AN AUTONOMOUS METHOD FOR MEASURING 3D JOINT KINEMATICS FROM 2D XRAY IMAGES

By

ANDREW JAMES JENSEN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTORATE OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2022

Dedication placeholder.

ACKNOWLEDGEMENTS

First, I would like to thank my parents for their continued support throughout my undergraduate and graduate studies. They fostered my work ethic and drive to get through something as ambitious as a doctorate.

Secondly, I would like to thank my advisor, Scott Banks. He has truly been something of a second dad to me as I navigate through graduate school and my future career. Though most conversations seem to revolve around either beer, music, or hiking, we seem to have gotten quite a lot done.

Third, I would like to thank my Fiancee, Lauren. Though busy with medical school, she still takes the time to understand my research and offer support outside of academics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TKA	Total Knee Arthroplasty. This is the complete or partial resurfacing of the articulating surfaces in the knee.
TSA	Total Shoulder Arthroplasty. This is the complete resurfacing of the articulating surfaces in the shoulder.
rTSA	Reverse Total Shoulder Arthroplasty. This is a TSA procedure where the ”ball and socket” mechanism is reversed.
ML	Machine Learning. This is the process of feeding a computer inputs and outputs in order to determine an algorithm that goes from input $\longrightarrow$ output
CNN	Convolutional Neural Network. This is a type of neural network that uses convolution kernels as the operation between each of the layers
HRNet	High Resolution Convolutional Neural Network. This is a specific CNN created by (ADD CITATION) (https://github.com/HRNet)

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctorate of Philosophy

AN AUTONOMOUS METHOD FOR MEASURING 3D JOINT KINEMATICS FROM 2D XRAY IMAGES

By

Andrew James Jensen

December 2022

Chair: Scott Banks
Major: Mechanical Engineering

The primary function of human synovial joints is to support the dynamic motion of the musculoskeletal system. The diseases that typically affect these systems manifest during movement, with mild to severe pain arising during specific activities or during particular ranges of motion. Unsurprisingly, the financial burden of musculoskeletal diseases is roughly USD 300 billion per year in direct and indirect costs [1]. One of the most common conditions affecting human joints is osteoarthritis, which involves the progressive loss of the cartilage between the joint surfaces over time [54]. A highly effective solution for osteoarthritis is arthroplasty, which involves a partial or complete removal and resurfacing of the affected joint with polymeric and metallic components intended to relieve pain and restore a degree of natural function and motion. Despite being highly effective, roughly 20% of patients receiving total knee arthroplasty express some form of dissatisfaction, usually manifested as pain, instability or stiffness during movement ([3], [52], [8]). Surprisingly, standard clinical musculoskeletal diagnostic methods are entirely static. That is, clinicians do not have at their disposal clinically practical ways to quantify skeletal motion during weight-bearing or dynamic movement when most pain symptoms occur. Unfortunately, most of the tools used to accurately quantify 3D dynamic motion (e.g., 3D motion capture, radiostereometry, fluoroscopic model/image registration) are prohibitively expensive or impractical to use in clinical settings. Methods using single-plane fluoroscopic or flat-panel imaging with 3D-to-2D model-image registration have been used since the 1990s. They have

