA) . Multiple urethane compounds of varying densities from the manufacturer were compared to the Butterboard material to find a compromise of dimensional stability to thermal expansions and radiographic contrast provided to the markers. Testing included preparing and embedding 3.0 mm stainless steel markers into three 3x10x100 mm samples of each material type. Intermarker distances were quantified using fluoroscopy after multiple freeze-thaw cycles of the sample blocks. Relative marker intensity contrasts to the background resin material were also calculated from the fluoroscopy images. To protect the calibration object, this resin core was press-fit into a polycarbonate tube. Machined fasteners embedded in the polycarbonate endplates allowed the object to be externally manipulated at a safe distance from the X-ray source. Photographs of the final spiral calibration objects are shown next to the original calibration cube (Figure 5.9). To obtain the nominal locations of the markers in the objects, a high-resolution CT scans were performed for each object on a North Star Imaging NSI X5000 scanning system with a resolution of 0.145 mm per voxel. Marker centroids were localized using a 3D intensity-weighted centroiding method in MATLAB and expressed in a local coordinate system. The natural cylindrical nature of the spiral

lends itself well to defining this local coordinate frame orientation from the long axis. A second vector is defined as the shortest distance between any marker centroid and the long axis. A cross-product generates the final orthogonal axis, and the origin may be defined at either a marker centroid or a weighted location along the axis or between a subset of markers. This flexibility allows for the spiral shape to be utilized from multiple biplane imaging angles, and the geometric information from the projections of the spiral may be used to determine camera orientation relative to the coordinate frame embedded in the object.

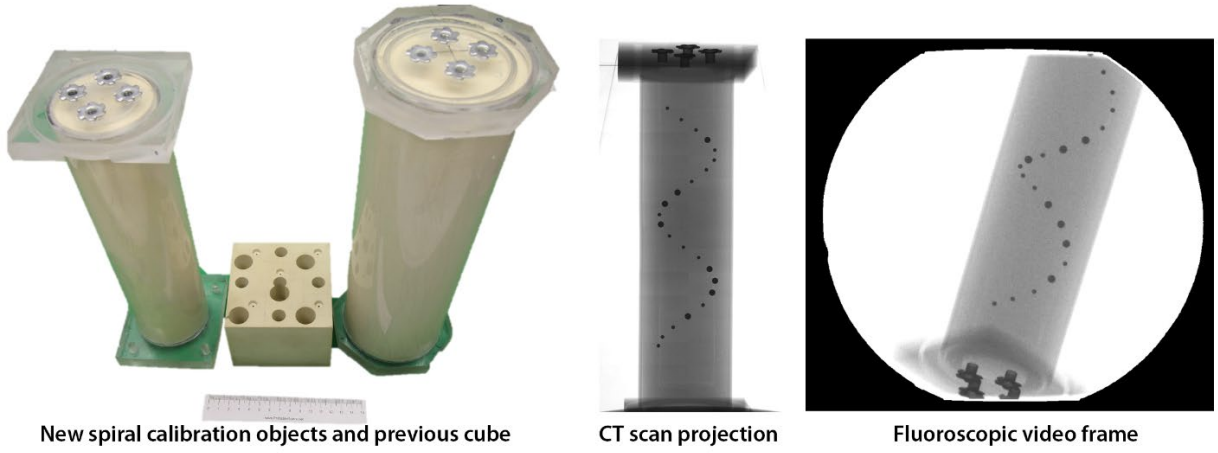

Figure 5.9 The new spiral calibration objects are shown next to the previous calibration cube for scale (left). The spiral objects have markers of two different diameters with adequate contrast to the background. Markers are distributed in a unique pattern that can be easily detected from the radiographic projections. The endplates include hardware that allows the spiral objects to be manipulated by a human or mechanical drive system through the biplane capture volume.

5.4 CALIBRATION ALGORITHM

The updated calibration objects allow for new calibration procedures that can utilize multiple frames of input data and a non-linear optimization routine for generating camera parameters that produce the smallest 3D reconstruction errors for the virtual biplane model. The following is an overview of that proposed framework.

5.4.1 *Image acquisition*

Most simply, the spiral objects may be used in a static calibration routine that uses a single image frame with the markers spanning as much of both views as possible and processed with the traditional DLT algorithm to obtain the camera parameters. A dynamic calibration is also possible where the spiral calibration object is moved through the intersecting capture volumes of the biplane system. The motion of the object can be prescribed with a motorized system and if recorded by encoders, the position and orientation of the spiral object can be known for each image pair. This information can, in turn, be used to generate initial guesses for camera extrinsic parameters [82] as part of a larger calibration optimization scheme. During this motion, the long axis of the spiral should lay as close as possible to the plane formed by the X-ray source and detector centers. Deviations from this recommendation are accommodatable but require additional manual user inputs in the subsequent processing steps. The X-ray parameters and video camera exposure parameters should be set to give the best possible marker contrast and clarity. Object motion can be limited to pure translation, but a more complete, pseudo-random sampling of the 3D lab space can be obtained by adding rotation of the spiral about the long axis. In any case, the camera frame rate and exposure times must be balanced with the object's motion to minimize marker blur.

5.4.2 *Marker localization and correspondence matching*

An intensity thresholding algorithm can be applied to the fluoroscopic images to extract the darker round regions of the marker projections (Figure 5.10). Spurious regions, such as the metallic mounting inserts, or other hardware can be eliminated by calculating the circularity and

size of each detected region. The remaining candidate regions' diameters and centroids are determined. A line passing through the cluster of centroids in each view (Figure 5.10) approximates the long axis of the spiral, and the order of the diameters projected onto the line can be used to determine the orientation of the spiral. This line can be determined via linear regression or more robustly using principal components analysis of the centroid pixel locations. Subsets of the projected marker diameter orderings can be compared like character strings to the known model of the calibration object, and these partial matches can seed initial guesses to the identity of other markers that may be overlapped or partially off-screen.

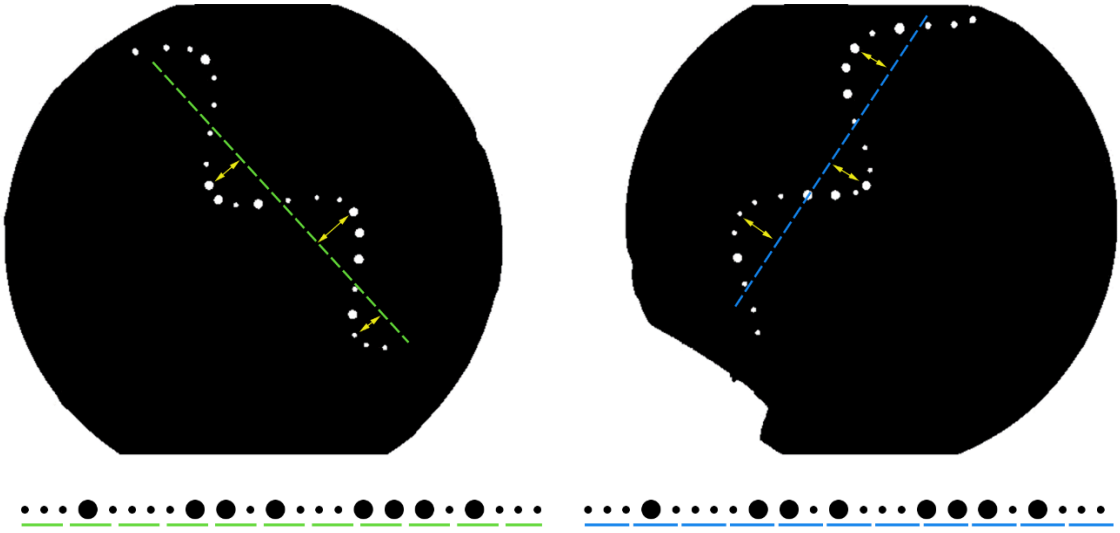

Figure 5.10 The markers of the spiral calibration object can be easily thresholded from the background of the fluoroscopic images, yielding circular regions of interest. A line (dashed) is fit to the centroids of these circular regions using linear regression or principal component analysis. The diameters of the detected regions are projected onto the line, and their ordering compared to the known object geometry to determine the orientation of the spiral. The position of the spiral may be determined with at least three of the markers.

Having extracted the marker centroids for each view, the correspondences between views must be determined. Mislabeling of the markers will cause image pixel coordinates to be input incorrectly into the system of equations for solving the camera matrices. This task was previously manually done by the user by manipulating the virtual model of the calibration cube projected onto the fluoroscopic images. Such a manual correspondence task is greatly simplified

with the unique spiral pattern that offers strings of patterns a human user could easily identify in each view. Initial guesses from the spiral-axis projection algorithm described above could also seed iterative methods for correspondence matching in stereo camera arrays, such as random sample consensus (RANSAC) [92, 101]. Match candidates may be further improved or filtered automatically based on the detected marker diameter.

5.4.3 *Camera parameter determination*

The RANSAC algorithm has been applied in iterative methods for determining the geometry of a stereo camera setup using as few as eight correspondence points [92, 102]. The stereo geometry describes the relative translation and rotation of one camera frame origin relative to the other and relates the 2D projections in the images to the corresponding 3D location. Since the biplane image system is static during image acquisition (the X-ray cameras and sources do not move relative to each other, and camera focal lengths and image centers do not change), the matrix describing this spatial relationship should be constant across all images of the calibration object. While the position and orientation of the calibration of the object may vary across frames of a dynamic sequence, the cameras themselves are still static, and either camera may be selected as the reference frame for the stereo rig. Rather than treating each camera as a separate system, calculating individual DLT matrices, and trying to rectify 3D reconstruction errors with downstream processing or data filtering, this spatial rigidity in the stereo camera rig may be used as a constraint for optimizing the camera parameters. This reduces the total number of parameters calculated at each stage of the bundle adjustment. Originally, two arms of the biplane system, each with 11 unique DLT coefficients, yield 22 total parameters to be solved and then decomposed into actual camera parameters using Melen's method. Using the new calibration object and the dynamic calibration procedure, the six intrinsic parameters of the stereo camera rig may be calculated using the seven-point algorithm [92, 102] to determine the epipolar geometry between the views. Establishing the epipolar geometry allows for finding the correspondences between views of partial or missing marker data for tasks like marker filtering, automatic labeling, and gap-filling. In the next calibration step, the lab coordinate system is temporarily moved to one of the arbitrarily selected cameras. The remaining camera's six extrinsic parameters expressing the rigid position and orientation of one camera relative to the

other may be solved for using the nonlinear optimization scheme of bundle adjustment. Finally, the relation between the master camera of the previous step and the lab coordinate system is solved for by a final optimization of those six extrinsic parameters expressing the camera with respect to the world coordinate frame. This final relative position and orientation of the rigid stereo rig may even be determined using a static pose of the calibration object that is separate and distinct from any of the dynamic calibration images. In total, the system may be calibrated, stepwise, using 18 parameters while exploiting the rigid stereo rig assumption. The improved calibration procedure offers flexibility and robustness not possible with the previous DLT implementation and will hopefully streamline data processing and reduce user-errors.

5.5 SUMMARY

The analyses in this chapter elucidated the relationships between the error sources and magnitudes explored and characterized in previous chapters and the final metric accuracy of the biplane system for reconstructing 3D points. The sensitivity study explained the suboptimal performance of the previous calibration object and, considering these lessons with the literature, provided guidelines for an improved set of calibration objects and procedures. Automating the calibration procedure in a manner that still generates reliable camera matrices is essential to removing data processing bottlenecks. Future work will include further development of the bundle adjustment optimization scheme for CLiMB and a calibration protocol that minimizes operator radiation exposure and acquisition and processing time.

Chapter 6. MODEL-BASED REGISTRATION

6.1 BACKGROUND

Model-based tracking attempts to reconstruct the 3D pose of a bone or object by utilizing computer vision algorithms to coregister simulated X-ray projections of the object to the synchronized, distortion-corrected (Chapter 3), stereo fluoroscopic image pairs acquired from the biplane system utilizing the camera projection models calculated from the calibration procedures (Chapter 5). A schematic of this process is depicted in

Figure 6.1 6.1. A computed tomography (CT) scan of the subject (bones) or object of interest is used to generate a 3D gridded scalar volume of X-ray attenuation data. Elements of the CT volume are called "voxels", from "volume elements", and are the 3D equivalent of a 2D pixel, or "picture element" in a slice of the CT volume. The scalar value of a CT voxel is represented in Hounsfield units, which provide a standardized measure of X-ray attenuation across different CT scanner makes, models and parameters. Digitally reconstructed radiographs (DRRs) are simulated X-ray projections that may be formed in a variety of ways like ray-casting, splatting, light fields, and Fourier volume slice theorem [103-112]. We choose to implement a ray-casting algorithm [18] for its accuracy, amenability to a parallel computing approach, and flexibility to include additional layers of physical complexity such as scatter modeling and beam-hardening approximations. Following this approach, DRRs are formed by virtually casting rays originating at each camera model X-ray source to the pixel centers of the corresponding detector (Chapter 5). The summation of the Hounsfield units along the rays produces an image of simulated X-ray attenuations (DRR) that is similar in appearance to a radiographic projection of the bone observed from the same viewpoint on a real X-ray camera. For a given bone, the computed tomography volume translations ($\mathbf{t}_x$,$\mathbf{t}_y$,$\mathbf{t}_z$) and orientations ($\boldsymbol{\theta}_x$, $\boldsymbol{\theta}_y$, $\boldsymbol{\theta}_z$) relative to the stereo camera geometry is manipulated, and new DRR pairs are calculated iteratively until a position of best-match is found. This should correspond to the true position of the bone in the laboratory. One or more sets of features (e.g., intensities, edges, textures) are compared between the DRR and fluoroscope images to derive a scalar score of image similarity that drives the optimization routine (Figure 6.1, bottom right). In the ideal case, when the bone model is at the "true" pose in the virtual biplane software, the DRRs projected onto the acquired fluoroscope images would match perfectly and produce a high score

of similarity. The final accuracy and repeatability of the derived bone kinematic measurements for our research studies and the processing time required for their calculation are directly related to the performance of the image similarity metric (ISM). The ISM is an essential component of the biplane bone tracking process and warrants investigation into the merits and performance of various candidate metrics for our case of foot and ankle bones.