been shown to provide sufficient accuracy for many clinical joint assessment applications , including natural and replaced knees ([4], [6], [39], [71]), natural and replaced shoulders ([41], [43], [70], [42], [30]), and extremities ([67], [36], [14], [13], [59]). One benefit of this approach is that suitable images can be acquired with equipment commonly found in most hospitals. The main impediment for this technology to be used clinically is the time and expense of human operators to supervise the model-image registration process. If the need for human supervision for model-image registration were eliminated, then fluoroscopic imaging could provide a reliable, inexpensive, and accurate method to provide 3D dynamic joint kinematics in a clinical setting. State-of-the-art techniques for generating kinematics using model-image registration involve numerical optimization techniques that iteratively match bone or implant model projections in dynamic x-ray images ([47], [19], [60]). These methods provide accurate 3D bone or implant kinematics when given a rough initial pose estimate for numerical optimization ([19]). However, these methods still require human input for an initial pose estimate, making them impractical for clinical use. Recent advancements in computational capabilities and machine learning algorithms provide tools that are well-suited to replace human supervision for a range of time-consuming tasks including model-image registration. In particular, convolutional neural networks can be trained to provide the image segmentation and pose-estimation capabilities required to autonomously extract knee implant kinematics from single-plane video fluoroscopy. Neural networks can be trained to segment the pixels belonging to a particular knee implant (femoral or tibial), and this pixel information can be used in a numerical optimizer to generate an implant's 3D pose. Alternatively, a neural network can be used directly for pose-regression, using image data as input values and 3D object pose as output. This latter technique relies on the network's ability to extract latent characteristics that determine the pose, not an object-oriented cost function to minimize pose error. This regression approach will be sensitive to study conditions, including implant geometry, projection distance and image size, all of which are "lost" when viewing a single-plane image as only a collection of pixels.

CHAPTER 1
INTRODUCTION

Total Knee Arthroplasty (TKA) is a standard procedure for alleviating symptoms related to osteoarthritis in the knee. In 2018, orthopaedic surgeons performed more than 715,000 TKA operations in the United States [2]. This number is projected to increase to 3.48 million by 2030 [33] due to an aging population and increased obesity rates. While TKA largely relieves symptomatic osteoarthritis, roughly 20% of TKA patients express postoperative dissatisfaction, citing mechanical limitations, pain, and instability as the leading causes [3, 8, 52]. Standard methods of musculoskeletal diagnosis cannot quantify the dynamic state of the joint, either pre- or post-operatively; clinicians must rely on static imaging (radiography, MRI, CT) or qualitative mechanical tests to determine the condition of the affected joint, and these tests cannot easily be performed during weight-bearing or dynamic movement when most pain symptoms occur. Unfortunately, most of the tools used to quantify 3D dynamic motion are substantially affected by soft-tissue artifacts [20, 55, 35], are prohibitively time-consuming or expensive [17], or cannot be performed with equipment available at most hospitals.

Model-image registration is a process where a 3D model is aligned to match an object's projection in an image [9]. Researchers have performed model-image registration using single-plane fluoroscopic or flat-panel imaging since the 1990s. Early methods used pre-computed distance maps [34, 71], or shape libraries [4, 61, 62] to match the projection of a 3D implant model to its projection in a radiographic image. With increasing computational capabilities, methods that iteratively compared implant projections to images were possible [39, 19, 37]. Most model-image registration methods provide sufficient accuracy for clinical joint assessment applications, including natural and replaced knees [6, 5, 31, 10], natural and replaced shoulders [30, 38, 42, 56], and extremities [14, 13, 18]. One of the main benefits of this single-plane approach is that suitable images can be acquired with equipment found in most hospitals. The main impediment to implementing this approach into a standard clinical workflow is the time and expense of human operators to supervise the model-image registration process. These methods require either (1) an initial pose estimate [19, 37], (2) a pre-segmented contour of the implant in the image [9, 34], or (3) a human operator to assist the optimization routine out of local minima [39]. Each of these requirements makes model-image registration methods impractical for clinical use. Even state-of-the-art model-image registration techniques [19] require human initialization or segmentation to perform adequately.

Machine learning algorithms automate the process of analytical model building, utilizing specific algorithms to fit a series of inputs to their respective outputs. Neural networks are a subset of machine learning algorithms that utilize artificial neurons inspired by the human brain's connections [40]. These networks have shown a great deal of success in many computer vision tasks, such as segmentation [15, 63, 51], pose estimation [66, 28], and classification [32, 48, 49]. These capabilities might remove the need for human supervision from TKA model-image registration. Therefore, we propose a three-stage data analysis pipeline where a convolutional neural network (CNN) is used to segment, or identify, the pixels belonging to either a femoral or tibial component. Then, an initial pose estimate is generated comparing the segmented implant contour to a pre-computed shape library. Lastly, the initial pose estimate serves as the starting point for a Lipschitzian optimizer that aligns the contours of a 3D implant model to the contour of the CNN-segmented image.

1.1 Background

1.1.1 Current Ortho Exams

1.1.2 Fluoroscopy

1.1.3 Kinematics from Fluoroscopy

1.2 Model-Image Registration

1.2.1 Geometric Transformations

Geometric primitives, such as points, are fundamental building blocks for representing shapes and objects in computer graphics. In this section, we will discuss how points can be represented in 2D and 3D space..

1.2.1.1 2D and 3D Points

In N-dimensional space, a point is represented as a set of N scalars, each representing a magnitude in a particular direction. This can be represented mathematically as a column vector, as shown in Equation 1-1. In the case of 2D space, a point is represented as a 2D column vector, $\mathbf{x} = [x, y]^T$. Similarly, in 3D space, a point is represented as a 3D column vector, $\mathbf{x} = [x, y, z]^T$..

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix} \in \mathbb{R}^N \tag{1-1}$$

Homogeneous coordinates provide a way to represent points with an additional scale factor, $\tilde{w}$. This allows us to perform rotations and translations simultaneously and successively using matrix multiplications. Homogeneous coordinates for a point in N-dimensional space are represented as a column vector with N+1 elements, as shown in Equation 1-2. The homogeneous coordinates for a 2D point and 3D point can be represented as $\tilde{\mathbf{x}} = \tilde{w}\bar{\mathbf{x}}$, where $\tilde{w}\bar{\mathbf{x}} = [\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_N}, \tilde{w}]^T$. In most model-image registration applications, the scale factor $\tilde{w}$ is set to 1..

$$\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{x_1} \\ \tilde{x_2} \\ \vdots \\ \tilde{x_N} \\ \tilde{w} \end{bmatrix} = \tilde{w} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \tilde{w}\bar{\mathbf{x}} \tag{1-2}$$

1.2.1.2 2D Transformations

Transformations are operations that change the position, orientation, or shape of an object in 2D space. One of the most basic transformations is a translation, which moves an object by adding a displacement vector to its position. This can be expressed mathematically as shown in Equation 1-3:

$$\mathbf{x}' = \mathbf{x} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \mathbf{x} + \mathbf{t}$$

Or, using homogeneous coordinates and matrix multiplication (1-3)

$$= \begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}}$$

Where I is the 2×2 identity matrix

In this equation, $\mathbf{x}$ is the original position of the object, $\mathbf{x}'$ is the transformed position, and $\mathbf{t}$ is the displacement vector that specifies the amount of translation in the x and y directions. The identity matrix $\mathbf{I}$ is used when expressing the transformation using matrix multiplication, as it represents the identity transformation in the x and y dimensions. Using homogeneous coordinates and matrix multiplication allows for the convenient representation of multiple transformations as a single matrix multiplication, as well as for the composition of transformations (i.e., performing multiple transformations in a specific order).

The next type of 2D transformation is a rotation, which changes the orientation of an object, but not its shape (Eq. 1-4, 1-5).

$$\mathbf{x}' = \mathbf{R}\mathbf{x} \tag{1-4}$$

where

$$\mathbf{R} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \tag{1-5}$$

This will rotate an object θ in the counter clockwise direction.

It's also possible to perform a rotation and a translation at the same time, by replacing the identity matrix $\mathbf{I}$ in Equation 1-3 with the rotation matrix $\mathbf{R}$ from Equation 1-5. This results in the transformation shown in Equation 1-6. This transformation preserves lengths and angles.

$$\mathbf{x}' = \begin{bmatrix} \mathbf{R}_{2\times 2} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}} \tag{1-6}$$

A scaled rotation will change the size of the object by some scaler factor, s (Eq. 1-7); this transformation preserves angles.

$$\mathbf{x}' = \begin{bmatrix} s\mathbf{R}_{2\times 2} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}} \tag{1-7}$$

An affine transformation preserves parallelism, and is simply a pre-multiplication by an arbitrary 2×3 matrix (Eq. 1-8).