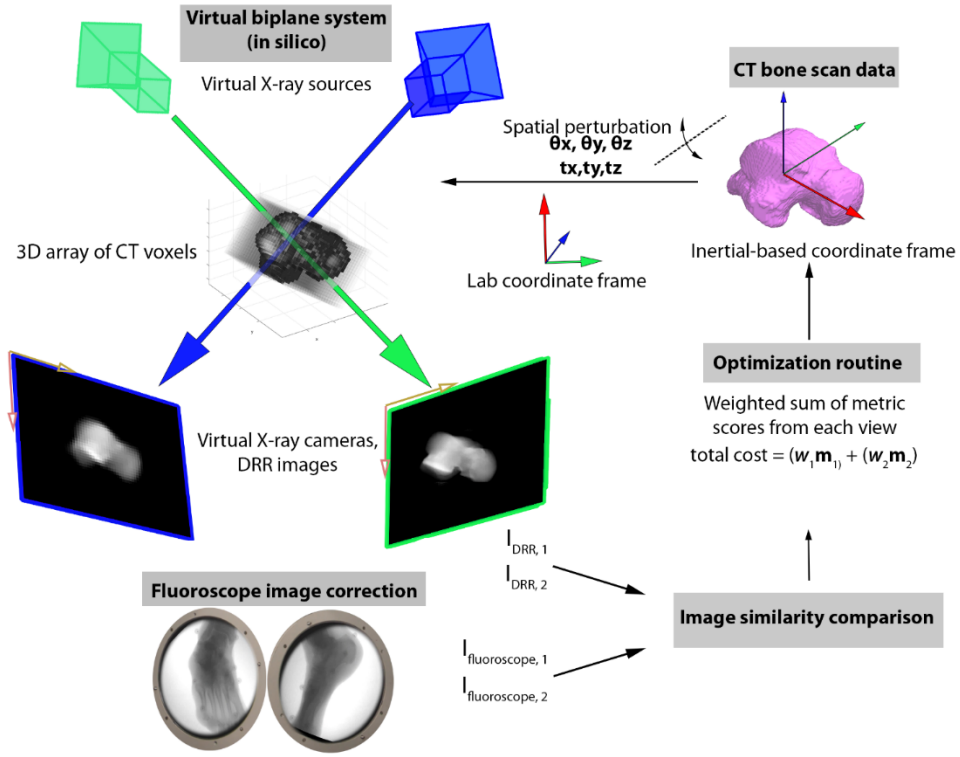

Figure 6.1 An overview of the model-based tracking process. The biplane system's cameras and X-ray sources are virtually reconstructed using the procedures of Chapter 5. Pre-processed computed tomography (CT) scan volumes of the bones of interest are imported into the virtual lab space. Digitally reconstructed radiograph (DRR) images are generated for a given bone pose and are compared to the pre-processed (Chapter 3) fluoroscope images using an image similarity measure. This generates a scalar value of image similarity that drives the optimization routine spatially perturbing the CT volume until the similarity scores are maximized for both views. The cost function may be formed as the weighted combination of multiple metrics for each view.

The following presents desirable traits of a similarity metric for bone tracking and highlights sources for differences between the fluoroscopic images and the DRR projections, and a variety

of ways for comparing these differences. Firstly, the ISM must be computable as a scalar value that is bounded or normalized. A given image, compared to an exact, spatially aligned copy of itself should have a maximum score which can be normalized to a maximum value of unity for comparison between different metric definitions and test cases. Dissimilar image pairs should tend towards scores of zero. For image pairs that fall between these two extreme conditions, the metric should vary smoothly and monotonically due to differences in image content or misalignment magnitude (Figure 6.2). Metrics that produce a flatter, non-varying (Figure 6.2, blue curve) or noisy responses with many local minima (Figure 6.2, red curve) are suboptimal for driving the coregistration optimization function.

Secondly, the metric should be as robust as possible to noise and bias artifacts in both image sets. Random high-frequency Poisson variations in the signal from various components of the fluoroscopic imaging chain affect image gray-levels directly. Low-frequency artifacts like X-ray scatter, soft tissue attenuation, and field inhomogeneity also degrade and obscure useful features like bone edges. Bone edge regions are also, by nature of attenuating more X-rays, photon-starved portions of the fluoroscopic images that are more susceptible to noise. Quantum noise and scatter artifact in the fluoroscope image and the idealized projections of the DRR data most-likely present the largest issues for an image similar metric.

In addition to noise artifacts, bias artifacts in the intensities may also be present. The DRR is derived from a CT scan, which has been segmented to extract the bones of interest. Segmentation errors, partial voxel effects, and differences in spatial resolution between the CT and fluoroscopic systems introduce aliasing in the DRRs that reduce their similarity. The simple pinhole, ray-casting DRR model does not model any of the non-finite spot size, scattering behavior, beam hardening, or polyenergetic nature of the real X-ray sources. Furthermore, CT scans are typically acquired at higher source potential voltages (ex: 120 kV) than fluoroscopic images (ex: 60 kV). Recall from Chapter 2, that the X-ray mass attenuation level of a tissue is a function of this potential energy level. This mismatch of acquisition voltages introduces a nonlinear transformation of the resulting image intensity values (Figure 6.3). Similarity metrics that assume a linear offset between datasets might require a correction of this nonlinearity. The

regions most affected by these voltage changes are the denser, cortical bone regions that provide edge content for the registration problem.

Fourthly, the similarity metric should be balanced as to not be overly prone to local minima while possessing a smooth optimizer response that is sensitive to translation and rotation of the bone from its “true” position. An operator or automated algorithm will provide the initial guess pose for a bone. Understanding the shape of this metric score-versus-perturbation curve will inform initial guess strategies and rule out ineffective image similarity metrics. These curves are generated from experimental data, described below, and are used to inform the selection of the metric(s) optimal for foot and ankle bone tracking.

Lastly, the image similarity metrics must be computationally frugal or amenable to a parallelizable computing solution. Depending on initial guess accuracy, the optimizer will need to perform dozens of iterations, each requiring a DRR generation and similarity calculation step, per bone, for each frame of data. Parallel algorithms allow for substantial computational savings and speedup, which in turn increases bone tracking throughput. In large, multi-bone, multi-condition studies, the data processing time is the limiting factor. The proper choice of image similarity metric, or a weighted combination of them, most likely varies with the type of bone and image conditions present. To make an informed selection of image similarity metrics and to gauge their performance in tracking foot and ankle bones, a series of experiments were conducted using custom-designed human bone phantoms.

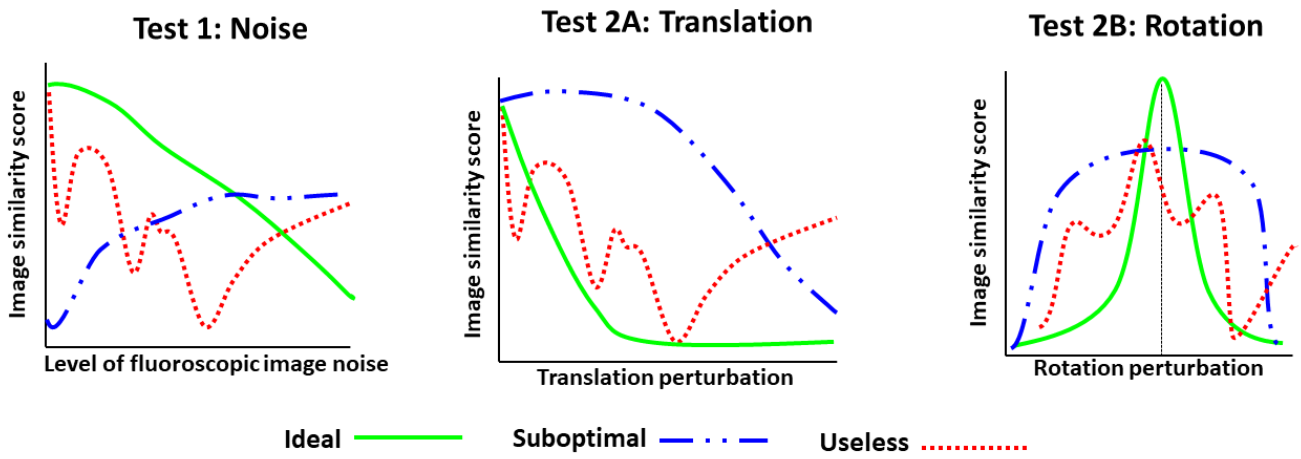

Figure 6.2 Expected responses of three example image similarity metrics to increasing amounts of image noise (left), and spatial translation (center) and rotation (right) errors relative to an ideally registered image. The response curves of an ideal (green), suboptimal (blue) and useless (red) metric are depicted for each test scenario. As in the ideal (green) curves, we seek metrics that vary smoothly and monotonically decay from the maximal score value at the co-registered position. Suboptimal metrics (blue) may not decay as rapidly and useless metrics are inherently noisy with multiple zero-crossings in the first derivative.

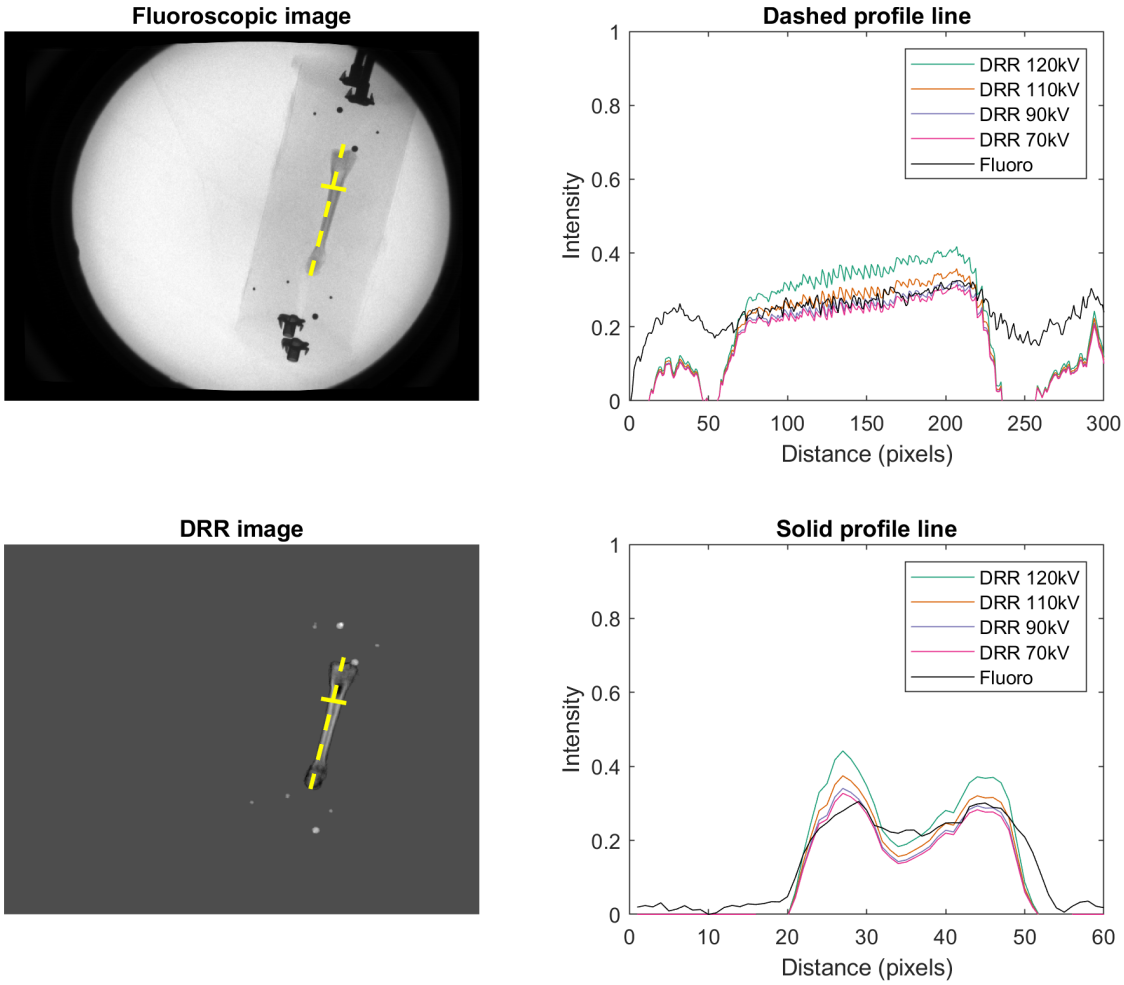

Figure 6.3 Differences in intensity profiles (right column) fluoroscope (top left) and DRR (bottom left) images at different X-ray source potential voltages. Since the attenuation coefficients of a material (μ) are a function of the voltage (Chapter 2) and affect the intensities of the radiographic projections. Note for both DRR and fluoroscope images how the relative intensity differences (height between peaks of profile plot) increase (compare the green curve to the red curve).

6.2 IMAGE SIMILARITY METRICS

This section provides a brief summary of the types of image similarity metrics considered for the registration problem, the expected advantages and any special considerations for each. The following operations are performed at each matching pixel index (row **i**, column **j**) in the aligned DRR and underlying fluoroscope image ($\mathbf{I}_{\text{fluoro}}$) over a domain ($\boldsymbol{\Omega}$) that is limited to the pixels in the DRR image ($\mathbf{I}_{\text{DRR}}$) that are bone. This domain constraint reduces computational time and limits analysis to only the pertinent regions for tracking, for example avoiding generating data for empty regions of air. Differences in image intensity are summed over $\boldsymbol{\Omega}$ (all the pixel elements) into a single value ($\boldsymbol{m}$) that may serve as the cost in the optimization function for registering the bone DRR to $\mathbf{I}_{\text{fluoro}}$. The image similarity metrics considered in this work belong may be thought of belonging to one of three families:

6.2.1 *Intensity-based metrics*

Intensity-based metrics attempt to register the images by performing pixel-wise gray-level comparisons between the images at each pixel position (**i**,**j**) in **Ω.** Images that are exactly coregistered will have a minimum difference, or metric distance, in intensity values of matching pixels. The sum of **squared intensity differences** (SID, Equation 6.1) and the sum of the **logarithm of absolute intensity differences** (LAD, Equation 6.2) are the simplest metrics used in this study. The squared intensity differences algorithm is computationally cheap and serves as a simple benchmark for the other metrics. It may be thought of as the L^2 (Euclidean) norm of the pixel-wise difference matrix between the images:

$$\boldsymbol{m}_{SID} = \sum_{[i,j] \in \boldsymbol{\Omega}} [\boldsymbol{I}_{DRR}(i,j) - \boldsymbol{I}_{fluoro}(i,j)]^2 \tag{6.1}$$

A few pixels in $\boldsymbol{\Omega}$ with significantly larger values from the central tendency can inflate $\mathbf{m}_{\text{SID}}$. These outliers may be caused by quantum noise or scatter phenomena in the relatively noisy fluoroscope images or to readout noise in the high-speed camera. As an alternative metric, the LAD method suppresses such large outlier differences between the two images due to the nature of the logarithm. The formulation (Equation 6.2) includes modifications to avoid ill-conditioning for the case of zero difference at a given pixel, and rescaling for a maximum score of unity:

$$\boldsymbol{m}_{LAD} = \sum_{[i,j] \in \boldsymbol{\Omega}} \frac{\log\left(abs\left(\boldsymbol{I}_{DRR}(i,j) - \boldsymbol{I}_{fluoro}(i,j)\right) + 1\right)}{\log(2)} \tag{6.5}$$

Normalized cross-correlation (NCC, Equation 6.3) [113] is another way to quantify the similarity between signals. NCC first normalizes each image by subtracting the mean intensity value of the image ($\bar{\mathbf{I}}$) and dividing by the standard deviation of intensity values. The pixel-wise correlations between the two normalized images ($\hat{\mathbf{I}}_{\text{DRR}}$ and $\hat{\mathbf{I}}_{\text{fluoro}}$) are summed over the domain $\boldsymbol{\Omega}$ to produce the scalar $\mathbf{m}_{\text{NCC}}$:

$$\hat{\boldsymbol{I}}_{DRR} = \frac{\boldsymbol{I}_{DRR} - \bar{\boldsymbol{I}}_{DRR}}{\sqrt{\sum\left(\boldsymbol{I}_{DRR} - \bar{\boldsymbol{I}}_{DRR}\right)^2}}, \qquad \hat{\boldsymbol{I}}_{fluoro} = \frac{\boldsymbol{I}_{fluoro} - \bar{\boldsymbol{I}}_{fluoro}}{\sqrt{\sum\left(\boldsymbol{I}_{fluoro} - \bar{\boldsymbol{I}}_{fluoro}\right)^2}}$$

$$\boldsymbol{m}_{NCC} = \sum_{[i,j] \in \boldsymbol{\Omega}} \hat{\boldsymbol{I}}_{DRR}(i,j)\ \hat{\boldsymbol{I}}_{fluoro}(i,j) \tag{6.6}$$

By normalizing the images first, NCC is invariant to linear transformations to the intensity data. It can handle large differences in amplitudes between the signals, and is computationally cheap, especially if calculated in the Fourier Domain. These characteristics have made it one of the most commonly used image similarity metrics in medical image registration and biplane radiography, in general [18, 41, 113, 114].

6.2.2 *Information-based metrics*

Information-based metrics compare the statistics of the intensity distributions in each image. Entropy is a measure of the information content in an image. An image with a single or a few repeated intensity values has low entropy, low information content to guide the registration. There is variation in X-ray projections of bones; attenuation lends to darker and brighter image regions corresponding to trabecular and cortical bone tissues. These irregularities increase image entropy and yield essential information content that may be exploited for the comparison of two images. Consider a floating-point representation of a grayscale image with intensity values $\mathbf{x} \in [0.0\ 1.0]$ and probability mass function, $p(\mathbf{x}) \in [0\ 1]$. In grayscale images, the marginal Shannon entropy ($\mathbf{H}$) of the intensity data may be calculated as:

$$\boldsymbol{H}(image) = -\sum p(\boldsymbol{x}) \log p(\boldsymbol{x}) \tag{6.7}$$

Equation 6.8 can be modified using the joint probabilities, $p(\mathbf{x}_{DRR,}\mathbf{x}_{fluoro})$, between image intensity occurrences, $\mathbf{x}_{DRR,}$ and $\mathbf{x}_{fluoro}$, to generate the joint entropy of the images: $\mathbf{H}(\mathbf{x}_{DRR,}\mathbf{x}_{fluoro})$. Combining the marginal and joint entropies of the DRR and fluoroscope image, the Shannon **mutual information** (MI) may then be defined as [115]:

$$\boldsymbol{m}_{MI} = \boldsymbol{H}_{DRR} + \boldsymbol{H}_{fluoro} - \boldsymbol{H}_{DRR,fluoro} \tag{6. 8}$$

In the optimal coregistered position, the joint entropy is minimized, marginal entropies are maximized, and the MI between the two is maximized. Mutual information metrics have been employed in the coregistration of multimodality datasets like magnetic resonance imaging and computed tomography. The same tissues appear with very different grayscale distributions in the respective modalities, however, the statistics of the occurrences of those values must be similar. While a multitude of both linear and non-linear differences between the DRR and fluoroscope images have been highlighted in the previous section, information-based metrics should be robust to many of these.

The histogram-based nature of MI lacks reference to any spatial pixel coordinates (**i**,**j**) in the formulas, making MI ignorant to the extra information that could be provided by observing similar patches at similar locations in both images. For example, imagine an image compared to a copy of itself in which the pixel locations were randomly scrambled. MI would be identical for all random re-orderings of the data; location within the image does not matter. **Regional mutual information** (RMI), described by Russakoff et al. [116] builds upon the robustness of MI to global intensity variations but considers multiple tiles of the local spatial intensity distributions. For each pixel in the image domain **Ω,** the intensity values are sampled within an **n** x **n** neighborhood yielding an $\mathbf{n}^2$-dimensional distribution as opposed to the usual one-dimensional distribution of intensities of an image sampled once at each pixel. By assessing the mutual information over the higher-dimensional space formed by this series of regional intensity distributions, the distributions of features like bone edges that are in proximity (i.e. exist in the same neighborhood) are directly compared rather than being lost in a single, global histogram. For a formal formulation of this method, see Section 2.4 of Russakoff [116].

Lastly, **pattern intensity** (PI), described by Weese [113, 117] acts on the assumption that at the perfectly coregistered position, the difference image $\mathbf{I}_{\text{difference}}$ formed by subtracting $\mathbf{I}_{\text{fluoro}}$ and $\mathbf{I}_{\text{DRR}}$ should be nearly uniform aside from a few pixel-wise differences that may exist. For each pixel in **Ω,** the intensity values are sampled at pixel locations **u**,**v** within a specified neighborhood radius of location **i**,**j** to form subdomain **ω**. The magnitude of the $\mathbf{I}_{\text{fluoro}}$ and $\mathbf{I}_{\text{DRR}}$ differences within each **ω** are compared using a variance sensitivity parameter, **σ**:

$$\boldsymbol{m}_{PI} = \sum_{[i,j]\in \boldsymbol{\Omega}} \sum_{[u,v]\in \boldsymbol{\omega}} \frac{\sigma^2}{[\boldsymbol{I}_{difference}(i,j) - \boldsymbol{I}_{difference}(u,v)] + \sigma^2} \tag{6. 9}$$

Weese et al. used a neighborhood radius = 3 and **σ** = 10 in [117]. After preliminary tests with our data, we used a radius = 2 and **σ** = 10 for foot and ankle bones.

6.2.3 *Gradient-based metrics*

Gradient-based metrics use cues from the local spatial variation of intensity (edges, bony features). Calculating the spatial gradient of the image intensities is essentially taking the derivative of an already noisy fluoroscope image and will tend to decrease the signal-to-noise ratio in the output. A pre-smoothing step is often employed to reduce local random fluctuations in intensity and preserve signal typically, a Gaussian filter. These filters may even be efficiently combined into a single processing step/kernel (e.g., Laplacian of Gaussian or derivative of Gaussian). The output of the gradient estimation gives two useful images: the gradient magnitude ($\mathbf{G}_{\text{mag}}$) and the gradient orientations ($\mathbf{G}_{\theta}$) of detected edges. Gradient magnitudes are larger at cortical bone edges and other dense objects. Orientations are expressed as angles relative to the image pixel grid (e.g. counterclockwise angle from the horizontal axis). The gradient orientations in the fluoroscope and DRR images that are exactly out-of-phase are treated as equivalent (e.g. $+0.25\pi \equiv -0.75\ \pi$).

The simplest gradient-based measure compared is **gradient difference** (GD) [113], which can be thought of as performing the SID algorithm (Equation 6.1) on the magnitudes of gradient images $\mathbf{G}_{\text{fluoro}}$ and $\mathbf{G}_{\text{DRR}}$:

$$\boldsymbol{m}_{GD} = \sum_{[i,j]\in \boldsymbol{\Omega}} [\boldsymbol{G}_{mag,DRR}(i,j) - \boldsymbol{G}_{mag,fluoro}(i,j)]^2 \tag{6.10}$$

The **gradient correlation** (GC) may be thought of as performing the NCC operation (Equation 6.3) on the magnitudes of the gradient images $\mathbf{G}_{\text{fluoro}}$ and $\mathbf{G}_{\text{DRR}}$:

$$\widehat{\boldsymbol{G}}_{DRR} = \frac{\boldsymbol{G}_{DRR} - \overline{\boldsymbol{G}}_{DRR}}{\sqrt{\sum\left(\boldsymbol{G}_{DRR} - \overline{\boldsymbol{G}}_{DRR}\right)^2}}, \qquad \widehat{\boldsymbol{G}}_{fluoro} = \frac{\boldsymbol{G}_{fluoro} - \overline{\boldsymbol{G}}_{fluoro}}{\sqrt{\sum\left(\boldsymbol{G}_{fluoro} - \overline{\boldsymbol{G}}_{fluoro}\right)^2}}$$

$$\boldsymbol{m}_{GC} = \sum_{[i,j] \in \boldsymbol{\Omega}} \widehat{\boldsymbol{G}}_{DRR}(i,j)\ \widehat{\boldsymbol{G}}_{fluoro}(i,j) \tag{6.11}$$

Gradient information (GI) [118] described by De Silva applies the min operator to both gradient magnitudes to suppress extraneously large gradients and a weighting function based on the angular difference in edge orientation:

$$\boldsymbol{m}_{GI} = \frac{1}{N}\sum_{[i,j] \in \boldsymbol{\Omega}}^{N} w(i,j)\ \min\left(\boldsymbol{G}_{mag,DRR}(i,j), \boldsymbol{G}_{mag,fluoro}(i,j)\right) \tag{6.12}$$

$$\text{where, } w(i,j) = \frac{\cos\left(\boldsymbol{G}_{\theta,DRR}(i,j) - \boldsymbol{G}_{\theta,fluoro}(i,j)\right)+1}{2}$$

Gradient orientation (GO) [118] also described by De Silva ignores pixels with gradients below the median gradient magnitude and replaces the cosine weighting penalty function with a faster-decaying natural log model:

$$\boldsymbol{m}_{GI} = \frac{1}{\max(N, N_{lb})}\sum_{[i,j] \in \{\boldsymbol{\Omega}:\, \boldsymbol{G}_{mag,DRR} > median(\boldsymbol{G}_{mag,DRR}) \cap \boldsymbol{G}_{mag,fluoro} > median(\boldsymbol{G}_{mag,fluoro})\}}^{N} w'(i,j) \tag{6.13}$$

$$\text{where, } w'(i,j) = \frac{2 - \left(\ln\left(\left|\cos^{-1}\left(\cos\left(\boldsymbol{G}_{\theta,DRR}(i,j) - \boldsymbol{G}_{\theta,fluoro}(i,j)\right)\right)\right|+1\right)\right)}{2}$$

6.3 IMAGE SIMILARITY METRIC COMPARISON

The following describes a cadaver-based imaging experiment for obtaining data characterizing the performance of the image similarity metrics outlined in Section 6.2.

6.3.1 *Aims*

The main goal of this experiment was to characterize the response of the image similarity metrics to conditions of image noise and errors in the initial position guess. Additional aims were to elucidate the performance limitations of the metrics in the previous implementation of our tracking software and explore the efficacy of metrics based on *information* between the two images, rather than only intensity or gradient correlations.

6.3.2 *Methods*

Two human cadaveric foot specimens were carefully dissected to extract the talus, calcaneus, and first metatarsals. These bones were selected given their anatomical and clinical importance to the foot and ankle, and the tracking challenges they present due to variations in their shape and size. Bones were manually denuded of soft tissues, then cast in the center of a minimally attenuating polyurethane resin cylinder (Figure 6.4, left). At the opposing ends of each cylinder, a unique pattern of four stainless steel bearings was embedded in a manner that prevented their overlap with the projection of the bone. These marker patterns allowed the exact position and orientation of the cylinder to be recovered independently of any bone information. Threaded inserts embedded into the cylinder endcaps facilitated pose manipulation of the phantoms. The locations of the stainless-steel markers relative to the bone tissues in each phantom were determined using CT scans. A clinical-grade CT scanner (GE Lightspeed, voxel size: 0.54/0.54/0.625 mm) was used to acquire data for generating DRRs, while a higher-resolution micro-CT scanner (Siemens, Inveon, isotropic voxel size: 0.074 mm) was used to determine the marker positions with higher precision.