$$\begin{gathered} \mathbf{x}' = \mathbf{A}\bar{\mathbf{x}} \\ \text{where} \\ \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \end{gathered} \tag{1-8}$$

A projection matrix (or perspective transformation) is one that operates on homogeneous coordinates (Eq. 1-9).

$$\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\tilde{\mathbf{x}} \tag{1-9}$$

To obtain inhomogeneous results, the resultant $\tilde{\mathbf{x}}'$ must be normalized (Eq. 1-10). A projective transformation preserves straight lines.

$$\begin{aligned} x' &= \frac{\tilde{x}}{\tilde{w}} & y' &= \frac{\tilde{y}}{\tilde{w}} \\ &= \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} & &= \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \end{aligned} \tag{1-10}$$

1.2.1.3 3D Transformations

3D transformations operate on 3D points (Eq. 1-1) in a similar way to 2D transformations. In general, all types of transformations, including translations, rotations, scaled rotations, and projections, can be represented by matrices in 3D space, with the only difference being that the dimensions are increased by one.

One aspect of 3D rotations that adds complexity is the possibility of multiple axes of rotation. This introduces the issue of non-commutativity, which means that the order in which rotations are applied matters. To handle this, we can use the concept of Euler angles, which represent the rotations as composite rotations about canonical axes. These rotations are represented by matrices (Eq. 1-11). Note that these matrices are post-multiplied by each other to determine the final rotation matrix for object-centered rotations. Additionally, given a rotation matrix and the corresponding Euler rotations, it is possible to decompose the matrix into its composite rotations about the canonical axes.

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_x & -s_x \\ 0 & s_x & c_x \end{bmatrix} \quad R_y = \begin{bmatrix} c_y & 0 & s_y \\ 0 & 1 & 0 \\ -s_y & 0 & c_y \end{bmatrix} \quad R_z = \begin{bmatrix} c_z & -s_z & 0 \\ s_z & c_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1-11}$$

3D to 2D Projections Geometric primitives, such as points, lines, and polygons, are the building blocks of computer graphics. These basic shapes can be transformed and combined in various ways to create more complex objects and scenes. In order to represent these 3D primitives and objects in 2D image space, we must apply transformations that manipulate their position, orientation, and other properties. By using these transformations, we can create the illusion of depth and spatial relationships on a flat display. Understanding how to work with primitives and transform them is crucial for creating a wide range of visual effects and graphics in computer graphics.

In computer graphics, one of the most basic methods of projecting three-dimensional objects onto a two-dimensional image plane is an orthographic projection. This projection simply drops the depth component of the object and flattens it onto the image plane (Eq. 1-12. This can be thought of as mimicking a camera with a long focal length, or when the depth of the object is shallow compared to its distance from the camera (also known as a weak perspective projection).

In this equation, $\mathbf{p}$ represents a point in 3D space and $\mathbf{x}$ represents the projected point in 2D image space.

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tilde{\mathbf{p}} \tag{1-12}$$

Perspective projection The most common 3D-2D projection is *perspective projeciton*, which mimics how human's normally see by taking depth-perception into account. This can be done by scaling each point by it's z position relative to the camera (Eq. 1-13). We can also perform this using homogeneous coordinates (Eq. 1-14).

$$\bar{\mathbf{x}} = \mathscr{P}_z(\mathbf{p}) = \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix} \tag{1-13}$$

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{\mathbf{p}} \tag{1-14}$$

The perspective projection is the cornerstone of model-image registration. If a matrix can be created that mimics the fluoroscopic imaging setup in the clinic, then this offers a strong possibility of re-creating a virtual scene with the same geometry as the actual scene.

1.2.2 Image Formation and Camera Properties

Using our knowledge of geometric transformations and projective geometries, we can build up a model of a camera step by step. We standardize our reference frame by having the z direction along the focal direction of the camera, the x direction to the right, and the y direction such that the right-hand rule is maintained. The origin is at the center of the camera.