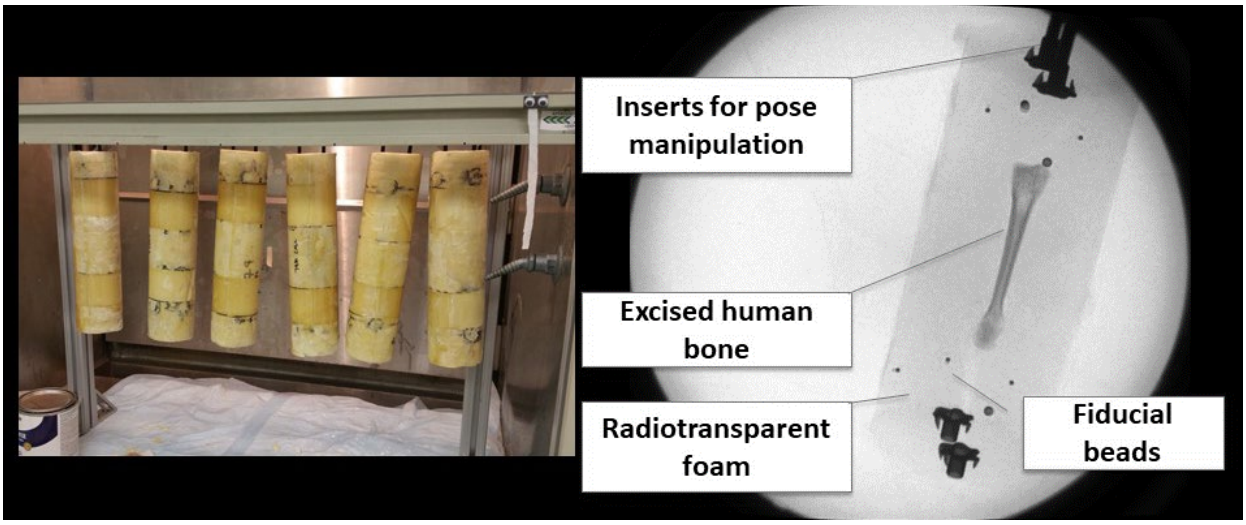

Figure 6.4 Fabrication (left) and fluoroscopic image (right) of bone phantoms using excised human specimens embedded in radiotransparent foam with fiducial markers for determining the ground-truth bone pose.

Bone phantoms were imaged in the biplane system using acquisition settings that are typical for in vivo data collection (Figure 6.4, right). Data were collected both with the bones roughly oriented in the position and orientations encountered during in vivo gait, and in spatially varying positions relative to the image intensifier faces. All fluoroscopic image data were corrected using the distortion and flatfield procedures outlined earlier (Chapter 3). The ground truth pose of the bone for each trial was determined by localizing the stainless beads in the ends of the phantoms.

First, the DRRs were generated with the bones in the ground truth position using the marker bead data. These DRR projections were considered the best-aligned cased that should yield the highest similar scores relative to the fluoroscopic images. At these truth positions, image similarity metrics were calculated for the DRRs after the fluoroscopic images were further corrupted by zero-mean, Gaussian noise of increasing variance. The mean response of each metric to increasing amounts of noise was calculated (Figure 6.5). The sensitivity of each metric to perturbations in the perfect coregistration of the images was assessed by adding displacements (n=1000) to the position and orientation of the DRR bone and recalculating the similarity scores. All transformations were applied relative to the projected center of mass of the bone volume. The

translation displacement vectors were uniformly-distributed between 0 and 10 pixels (about 3 mm in world space).

6.3.3 *Results*

The addition of image noise affected each metric score differently (Figure 6.5). Gradient Information (GI) and Pattern Intensity (PI) were unaffected by increasing levels of noise, while all other metric scores indicated a greater dissimilarity between the fluoroscopic images and the DRRs. After an initial amount of added noise, Gradient Orientation (GO) and Gradient Difference (GD) scores were both nearly constant (Figure 6.5). The Mutual Information (MI) metric score did not monotonically vary with increasing noise. All other metrics varied as expected with increasing noise; smoothly indicating an increased difference between the images.

Metric score responses to errors in translation (Figure 6.6) and rotational (Figure 6.7) misalignments varied by the ISM algorithm. Intensity-based measures (SID, LAD, NCC) had flat responses for increasing amounts of translation error (Figure 6.6). MI, GC, and GO smoothly decreased with larger perturbation error, as desired. A similar pattern of behaviors was observed for rotational perturbations (Figure 6.7).

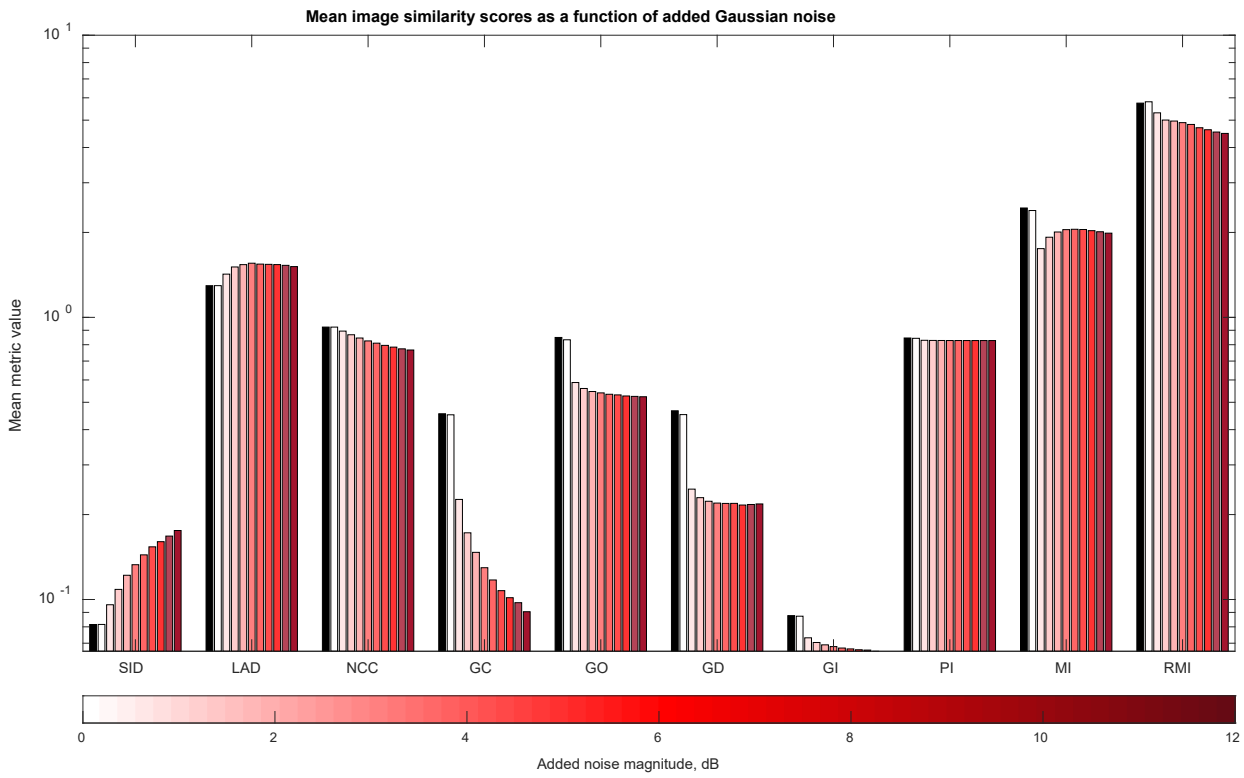

Figure 6.5 Responses of each similarity metric type with increasing amounts (darker red) Gaussian noise added to the fluoroscopic image data. Each bar is the average across all bones and all image samples. The black bars are the average truth position scores with no noise added.

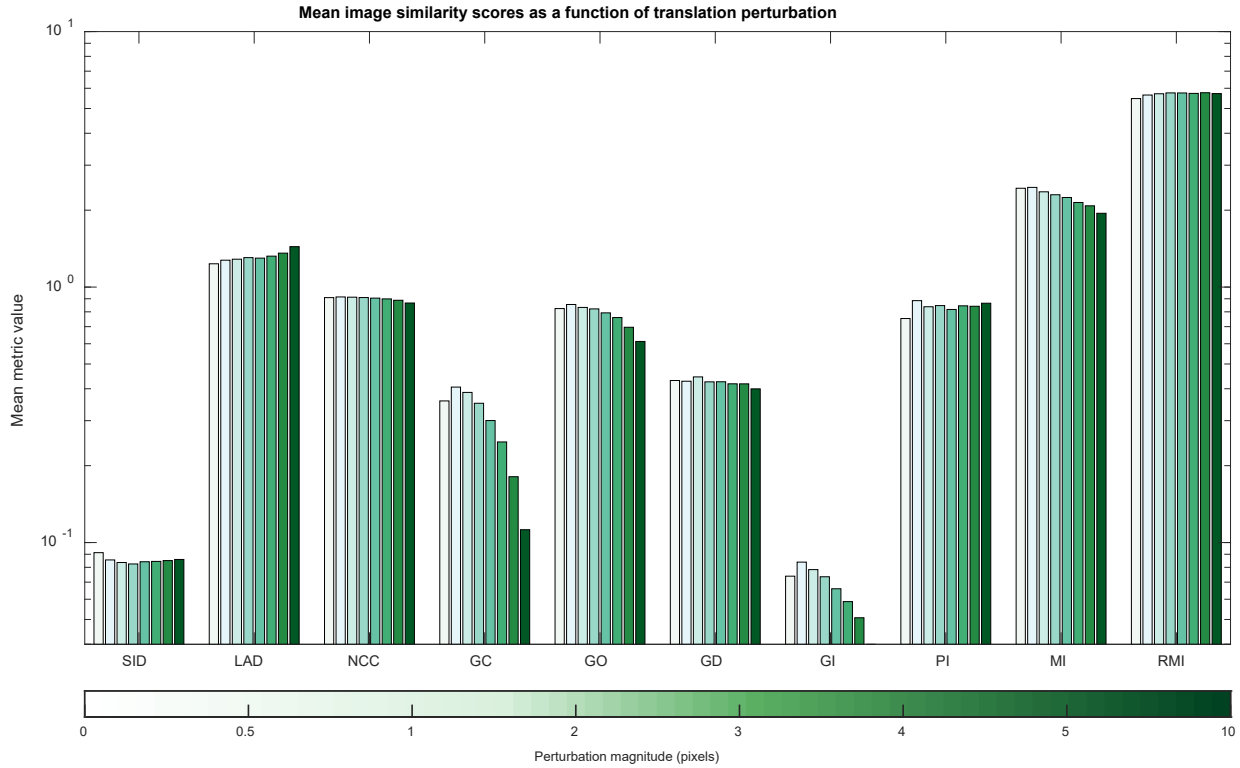

Figure 6.6 Responses of each similarity metric type with increasing amounts (darker green) translation perturbation of the DRR from the true position.

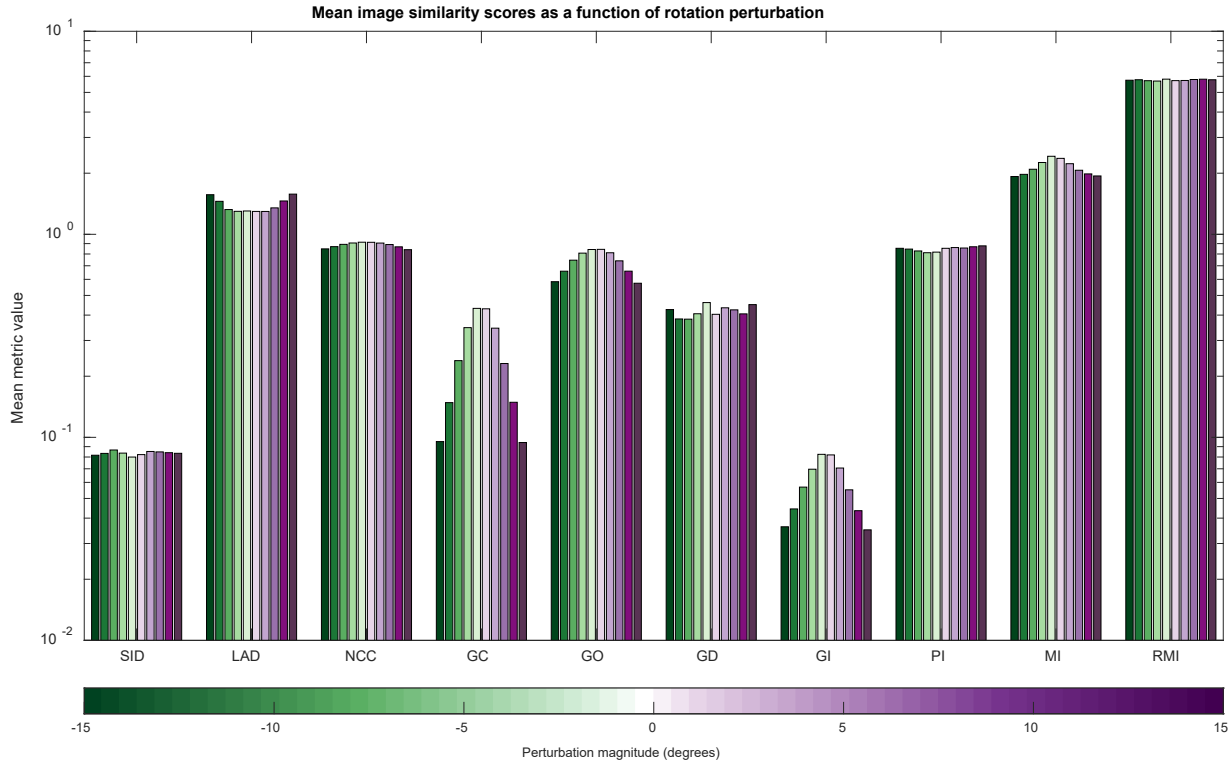

Figure 6.7 Responses of each similarity metric type with rotational perturbation of the DRR from the true position.