We will describe the object of interest as a collection of points, $\mathbf{p} = [x_i, y_i, z_i, 1]$ for $i = 1, 2, \cdots, N$. For a complete picture, any given operation will be performed on all points. For simplicity, the following equations will demonstrate the process on a single point of the object.

First, we need to describe the location and orientation of the object with respect to the camera. This can be done with a 3D homogeneous transformation matrix (translation and rotation) (Eq. 1-15).

$$\tilde{\mathbf{p}}' = \begin{bmatrix} R_{3\times 3} & \mathbf{t}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \tilde{\mathbf{p}} \tag{1-15}$$

Then, we use a projective transformation that determines the perspective scaling between the objects actual location and the image place at the focal distance. Geometrically, this relationship can be visualized with similar triangles (Fig. 1-1), and quantified using the ratio of lengths (Eq. 1-16). In these equations, f' is the focal distance in units of length.

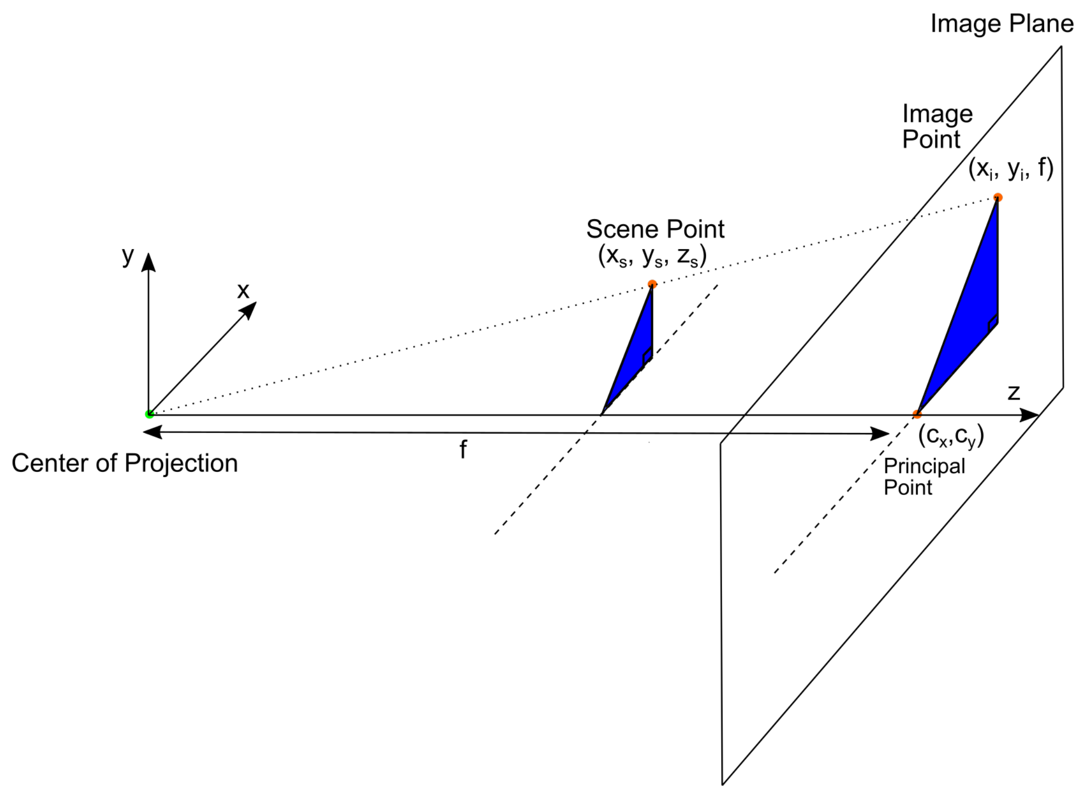

Figure 1-1. The geometry of perspective projection can be visualized by using similar triangles. The overall scaling of the image is based on the ratio of the focal length to the depth of the object.