6.3.4 *Discussion*

Most of the image similarity metrics decreased in value with added noise, as expected, with a few notable exceptions. Squared intensity differences (SID) scores increased with noise, as the magnitude in differences between images increased. Pattern intensity (PI) and Gradient information (GI) exhibited flat response curves to the added noise. Except for mutual information (MI), all of the other metrics exhibited the mostly smooth, monotonic variation in score with increasing error magnitude. Counter-intuitively, after an initial decrease, mutual information score increased with added Gaussian noise, suggesting that the intensity distributions became more similar only after the fluoroscopic image was further-corrupted with Gaussian noise. Regional mutual information (RMI) was more performant than MI in situations of added noise, perhaps due to the added spatial-information component to the function that accounts for local differences in intensity distributions.

The similarity metrics exhibited different responses to translational and rotational perturbations. SID, PI, and RMI were insensitive to perturbations in image alignment, as indicated by the flat responses in Figure 6.6 and Figure 6.7. NCC varied smoothly, but subtly for both perturbation types. With these incremental differences in NCC, scores may not be lost to numerical round-off errors in the optimization engine, a larger gradient in the similarity score response may drive the registration to faster or more stable convergence. MI, GC, GI, and GO exhibited the best response characteristics to spatial perturbations by tapering smoothly towards zero for the extreme misalignments. Presently, these three metrics have the greatest potential for driving the foot and ankle DRR-fluoroscope registration problem.

The experiments and analyses presented in this section are an idealized set of test conditions with many limitations, including a limited sample of bones imaged under the best conditions (without soft tissue attenuation or scatter artifact). The noise perturbation testing assumes that the added Gaussian noise is a valid noise model. Additionally, the processing time for calculating the similarity scores was neglected in this exploratory analysis. Specific computational optimizations and short-cuts may be implemented in the metric(s) that are deemed potentially useful for the DRRACO optimizer. Lastly, these scores could be combined as linear combinations either in parallel, or serially in a hierarchical tracking procedure. Future work will explore combining these metrics and quantifying their response in situations that are high in X-ray scatter noise.

6.4 IMAGE GRADIENT CALCULATIONS

In the gradient-based family of image similarity metrics, features like bony edges in the images are emphasized by taking the spatial derivative of the image signal. As previously mentioned, taking the derivative of a noise-contaminated signal will lower the signal-to-noise ratio and may introduce spurious signal artifacts that may be mistaken for real image features. Striking a balance between signal (detected bone edge magnitude) and noise (background of the image, e.g. soft tissue, air, shoes, walkway flooring) in the gradient images is essential to the smooth, efficient, and robust DRR to fluoroscope image registration. In the experiments outlined in the previous chapters and the current implementation of the bone tracking software, DRRACO [18],

the Sobel method for extracting image edges was utilized for each test case. However, much like the localization algorithms and image similarity metrics, the edge detection scheme is an important enough component of the system to warrant its own investigation. Given the promising results of the gradient-based image similarity metrics in response to perturbations, an experiment was designed to explore alternative means for detecting bony features (edges) in the fluoroscopic images given the typical levels of quantum noise, false edges, and scatter artifacts present in a foot and ankle gait fluoroscopic image pair. Additionally, quantitative data characterizing the magnitude and cleanliness of detected edges in both an idealized test case using a line pair object and a simulation of in vivo conditions using a foot phantom were sought to inform revisions to the gradient image calculation paradigm in future updates to the bone tracking software.

6.4.1 *Methods*

Fluoroscopic images using the standard laboratory camera configurations were produced for two test cases. The first test case assesses edge detection of a radiographic line pair object at three modulation frequencies: 1.5, 0.75, and 0.5 line pairs per millimeter. Recall from the Chapter 3 discussion of the modulation transfer function that the resolving power of the optical chain was determined to be limited at around 1.5 line pairs per millimeter. Thus, edges that are closer together than this limit are not resolvable. Typical healthy articular joint spacing (ergo, the distance between bone edges of interest) in the foot and ankle is on the order of 1.0 millimeters [119]. In the second test case of the foot and ankle during gait (Figure 6.8) using the following algorithms applied via spatial convolutional operations across the image domain. Kernel size was limited to a 3 x 3 neighborhood for all cases. For the current hardware setup, this kernel size in pixels corresponds to a 1 mm^2 region on the image. Each edge detection kernel described in the following section was implemented in MATLAB software via convolution spatial operations, and the output gradient images were compared qualitatively and via intensity profile plots. The intensity profile indicates the magnitude of the detected edges as a function of a location along the sampling line. In our fluoroscopic images, bony edges are dark regions (low regions of the intensity profile), while areas of less attenuation like air are brighter (higher on the intensity profile plot).

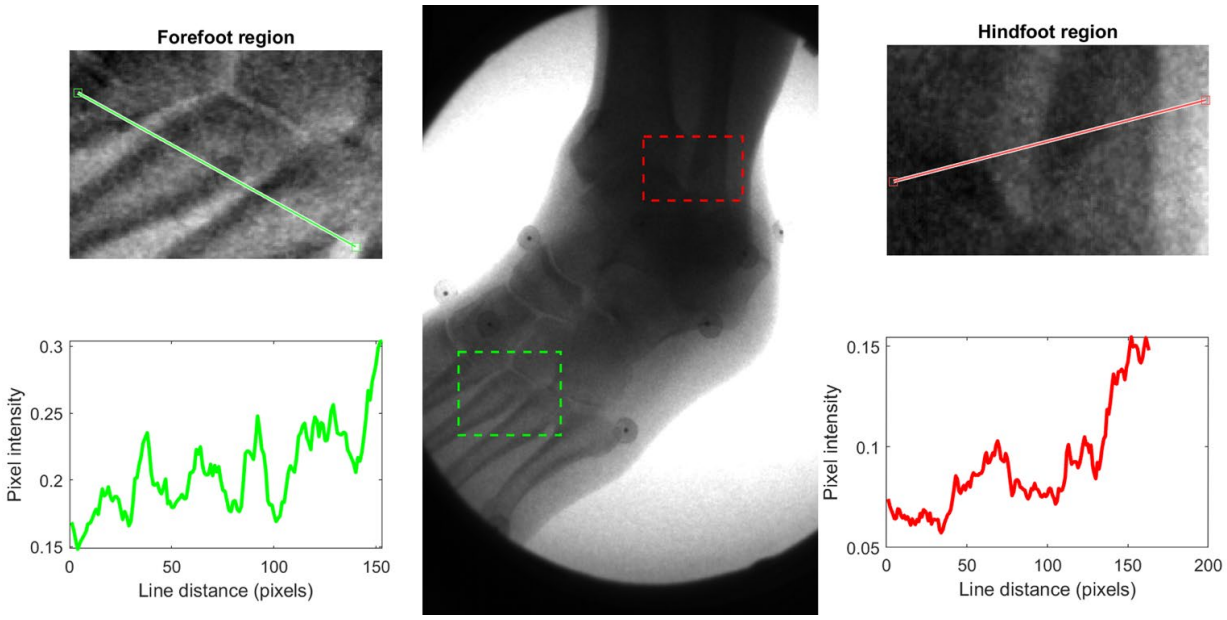

Figure 6.8: Example fluoroscopic image of a human foot during walking. Two regions of interest: a forefoot region (green) and hindfoot region (red) with intensity profiles are depicted. The local variations in intensity (peaks on the profile lines) provide essential information for the gradient-based image similarity metrics. The smoothness and prominence characteristics of the peaks are functions of the local image noise, contrast, and X-ray scatter.

6.4.1.1 Central difference

The central difference is the simplest approximation to the first spatial derivative of the image, utilizing a pair of kernels for detecting neighboring pixel differences for the rows and columns, respectively:

$$H_{rows} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad H_{columns} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

6.4.1.2 Separated central difference

Most bony edges are thicker than one image pixel. As a result, the central difference operator may be too small and inflate the noise across the edge region. To reduce these negative effects,

the separated central difference kernel introduces a gap in the middle of the operator to enhance thicker edges (indicated by the “zero” in the center”):

$$H_{rows} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad H_{columns} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

6.4.1.3 Roberts

The Roberts cross operator is one of the earliest and simplest feature detectors from the computer vision field [120] that attempts to detect diagonal signal changes as edges:

$$H_{rows} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad H_{columns} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

6.4.1.4 Prewitt

The Prewitt operator is another simple detector that may be thought of as performing the central difference operator while sampling the entire row and column of the kernel. The final kernels must be normalized to the total energy (sum of all elements) equals zero:

$$H_{rows} = \frac{1}{3} * \begin{bmatrix} 3 & 0 & -3 \\ 3 & 0 & -3 \\ 3 & 0 & -3 \end{bmatrix} \qquad H_{columns} = \frac{1}{3} * \begin{bmatrix} 3 & 3 & 3 \\ 0 & 0 & 0 \\ -3 & -3 & -3 \end{bmatrix}$$

6.4.1.5 Sobel

The Sobel operator may be thought of as a hybrid between the Prewitt and central difference operators. As in the Prewitt filter, the entire row/column is sampled but the end pixel values have lower weightings compared to the central pixel:

$$H_{rows} = \frac{1}{4} * \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \qquad H_{columns} = \frac{1}{4} * \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

6.4.1.6 Laplacian point

The Laplacian point operator is a rotationally-invariant, symmetrical second-order differentiator based on the divergence of the gradient field. For a discrete signal, the center pixel of the kernel is up-weighted relative to the neighboring values via the operator approximation:

$$H_{Laplacian\ point} = \frac{1}{8} * \begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix}$$

6.4.1.7 Laplacian line

True bone edges are contiguous and have a length greater than a few pixels. The point-wise Laplacian operator considers only the center pixel value and does not preferentially treat connected regions edge regions. As a modification to the previous method, a set of four Laplacian line kernels are convolved with the signal giving the gradients in the horizontal, vertical, and two diagonal directions:

$$H_{vertical\ line} = \frac{1}{6} * \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}$$

$$H_{horizontal\ line} = \frac{1}{6} * \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$$

$$H_{diagonal\ line} = \frac{1}{6} * \begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix}$$

$$H_{diagonal\ line} = \frac{1}{6} * \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

6.4.1.8 Abdou

Abdou described a 7 x 7 element kernel for edge detection [121]. This “truncated pyramid” linearly weights pixels closer to the center of the kernel:

$$H_{rows} = \frac{1}{34} * \begin{bmatrix} 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 2 & 2 & 0 & -2 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 2 & 0 & -2 & -2 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \end{bmatrix}$$

$$H_{columns} = \frac{1}{34} * \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -2 & -2 & -2 & -2 & -2 & -1 \\ -1 & -2 & -3 & -3 & -3 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 3 & 3 & 2 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

6.4.1.9 Frei-Chen

The Frei-Chen operator is a combination of kernels that represent different local features in images. Examples of these basic features can be depicted graphically (Figure 6.9). Each feature space kernel is analogous to the convolution kernels described above. The responses from each kernel are summed together in weighted combinations of the nine feature types. Specific features may be up- or down-weighted as needed to enhance or suppress features of a certain type or orientation. This strategy may be useful when tracking bone edges superimposed against another radiographic grid caused by the X-ray attenuation of an electronic plantar pressure sensor or a dense shoe sole, two common scenarios during in vivo clinical trials.

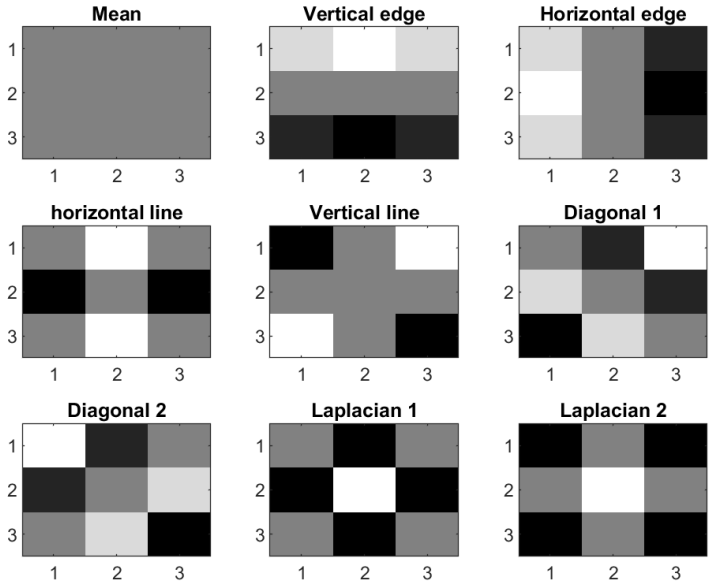

Figure 6.9: Examples of the Frei-Chen feature spaces that represent local neighbor "objects" like lines, edges, and ripples.