$$\tilde{\mathbf{x}}_i = \begin{bmatrix} f' & 0 & 0 \\ 0 & f' & 0 \\ 0 & 0 & 1 \end{bmatrix} \tilde{\mathbf{p}}'$$

where

$$x_i = p'_x \frac{f'}{p'_z}$$

$$y_i = p'_y \frac{f'}{p'_z} \tag{1-16}$$

In computer graphics, the principal point, (c_x, c_y) is a key element in the mathematical model of a camera. It is the point where the optical axis of the camera intersects the image plane, and is typically near the center of the image. The principal point is important because it determines the perspective of the image: objects that are farther away from the principal point will appear smaller in the image, while objects that are closer to the principal point will appear larger. When dealing with the shift between the reference frame embedded in the object and the reference frame embedded in the image plane, the principal point acts as a translation (Eq. 1-17).

$$\tilde{\mathbf{x}}_i = \begin{bmatrix} f' & 0 & c_x \\ 0 & f' & c_y \\ 0 & 0 & 1 \end{bmatrix} \tilde{\mathbf{p}}'$$

where

$$c_x \approx \frac{W_{image}}{2}$$

$$c_y \approx \frac{H_{image}}{2} \tag{1-17}$$

Finally, we need to convert the image coordinates, $\tilde{\mathbf{x}}_i$, to pixel coordinates using the pixel scale factor (Eq.1-18). The pixel scale factor is defined by the parameters k_x and k_y, which represent the number of pixels per unit distance in the x and y directions, respectively. By multiplying the perspective projection matrix by the pixel scale factor, we can obtain a new matrix that maps 3D points in world coordinates directly to pixel coordinates on the image.

$$\tilde{\mathbf{x}}_{pix} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f' & 0 & c_x \\ 0 & f' & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1-18}$$

where

$f_x = k_x f'$ and $f_y = k_y f'$

Are focal distances in units of pixels

1.2.3 Image Processing

Digital image processing is a field of computer vision that deals with the manipulation, analysis, and interpretation of digital images. It focuses on algorithms and techniques that extract meaningful information from images and to enhance visual quality. In fluoroscopy, image processing can be used to enhance the digital image, or determine important information about it such that a the model-image registration task can be performed efficiently and autonomously.

1.2.3.1 Filtering and Convolution

We have already seen how image formation yields a collection of 2D points, $\mathbf{x}_{pix}$. We can write the intensity values at each pixel locations as a function, $f(\mathbf{x}_{pix}) = f(i, j)$, where (i, j) represent locations in the image, and the function returns the intensity of the image at that particular pixel location. This allows us to treat images as functions, and perform similar functional operations and analysis to extract meaningful information from them.

The most widely used filter is a linear filter [57], where the output is some linear operation on the neighboring pixels (Eq. 1-19), also known as a *convolution*. In a convolution, the kernel, h is shifted along the input image, f, and the resultant image, g, is the dot product of those two matrices at that specific location.

$$
\begin{aligned}
g(i,j) &= \sum_{k,l} f(i-k, j-l)h(k,l) \\
&= \sum_{k,l} f(k,l)h(i-k, j-l)
\end{aligned}
\tag{1-19}
$$

1.2.4 Image Similarity Metrics

One of the key components in model image registration is image similarity. Fundamentally, this is the method of determining how well the user's synthetic image matches with the actual fluoroscopic image. The choice of similarity metric is going to be determined by many key factors such as the a-priori availability of implant/bone geometry and knowledge of the image quality and contrast. Broadly, there are two classes of image similarity when performing model-image registration: intenisty-based and feature-based.

1.2.4.1 Intensity Based

Intensity based measures are those that utilize specific pixel information in order to determine the difference between two images. This can be either a global image similarity metric, or measure the specfic regions of interest in the given image.

A canonical difference between two images would be the p-norm separating them (Eq. 1-20), which iterates through each pixel of the two images and finds the p-norm difference each intensity for the pixel pair. Common p-norms are the L_1 norm (*absolute intensity differences* or *mean absolute difference*) [26] ($p = 1$) and the L_2, or Euclidean, norm (*squared intensity differences* or *mean squared difference*) [22]($p = 2$).

Intensity-based measures use the pixel values of the images to determine their similarity. These measures can be global, meaning they consider the entire image, or they can focus on specific regions of interest. A common intensity-based measure is the p-norm (Eq. 1-20), which calculates the difference between the intensity of corresponding pixels in the two images. The L_1 norm, also known as the absolute intensity differences or mean absolute difference, uses $p = 1$ in the equation [26], while the L_2 norm, also known as the squared intensity differences or mean squared difference, uses $p = 2$ [22].

$$\|A-B\|_p = \left(\sum_{x=0}^{w}\sum_{y=0}^{h}|a_{xy}-b_{xy}|^p\right)^{\frac{1}{p}} \tag{1-20}$$

where A and B are the two images being compared, w and h are the width and height of the images, and a_{xy} and b_{xy} are the intensity values at pixel (x,y) in the two images, respectively.

While conceptually easy to use, the main limitation of p-norm measures is their lack of spatial information. For example, an image that has been shifted by a linear transformation would not score well using a p-norm, despite the two images containing only a minor shift, scale, or rotation. One method for overcoming this limitation is to use the cross-correlation, or sliding dot product, between images [7, 22] (Eq. 1-21). When used in conjunction with projective geometry, this can help locate regions of interest for a model-based registration pipeline. The cross-correlation is calculated using the following equation:

$$\begin{aligned}(A\star B)[x,y] &= E[A_{xy}\cdot B_{x+\tau_x,y+\tau_y}]\\ &= \sum_{\tau_x=-\infty}^{\infty}\sum_{\tau_y=-\inf}^{\infty} a_{xy}b_{x+\tau_x,y+\tau_y}\end{aligned} \tag{1-21}$$

This will have the effect of determining the regions of each image that are similar, causing the correlation function to “light up” at those areas in a similar way to the convolutional operation between two images. The normalized cross-correlation can also be used (Eq. 1-22), which removes noise coming from each of the original images.

$$\text{normalized cross correlation}(A,B) = \frac{A\star B}{(A\star A)(B\star B)} \tag{1-22}$$

1.2.4.2 Feature Based

Feature based image similarity metrics involve some method of determining key features in images, and using those notable features for measuring the differences between two images. These types of methods almost always involve some type of feature-extraction step, where the various features of interest are calculated and deterimined for subsequent use. The two main classes of features are *keypoints* and *edges*. The simplest method of keypoint detection is using a similar method to intensity-based matching, but having one of the "images" as a patch of the desired feature. With keypoints detected in the input image, one could determine the error of the current pose estimate by taking the Euclidean distance between all image keypoints and all projected keypoints: [10] (Eq. 1-23). With a-priori information about the keypoints, one could attach a weight to every keypoint in order to emphasize specific regions on the image and the model (Eq. 1-24)

$$\text{Keypoint Error} = (\sum_{i=0}^{N} (KP_{image,i} - KP_{proj,i})^2)^{\frac{1}{2}} \tag{1-23}$$

$$\text{Weighted Keypoint Error} = (\sum_{i=0}^{N} w_i (KP_{image,i} - KP_{proj,i})^2)^{\frac{1}{2}} \tag{1-24}$$

Keypoints are particularly useful when there are invariant features in images and 3D models that will always be present. However, if these features will not, or cannot always be deteted, then other measures must be utilized.

Finding Edges

Edge- and contour-based matching algorithms make use of the edges that are present in the image, and aligning that with a the projected edges of the 3D model. However, we must first consider the determination of edges in an images. For a human operator, it can be rather easy to find edges of interest, but how much this be incorporated computationally? The first approach might be in viewing an image topologically, with regions of different colors and intensity represented by different "heights". Then, an edge simply becomes an area with a steep gradient (Eq. 1-25).

$$\mathbf{J}(\mathbf{x}) = \nabla I(\mathbf{x}) = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})(\mathbf{x}) \tag{1-25}$$