6.4.1.1 Calculating magnitude and orientations of gradients

By convolving the above operators with the image, the edge response along the direction of interest is calculated. The magnitudes of the *n* directionally-dependent edge images, $\mathbf{E}_n$ (e.g. horizontal, vertical, diagonal edges), are averaged together to provide an overall edge magnitude ($\mathbf{E}_{magnitude}$) of the local edge using the following equation:

$$\boldsymbol{E}_{magnitude} = \frac{1}{n}\sum_{i=1}^{n} abs(\boldsymbol{E}_i) \tag{6.14}$$

Estimations on the orientations of the edges at each edge pixel are required for the Gradient Orientation similarity metric (Section 6.3).

$$\theta = \tan^{-1}\left(\frac{E_{columns}}{E_{rows}}\right) \tag{6.15}$$

6.4.2 *Results*

The resulting edge images detected in the forefoot region (Figure 6.8) images for each operator are depicted (Figure 6.10). Qualitatively, the central difference operator and Roberts operator generated weak bone edges, and both Laplacian kernels generated extremely noisy signals. Quantitatively, the Abdou, Sobel, Prewitt, and separated central difference kernels extracted larger magnitude edges relative to the background than the Laplacian-based kernels and the Roberts and central difference kernels (Figure 6.11). Frei-Chen detected edge regions (pink shaded area Figure 6.11) corresponded well with peaks detected from most of the other kernels. The wider Abdou kernel produced more contiguous edge maps, while Frei-Chen reduced spurious edges in the interior region of the bone. In the test case involving the line pair object, the Laplacian-based kernels were again the worst-performing except for the finest edge frequency case (1.5 line pairs per millimeter, bottom Figure 6.12). For edge frequencies of 0.3 and 0.75 line pairs per millimeter, there was a greater similarity in the detected edge magnitudes for the non-Laplacian kernels. The Abdou kernel could not detect edge frequencies at 1.5 line pairs per millimeter (Figure 6.12).

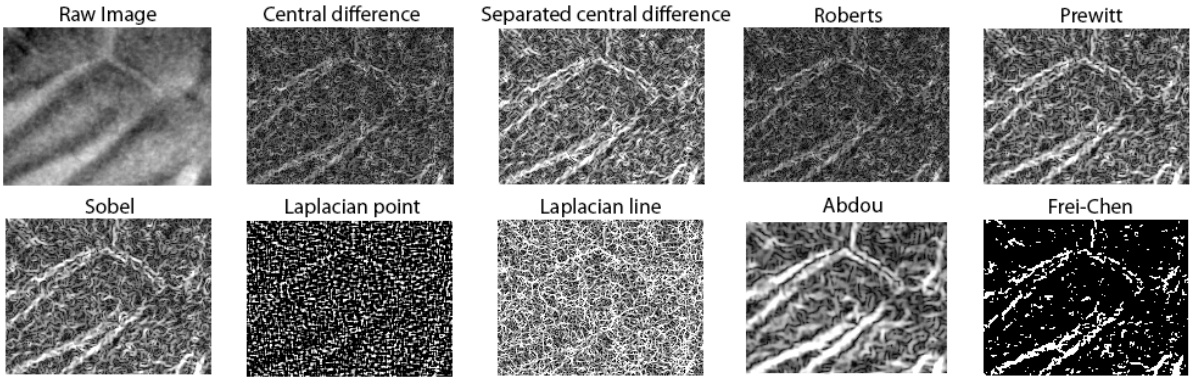

Figure 6.10: The forefoot region of the test image and the resulting gradient images from each of the operators. The separated central difference operator detected more pronounced metatarsal edges compared to the central difference operator and the Roberts kernel. The wider Abdou kernel generated an even more contiguous and pronounced edge map. Both Laplacian operators yielded poor results, likely due to the required extra derivative of the noisy fluoroscopy image. The Frei-Chen feature detection reduced the spurious false edges present inside the middle of the bone and air regions.

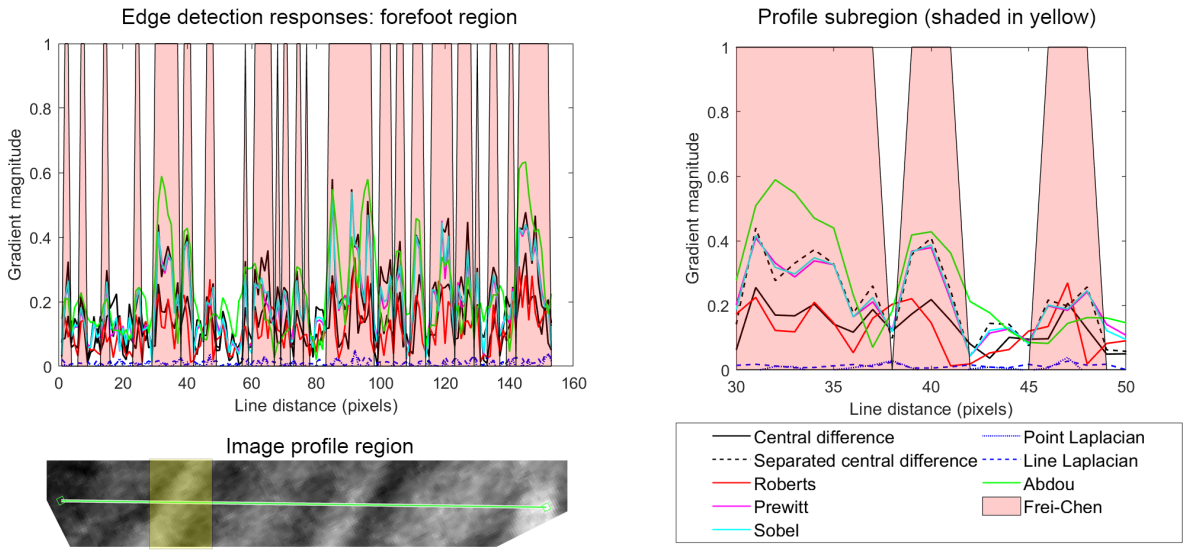

Figure 6.11: Profile plots (left top) for each method the detected gradient magnitudes in the forefoot subregion (bottom left) of the sample images in Figure 6.8. For clarity, a zoomed-in portion of the profile line is depicted on the right side. The pink shaded regions correspond to edge regions detected by the Frei-Chen operator.

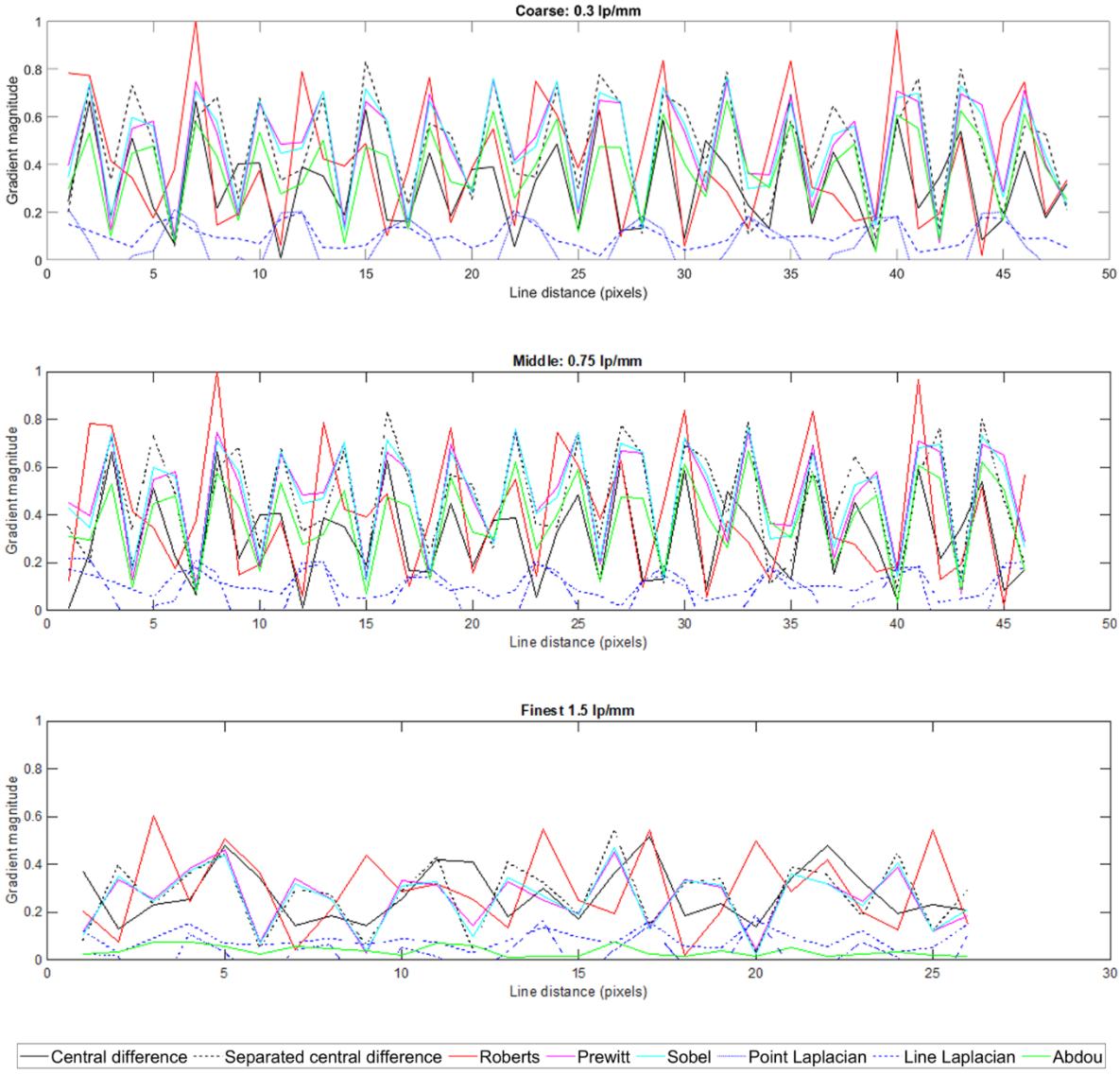

Figure 6.12: Intensity profiles detected edges of a line pair object at three modulation frequencies (0.3, 0.75, 1.5 line pairs per millimeter). The Frei-Chen detector results are omitted for clarity: every modulation edge was detected across all cases.

6.4.3 *Discussion*

For each of these simple test cases, the best and worst attributes of each kernel are displayed in the resultant edge profiles on the gradient images. The Laplacian-based kernels, requiring higher-

order derivatives, decreased the signal-to-noise ratio of the edges and led to noisy gradient images. Those kernels were useful in the detection of very fine edges (0.3-line pairs per millimeter) in the line pair object and could be potentially be utilized for image enhancement of fine markers, meshes, or other fiducials of interest in fluoroscopic images. The larger sampling size of the Abdou kernel leads to a stronger detected bone edge, so long as the edge frequency is not too high, as seen in the line pair object case. The separated central difference kernel, a smaller kernel that is in a similar form to the Abdou kernel struck a balance of the best and worst attributes of the Abdou kernel. Future work should further investigate performance improvements from increasing kernel size and pre-filtering the images before edge detection, although care must be taken to balancing the number of computations required which will directly influence the amount of time required to perform the DRR-to-fluoroscope registration optimization. An advantage to the kernels presented here is their separability and ability to be implemented on a parallel processing computing architecture such as a graphics processing unit. For example, each of the Frei-Chen feature responses may be separately computed and fused into a weighted vector score of edge likelihood for each pixel location in the image. Ultimately, the ideal recipe for detecting usable edges in fluoroscopic images will be dependent upon the anatomy of interest and the imaging parameters. By having a toolbox of edge detection algorithms at hand and an understanding of their best application cases, the most efficient and accurate object tracking may be possible.

Chapter 7. CONCLUSIONS

Accurate quantification of human joint motion will improve clinical understanding of a variety of issues that affect our daily lives. With continual advances in technology, the future of healthcare will revolve around individualized care and so-called “big data” approaches. Concerning biplane systems, computed tomography systems and X-ray detector panels are now available that can produce better images with less required radiation exposure to test subjects. The burden now shifts to making robust software to specifically address the challenges that have slowed the application of biplanar fluoroscopic imaging analysis. This thesis explored some of the root causes of algorithmic issues that are preventing a workflow of automated, robust processing pipelines that are required for handling the huge volumes of data produced by even the most modest research study. Beginning with a breakdown of the imaging chain sources of error and quantitative assessment of various solutions to the problems of image blur, defocusing, and spatial distortions. Having corrected the images to an adequate level, the next focus was fiducial localization; a core component of the calibration procedure for determining camera parameters, and the accepted gold-standard method for assessing bone kinematics via embedded markers. This investigation raised the question of how sensitive the calibration procedure was to the expected localization errors, and a re-development of the calibration object and algorithm to improve the geometric accuracy of measurements and object localization in the capture volume. Lastly, the core of the 3D-to-2D DRR and fluoroscopic image registration optimization routine, the image similarity metric, was investigated in the context of foot and ankle bones using custom-developed phantom objects. The results of these experiments have improved the laboratory’s understanding of the components of the biplane system, informed future research directions, and narrowed the field of the myriad algorithmic choices at each turn of the software development labyrinth.

BIBLIOGRAPHY

[1] Shibuya, N., Jupiter, D. C., Ciliberti, L. J., Vanburen, V., and La Fontaine, J., 2011, "Prevalence of Podiatric Medical Problems in Veterans versus Nonveterans.," Journal of the American Podiatric Medical Association, 101, pp. 323-330.
[2] Louie, P. K., Sangeorzan, B. J., Fassbind, M. J., and Ledoux, W. R., 2014, "Talonavicular joint coverage and bone morphology between different foot types," Journal of Orthopaedic Research, 32, pp. 958-966.
[3] Boyko, E. J., Seelig, A. D., and Ahroni, J. H., 2018, "Limb- and Person-Level Risk Factors for Lower-Limb Amputation in the Prospective Seattle Diabetic Foot Study," Diabetes Care, 41(4), pp. 891-898.
[4] Cowley, M. S., Boyko, E. J., Shofer, J. B., Ahroni, J. H., and Ledoux, W. R., 2008, "Foot ulcer risk and location in relation to prospective clinical assessment of foot shape and mobility among persons with diabetes," Diabetes Res Clin Pract, 82(2), pp. 226-232.
[5] Ledoux, W. R., Shofer, J. B., Ahroni, J. H., Smith, D. G., Sangeorzan, B. J., and Boyko, E. J., 2003, "Biomechanical differences among pes cavus, neutrally aligned, and pes planus feet in subjects with diabetes," Foot Ankle Int, 24(11), pp. 845-850.
[6] Muybridge, E., 1887, " Muybridge's Complete Human and Animal Locomotion: All 781 Plates from the 1887 Animal Locomotion by Eadweard Muybridge. , 1979. Print.," A. V. Mozley, ed., Dover Publishing, New York.
[7] Leardini, A., Stagni, R., and O'Connor, J. J., 2001, "Mobility of the subtalar joint in the intact ankle complex," Journal of Biomechanics, 34, pp. 805-809.
[8] Sancisi, N., Parenti-Castelli, V., Corazza, F., and Leardini, A., 2009, "Helical axis calculation based on Burmester theory: Experimental comparison with traditional techniques for human tibiotalar joint motion," Medical and Biological Engineering and Computing, 47, pp. 1207-1217.
[9] Colyer, S. L., Evans, M., Cosker, D. P., and Salo, A. I. T., 2018, "A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System," Sports Med Open, 4(1), pp. 24-24.
[10] Teufl, W., Miezal, M., Taetz, B., Fröhlich, M., and Bleser, G., 2018, "Validity, Test-Retest Reliability and Long-Term Stability of Magnetometer Free Inertial Sensor Based 3D Joint Kinematics," Sensors (Basel, Switzerland), 18(7), p. 1980.
[11] Teufl, W., Miezal, M., Taetz, B., Fröhlich, M., and Bleser, G., 2019, "Validity of inertial sensor based 3D joint kinematics of static and dynamic sport and physiotherapy specific movements," PloS one, 14(2), pp. e0213064-e0213064.
[12] Fassbind, M. J., Rohr, E. S., Hu, Y., Haynor, D. R., Siegler, S., Sangeorzan, B. J., and Ledoux, W. R., 2011, "Evaluating Foot Kinematics Using Magnetic Resonance Imaging: From Maximum Plantar Flexion, Inversion, and Internal Rotation to Maximum Dorsiflexion, Eversion, and External Rotation," Journal of Biomechanical Engineering, 133, p. 104502.
[13] Borotikar, B. S., and Sheehan, F. T., 2013, "In vivo patellofemoral contact mechanics during active extension using a novel dynamic MRI-based methodology," Osteoarthritis Cartilage, 21(12), pp. 1886-1894.
[14] Siegler, S., Udupa, J. K., Ringleb, S. I., Imhauser, C. W., Hirsch, B. E., Odhner, D., Saha, P. K., Okereke, E., and Roach, N., 2005, "Mechanics of the ankle and subtalar joints revealed through a 3D quasi-static stress MRI technique," Journal of Biomechanics, 38, pp. 567-578.

[15] Crisco, J. J., Halilaj, E., Moore, D. C., Patel, T., Weiss, A.-P. C., and Ladd, A. L., 2015, "In Vivo kinematics of the trapeziometacarpal joint during thumb extension-flexion and abduction-adduction," J Hand Surg Am, 40(2), pp. 289-296.
[16] Caputo, A. M., Lee, J. Y., Spritzer, C. E., Easley, M. E., Deorio, J. K., Nunley, J. A., and DeFrate, L. E., 2009, "In vivo kinematics of the tibiotalar joint after lateral ankle instability," American Journal of Sports Medicine, 37, pp. 2241-2248.
[17] Defrate, L. E., Papannagari, R., Gill, T. J., Moses, J. M., Pathare, N. P., and Li, G., 2006, "The 6 degrees of freedom kinematics of the knee after anterior cruciate ligament deficiency: An in vivo imaging analysis," American Journal of Sports Medicine, 34, pp. 1240-1246.
[18] Iaquinto, J. M., Kindig, M. W., Haynor, D. R., Vu, Q., Pepin, N., Tsai, R., Sangeorzan, B. J., and Ledoux, W. R., 2018, "Model-based tracking of the bones of the foot: A biplane fluoroscopy validation study," Comput Biol Med, 92, pp. 118-127.
[19] Iaquinto, J. M., Tsai, R., Haynor, D. R., Fassbind, M. J., Sangeorzan, B. J., and Ledoux, W. R., 2014, "Marker-based validation of a biplane fluoroscopy system for quantifying foot kinematics," Med Eng Phys, 36(3), pp. 391-396.
[20] Thorhauer, E., and Tashman, S., 2015, "Validation of a method for combining biplanar radiography and magnetic resonance imaging to estimate knee cartilage contact," Med Eng Phys, 37(10), pp. 937-947.
[21] Nichols, J., Roach, K., and Anderson, A., 2016, "Assessing the predictive accuracy of hindfoot models with dual-fluoroscopy as a reference," Foot and Ankle Surgery, 22, pp. 41-42.
[22] Pothrat, C., Goislard de Monsabert, B., Vigouroux, L., Viehweger, E., Berton, E., and Rao, G., 2015, "Quantifying foot deformation using finite helical angle," Journal of biomechanics, 48, pp. 3716-3719.
[23] Rouhani, H., Favre, J., Crevoisier, X., Jolles, B. M., and Aminian, K., 2012, "A comparison between joint coordinate system and attitude vector for multi-segment foot kinematics," Journal of Biomechanics, 45, pp. 2041-2045.
[24] Ahl, T., Dalen, N., Lundberg, A., and Selvik, G., 1987, "Mobility of the ankle mortise. A roentgen stereophotogrammetric analysis," Acta Orthop Scand, 58(4), pp. 401-402.
[25] Karrholm, J., Hansson, L. I., and Selvik, G., 1985, "Mobility of the lateral malleolus. A roentgen stereophotogrammetric analysis," Acta Orthop Scand, 56(6), pp. 479-483.
[26] Selvik, G., 1990, "Roentgen stereophotogrammetric analysis," Acta Radiol, 31(2), pp. 113-126.
[27] Sturesson, B., Selvik, G., and Uden, A., 1989, "Movements of the sacroiliac joints. A roentgen stereophotogrammetric analysis," Spine (Phila Pa 1976), 14(2), pp. 162-165.
[28] Nistor, L., Blaha, J. D., Kjellstrom, U., and Selvik, G., 1991, "In vivo measurements of relative motion between an uncemented femoral total hip component and the femur by roentgen stereophotogrammetric analysis," Clin Orthop Relat Res(269), pp. 220-227.
[29] Ryd, L., Lindstrand, A., Rosenquist, R., and Selvik, G., 1987, "Micromotion of conventionally cemented all-polyethylene tibial components in total knee replacements. A roentgen stereophotogrammetric analysis of migration and inducible displacement," Arch Orthop Trauma Surg, 106(2), pp. 82-88.
[30] Ryd, L., Lindstrand, A., Stenstrom, A., and Selvik, G., 1990, "Cold flow reduced by metal backing. An in vivo roentgen stereophotogrammetric analysis of unicompartmental tibial components," Acta Orthop Scand, 61(1), pp. 21-25.
[31] Ryd, L., Lindstrand, A., Stenstrom, A., and Selvik, G., 1992, "The influence of metal backing in unicompartmental tibial component fixation. An in vivo roentgen

stereophotogrammetric analysis of micromotion," Arch Orthop Trauma Surg, 111(3), pp. 148-154.
[32] Wykman, A., Selvik, G., and Goldie, I., 1988, "Subsidence of the femoral component in the noncemented total hip. A roentgen stereophotogrammetric analysis," Acta Orthop Scand, 59(6), pp. 635-637.
[33] Borlin, N., Thien, T., and Karrholm, J., 2002, "The precision of radiostereometric measurements. Manual vs. digital measurements," J Biomech, 35(1), pp. 69-79.
[34] Bonnan, M. F., Shulman, J., Varadharajan, R., Gilbert, C., Wilkes, M., Horner, A., and Brainerd, E., 2016, "Forelimb Kinematics of Rats Using XROMM, with Implications for Small Eutherians and Their Fossil Relatives," PLoS One, 11(3), p. e0149377.
[35] Brainerd, E. L., Moritz, S., and Ritter, D. A., 2016, "XROMM analysis of rib kinematics during lung ventilation in the green iguana, Iguana iguana," J Exp Biol, 219(Pt 3), pp. 404-411.
[36] Brocklehurst, R. J., Moritz, S., Codd, J., Sellers, W. I., and Brainerd, E. L., 2017, "Rib kinematics during lung ventilation in the American alligator (Alligator mississippiensis): an XROMM analysis," J Exp Biol, 220(Pt 17), pp. 3181-3190.
[37] Camp, A. L., and Brainerd, E. L., 2015, "Reevaluating Musculoskeletal Linkages in Suction-Feeding Fishes with X-Ray Reconstruction of Moving Morphology (XROMM)," Integr Comp Biol, 55(1), pp. 36-47.
[38] Menegaz, R. A., Baier, D. B., Metzger, K. A., Herring, S. W., and Brainerd, E. L., 2015, "XROMM analysis of tooth occlusion and temporomandibular joint kinematics during feeding in juvenile miniature pigs," J Exp Biol, 218(Pt 16), pp. 2573-2584.
[39] You, B. M., Siy, P., Anderst, W., and Tashman, S., 2001, "In vivo measurement of 3-D skeletal kinematics from sequences of biplane radiographs: Application to knee kinematics," IEEE Transactions on Medical Imaging, 20, pp. 514-525.
[40] Cross, J. A., McHenry, B. D., Molthen, R., Exten, E., Schmidt, T. G., and Harris, G. F., 2017, "Biplane fluoroscopy for hindfoot motion analysis during gait: A model-based evaluation," Med Eng Phys, 43, pp. 118-123.
[41] Arner, J. W., Irvine, J. N., Zheng, L., Gale, T., Thorhauer, E., Hankins, M., Abebe, E., Tashman, S., Zhang, X., and Harner, C. D., 2016, "The Effects of Anterior Cruciate Ligament Deficiency on the Meniscus and Articular Cartilage: A Novel Dynamic In Vitro Pilot Study," Orthop J Sports Med, 4(4), p. 2325967116639895.
[42] Irvine, J. N., Arner, J. W., Thorhauer, E., Abebe, E. S., D'Auria, J., Schreiber, V. M., Harner, C. D., and Tashman, S., 2016, "Is There a Difference in Graft Motion for Bone-Tendon-Bone and Hamstring Autograft ACL Reconstruction at 6 Weeks and 1 Year?," Am J Sports Med, 44(10), pp. 2599-2607.
[43] Tang, J., Thorhauer, E., Bowman, K., Fu, F. H., and Tashman, S., 2017, "In vivo posterior cruciate ligament elongation in running activity after anatomic and non-anatomic anterior cruciate ligament reconstruction," Knee Surg Sports Traumatol Arthrosc, 25(4), pp. 1177-1183.
[44] Tashiro, Y., Sundaram, V., Thorhauer, E., Gale, T., Anderst, W., Irrgang, J. J., Fu, F. H., and Tashman, S., 2017, "In Vivo Analysis of Dynamic Graft Bending Angle in Anterior Cruciate Ligament-Reconstructed Knees During Downward Running and Level Walking: Comparison of Flexible and Rigid Drills for Transportal Technique," Arthroscopy, 33(7), pp. 1393-1402.
[45] Zheng, L., Carey, R., Thorhauer, E., Tashman, S., Harner, C., and Zhang, X., 2018, "In vivo tibiofemoral skeletal kinematics and cartilage contact arthrokinematics during decline walking after isolated meniscectomy," Med Eng Phys, 51, pp. 41-48.

[46] Hume, D. R., Kefala, V., Harris, M. D., and Shelburne, K. B., 2018, "Comparison of Marker-Based and Stereo Radiography Knee Kinematics in Activities of Daily Living," Ann Biomed Eng, 46(11), pp. 1806-1815.
[47] Nagai, K., Gale, T., Irrgang, J. J., Tashman, S., Fu, F. H., and Anderst, W., 2018, "Anterior Cruciate Ligament Reconstruction Affects Tibiofemoral Joint Congruency During Dynamic Functional Movement," Am J Sports Med, 46(7), pp. 1566-1574.
[48] Englander, Z. A., Martin, J. T., Ganapathy, P. K., Garrett, W. E., and DeFrate, L. E., 2018, "Automatic registration of MRI-based joint models to high-speed biplanar radiographs for precise quantification of in vivo anterior cruciate ligament deformation during gait," J Biomech, 81, pp. 36-44.
[49] Wang, J., and Blackburn, T. J., 2000, "The AAPM/RSNA physics tutorial for residents: X-ray image intensifiers for fluoroscopy," Radiographics, 20(5), pp. 1471-1477.
[50] Bey, M. J., Kline, S. K., Tashman, S., and Zauel, R., 2008, "Accuracy of biplane x-ray imaging combined with model-based tracking for measuring in-vivo patellofemoral joint motion," J Orthop Surg Res, 3, p. 38.
[51] Hofbauer, M., Thorhauer, E. D., Abebe, E., Bey, M., and Tashman, S., 2014, "Altered tibiofemoral kinematics in the affected knee and compensatory changes in the contralateral knee after anterior cruciate ligament reconstruction," Am J Sports Med, 42(11), pp. 2715-2721.
[52] Storti, D., and Yurtoglu, M., 2016, CUDA for engineers : an introduction to high-performance parallel computing, Addison-Wesley, New York.
[53] Torry, M. R., Shelburne, K. B., Peterson, D. S., Giphart, J. E., Krong, J. P., Myers, C., Steadman, J. R., and Woo, S. L., 2011, "Knee kinematic profiles during drop landings: a biplane fluoroscopy study," Med Sci Sports Exerc, 43(3), pp. 533-541.
[54] Pepin, N. R., and Ledoux, W. R., 2016, "Hindfoot kinematics of healthy subjects using lateral wedge insoles in a biplane fluoroscopy system," University of Washington Libraries ,, Seattle, pp. 1 online resource (xiii, 218 unnumbered pages).
[55] Marchelli, G., 2014, "GPU-Accelerated Tools for Medical Image Registration and Biomechanical Modeling," University of Washington.
[56] Tashman, S., Thorhauer, E., Fu, F. H., and Irrgang, J. J., 2016, "Geospatial Analysis of Localized Relationships between Knee Mechanics and Cartilage Changes after ACL Reconstruction," Proceedings of the 2016 Annual Meeting of the Orthopaedic Research Society, p. 56630.
[57] Chen, L., Armstrong, C. W., and Raftopoulos, D. D., 1994, "An investigation on the accuracy of three-dimensional space reconstruction using the direct linear transformation technique," J Biomech, 27(4), pp. 493-500.
[58] Hinrichs, R. N., and McLean, S. P., 1995, "NLT and extrapolated DLT:3-D cinematography alternatives for enlarging the volume of calibration," J Biomech, 28(10), pp. 1219-1223.
[59] Wood, G. A., and Marshall, R. N., 1986, "the Accuracy of Dlt Extrapolation in 3-Dimensional Film Analysis," Journal of Biomechanics, 19, pp. 781-&.
[60] Wood, G. A., and Marshall, R. N., 1986, "The accuracy of DLT extrapolation in three-dimensional film analysis," Journal of Biomechanics, 19.
[61] Aiyangar, A. K., Zheng, L., Tashman, S., Anderst, W. J., and Zhang, X., 2014, "Capturing three-dimensional in vivo lumbar intervertebral joint kinematics using dynamic stereo-X-ray imaging," J Biomech Eng, 136(1), p. 011004.

[62] Anderst, W. J., Donaldson, W. F., Lee, J. Y., and Kang, J. D., 2013, "Cervical spine intervertebral kinematics with respect to the head are different during flexion and extension motions," J Biomech, 46(8), pp. 1471-1475.
[63] Rossmann, K., "Point spread-function, line spread-function, and modulation transfer function. Tools for the study of imaging systems," (0033-8419 (Print)).
[64] Bushberg, J. T., Seibert, J. A., Leidholdt, E. M., Boone, J. M., and Goldschmidt, E. J., 2003, "The Essential Physics of Medical Imaging," Medical Physics, 30, pp. 1936-1936.
[65] Boone, J. M., Arnold, B. A., and Seibert, J. A., 1986, "Characterization of the point spread function and modulation transfer function of scattered radiation using a digital imaging system," Med Phys, 13(2), pp. 254-256.
[66] Richardson, W. H., 1972, "Bayesian-Based Iterative Method of Image Restoration*," J. Opt. Soc. Am., 62(1), pp. 55-59.
[67] Boone, J. M., Seibert, J. A., Barrett, W. A., and Blood, E. A., 1991, "Analysis and correction of imperfections in the image intensifier-TV-digitizer imaging chain," Med Phys, 18(2), pp. 236-242.
[68] Chakraborty, D. P., 1987, "Image intensifier distortion correction," Medical Physics, 14, pp. 249-252.
[69] Pietka, E., and Huang, H. K., 1992, "Correction of aberration in image-intensifier systems," Comput Med Imaging Graph, 16(4), pp. 253-258.
[70] Gronenschild, E., 1997, "The accuracy and reproducibility of a global method to correct for geometric image distortion in the x-ray imaging chain.," Medical physics, 24, pp. 1875-1888.
[71] Cerveri, P., Forlani, C., Borghese, N. A., and Ferrigno, G., 2002, "Distortion correction for x-ray image intensifiers: local unwarping polynomials and RBF neural networks," Med Phys, 29(8), pp. 1759-1771.
[72] Cerveri, P., Forlani, C., Pedotti, a., and Ferrigno, G., 2003, "Hierarchical radial basis function networks and local polynomial un-warping for X-ray image intensifier distortion correction: a comparison with global techniques.," Medical & biological engineering & computing, 41, pp. 151-163.
[73] Soimu, D., Badea, C., and Pallikarakis, N., 2003, "A novel approach for distortion correction for X-ray image intensifiers," Computerized Medical Imaging and Graphics, 27, pp. 79-85.
[74] Holdsworth, D. W., Pollmann, S. I., Nikolov, H. N., and Fahrig, R., 2005, "Correction of XRII geometric distortion using a liquid-filled grid and image subtraction," Med Phys, 32(1), pp. 55-64.
[75] Gutiérrez, L. F., Ozturk, C., McVeigh, E. R., and Lederman, R. J., 2008, "A practical global distortion correction method for an image intensifier based x-ray fluoroscopy system.," Medical Physics, 35, pp. 997-1007.
[76] Rohr, K. S., S. , Sprengel, R. ,Kuhn, M., 2001, "Landmark-based elastic registration using approximating thin-plate splines," IEEE Transactions on Medical Imaging, 20(6), pp. 526-534.
[77] Fahrig, R., Moreau, M., and Holdsworth, D. W., 1997, "Three-dimensional computed tomographic reconstruction using a C-arm mounted XRII: correction of image intensifier distortion," Med Phys, 24(7), pp. 1097-1106.
[78] Kedgley, A. E., Fox, A. M., and Jenkyn, T. R., 2012, "Image intensifier distortion correction for fluoroscopic RSA: the need for independent accuracy assessment," J Appl Clin Med Phys, 13(1), p. 3441.

[79] Guan, S., Gray, H. A., Keynejad, F., and Pandy, M. G., 2016, "Mobile biplane X-Ray imaging system for measuring 3D dynamic joint motion during overground gait," IEEE Transactions on Medical Imaging, 35, pp. 326-336.
[80] Brack, C., Otte, H. G., Goss, F., and Schweikard, A., 1996, "Towards Accurate X-Ray-Camera Calibration in Computer-Assisted Robotic Surgery."
[81] Jaeger, T., 1990, "Methods for rectification of geometric distortion in radioscopic images.," Technical University of Berlin.
[82] Mery, D., 2015, Computer Vision for X-Ray Testing: Imaging, Systems, Image Databases, and Algorithms, Springer International Publishing.
[83] Lichti, D. D., Sharma, G. B., Kuntze, G., Mund, B., Beveridge, J. E., and Ronsky, J. L., 2015, "Rigorous geometric self-calibrating bundle adjustment for a dual fluoroscopic imaging system," IEEE Transactions on Medical Imaging, 34, pp. 589-598.
[84] Richard Hartley, A. Z., 2003, "Multiple View Geometry," Journal of Chemical Information and Modeling, 53, pp. 1689-1699.
[85] Kedgley, A. E., Fox, A. M. V., and Jenkyn, T. R., 2012, "Image intensifier distortion correction for fluoroscopic RSA: The need for independent accuracy assessment," Journal of Applied Clinical Medical Physics, 13, pp. 197-204.
[86] Mery, D., 2003, "Explicit geometric model of a radioscopic imaging system," NDT & E International, 36(8), pp. 587-599.
[87] Marco, A., and Martı´nez, J.-J., 2010, "Polynomial least squares fitting in the Bernstein basis," Linear Algebra and its Applications, 433(7), pp. 1254-1264.
[88] Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R., 2008, "Efficient subpixel image registration algorithms," Opt Lett, 33(2), pp. 156-158.
[89] Knorlein, B. J., Baier, D. B., Gatesy, S. M., Laurence-Chasen, J. D., and Brainerd, E. L., 2016, "Validation of XMALab software for marker-based XROMM," J Exp Biol, 219(Pt 23), pp. 3701-3711.
[90] Zhang, Z., 2000, "A flexible new technique for camera calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), pp. 1330-1334.
[91] Tsai, R., 1987, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IEEE Journal on Robotics and Automation, 3(4), pp. 323-344.
[92] Hartley, R., and Zisserman, A., 2003, "in computervision Multiple View Geometry in Computer Vision," Computer-Aided Design, 16, p. 672.
[93] Alem, N. M., Melvin, J. W., and Holstein, G. L., 1978, "Biomechanics applications of direct linear transformation in close-range photogrammetry," Proceedings of the Sixth New England Bioengineering Conference, D. Jaron, ed., Pergamon, pp. 202-206.
[94] Melen, T., 1994, "Extracting physical camera parameters from the 3x3 direct linear transformation matrix," Optical 3D Measurement Techniques II: Applications in Inspection, Quality Control, and Robotics, SPIE, p. 11.
[95] Choo, A. M., and Oxland, T. R., 2003, "Improved RSA accuracy with DLT and balanced calibration marker distributions with an assessment of initial-calibration," J Biomech, 36(2), pp. 259-264.
[96] Zhao, Z., Ye, D., Zhang, X., Chen, G., and Zhang, B., 2016, "Improved direct linear transformation for parameter decoupling in camera calibration," Algorithms, 9.
[97] Adams, B. M. E., M.S., Bohnoff, W.J., Dalbey, K.R., Eddy, J.P. Hu, K.T., 2015, "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter

Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.3 User's Manual," Sandia National Labs.
[98] Saltelli, A., Ratto, M., Andres, T., Campolongo F., Cariboni, J., Gatelli D., Saisana, M., Tarantola, S., 2007, Global Sensitivity Analysis: The Primer, Wiley.
[99] Renaudin, C., Magnin, I. E., Picard, C., Trousset, Y., Sergent, M., and Amiel, M., 1993, "Image quality study in 3D X-ray angiography: a first approach using the experimental design strategy," Comput Med Imaging Graph, 17(4-5), pp. 309-322.
[100] Rougee, A., Picard, C., Ponchut, C., and Trousset, Y., 1993, "Geometrical calibration of X-ray imaging chains for three-dimensional reconstruction," Comput Med Imaging Graph, 17(4-5), pp. 295-300.
[101] Fischler, M. A., and Bolles, R. C., 1981, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, 24(6), pp. 381–395.
[102] Longuet-Higgins, H. C., 1981, "A computer algorithm for reconstructing a scene from two projections," Nature, 293(5828), pp. 133-135.
[103] Abdellah, M., Eldeib, A., and Owis, M. I., 2015, "GPU acceleration for digitally reconstructed radiographs using bindless texture objects and CUDA/OpenGL interoperability," Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015-Novem, pp. 4242-4245.
[104] Dorgham, O. M., Laycock, S. D., and Fisher, M. H., 2012, "GPU accelerated generation of digitally reconstructed radiographs for 2-D/3-D image registration," IEEE Transactions on Biomedical Engineering, 59, pp. 2594-2603.
[105] Ghafurian, S., Metaxas, D. N., Tan, V., and Li, K., 2016, "Fast generation of digitally reconstructed radiograph through an efficient preprocessing of ray attenuation values," Progress in Biomedical Optics and Imaging - Proceedings of SPIE, 9786, pp. 1-8.
[106] Milickovic, N., Baltas, D., Giannouli, S., Lahanas, M., and Zamboglou, N., 2000, "CT imaging based digitally reconstructed radiographs and their application in brachytherapy," Physics in Medicine and Biology, 45, pp. 2787-2800.
[107] Neuner, M., med Rainer Schubert, U.-P., and Philipp Steininger, B., 2009, "Implementation and Evaluation of DRR algorithms (ray casting, wobbled splatting) for the ITK."
[108] Russakoff, D. B., Rohlfing, T., Mori, K., Rueckert, D., Ho, A., Adler, J. R., and Maurer, C. R., 2005, "Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2D-3D image registration," IEEE Transactions on Medical Imaging, 24, pp. 1441-1454.
[109] Russakoff, D. B., Rohlfing, T., Rueckert, D., Shahidi, R., Kim, D., and Maurer, J., Calvin R., 2003, "Fast calculation of digitally reconstructed radiographs using light fields," Proceedings of SPIE, 5032, pp. 684-695.
[110] Shen, A., and Luo, L., 2008, "Point-based digitally reconstructed radiograph," Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pp. 1-4.
[111] Staub, D., and Murphy, M. J., 2012, "A digitally reconstructed radiograph algorithm calculated from first principles," Medical Physics, 40, p. 011902.
[112] Abdellah, M., Eldeib, A., and Sharawi, A., 2015, "High performance GPU-Based Fourier volume rendering," International Journal of Biomedical Imaging, 2015.

[113] 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 2-D-3-D medical image registration.," IEEE transactions on medical imaging, 17, pp. 586-595.
[114] Brainerd, E. L., Baier, D. B., Gatesy, S. M., Hedrick, T. L., Metzger, K. A., Gilbert, S. L., and Crisco, J. J., 2010, "X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research," J Exp Zool A Ecol Genet Physiol, 313(5), pp. 262-279.
[115] Viola, P., and Wells III, W. M., 1997, "Alignment by Maximization of Mutual Information," International Journal of Computer Vision, 24(2), pp. 137-154.
[116] Russakoff, D. B., Tomasi, C., Rohlfing, T., and Maurer, C. R., 2004, "Image Similarity Using Mutual Information of Regions," Computer Vision - ECCV 2004, T. Pajdla, and J. Matas, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 596-607.
[117] Weese, J., Buzug, T. M., Lorenz, C., and Fassnacht, C., 1997, "An approach to 2D/3D registration of a vertebra in 2D x-ray fluoroscopies with 3D CT images," Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 119-128.
[118] De Silva, T., Uneri, A., Ketcha, M. D., Reaungamornrat, S., Kleinszig, G., Vogt, S., Aygun, N., Lo, S. F., Wolinsky, J. P., and Siewerdsen, J. H., 2016, "3D-2D image registration for target localization in spine surgery: Investigation of similarity metrics providing robustness to content mismatch," Physics in Medicine and Biology, 61, pp. 3009-3025.
[119] Imai, K., Ikoma, K., Kido, M., Maki, M., Fujiwara, H., Arai, Y., Oda, R., Tokunaga, D., Inoue, N., and Kubo, T., 2015, "Joint space width of the tibiotalar joint in the healthy foot," Journal of foot and ankle research, 8, p. 26.
[120] Roberts, L. G., 1969, Machine Perception of Three-Dimensional Solids, U.S. Government Printing Office.
[121] Abdou, I. A., 1979, "Quantitative design and evaluation of enhancement/thresholding edge detectors," IEEE, pp. 753-766.