Finding the direction of the steepest ascent/descent at any given location will give use the normal to the local edge at that point. However, the derivative operator will accentuate and amplify high frequencies in the image, causing noise to overpower the signal. Removing the high-frequency information (a low-pass filter) in the image results in gradient detection that is much more aligned with the salient edges of the image. The Gaussian kernel is a good option for an isotropic low-pass filter on a 2D signal (image) (Eq. 1-26)

$$
\begin{aligned}
\mathbf{J}_\sigma(\mathbf{x}) &= \nabla[G_\sigma(\mathbf{x} * I(\mathbf{x}))] \\
&= \nabla G_\sigma(\mathbf{x}) * I(\mathbf{x}) \\
&\text{where} \\
\nabla G_\sigma(\mathbf{x}) = (\frac{\partial G_\sigma}{\partial x}, \frac{\partial G_\sigma}{\partial x}) &= [-x - y]\frac{1}{\sigma^2}\exp(\frac{-(x^2+y^2)}{2\sigma^2})
\end{aligned}
\tag{1-26}
$$

The ubiquitous edge detection algorithm was proposed by John Canny in 1986 [12], which utilizes a five-step process. First, a Gaussian kernel is applied as a low-pass filter (Eq. 1-26), second, directional filters are used to find the gradients in each direction of the image, third, a gradient magnitude threshold is applied to remove noise, fourth, a double threshold is applied to remove both strong and weak edges, and lastly, edges are determined from hysteresis.

Using Edges for Image Similarity

In model-image registration, the similarity of two contours is used as a heuristic for the correct pose. When the projected model's contour aligns accurately with the edges in the fluoroscopic image, one can say that the model is *properly registered* to the image. The main question becomes: how can we computationally determine when two contours are aligned?

As always, the simplest approach is to take the p-norm between the model and image contours (Eq. 1-20), where instead of taking the difference between the two original images, one is taking the difference of the edges of the images. This function will be minimized when there is complete overlap between image and model contours. The primary issue with this formulation is the sensitivity to slight perturbations in the model. This is due to the width of the contour being a single pixel, which would render an extremely high error if the model is shifted just a single pixel in any direction. Because the edge-detected images are binary (0-no edge, 1-edge), we can take advantage of binary morphological operations to change the images to better suit the model-image registration pipeline. The primary operation is dilation (Eq. 1-27), which is simply the convolutional operation with the kernel containing all 1s.

$$I_{dil} = (I \circledast g)$$

where

$$g = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \tag{1-27}$$

The dilation operator is useful because it decreases the sensitivity of the p-norm metric for image similarity, allowing for a smoother curve for optimization routines to find a global minima.

1.2.4.3 Symmetry Traps

Objects with rotational or mirror symmetry cause pathological solutions to many of the image similarity metrics when used for optimizing the pose of the object relative to the image. The simplest example of a symmetry trap can be posed as follows: given the shadow of a basketball, which direction was the logo facing? It is quickly apparent that this is an impossible question to answer with just the information given by the image and the 3D model. This problem arises when performing optimizing for the post of mediolaterally symmetric tibial implants. Additional information must be used to find the correct pose of the implant.

However, with the knowledge of the direction of symmetry, it is possible to determine the "dual pose" of the current orientation, that is, the pose that produces indistinguishable projective geometry.

Algorithm for Determining the Dual-Pose of a Symmetric Object

1. Determine the viewing ray from camera $\rightarrow$ object (Eq. 1-28).
2. Determine the axis-angle (m, θ) rotation between the viewing ray and the symmetric axis of the object (Eq. 1-29, 1-30).
3. Rotate the object -2θ about the same axis, reflecting the rotation about the viewing ray

(Eq. **??**).

4. The final orientation of the object is exactly the "dual pose", producing indistinguishable projective geometry (Eq. 1-32).

If T is the homogenous transformation matrix describing the object

$$\vec{v}' = T_{1:3,4} \qquad (1\text{-}28)$$
$$\vec{v} = \frac{\vec{v}'}{\|\vec{v}'\|}$$

We can use trigonometry to determine the angle (Eq. 1-29) and perpindicular axis (Eq. 1-30) between two vectors. For our example, we use the normalized viewing ray and the z-axis (symmetric axis) as the two vectors.

$$cos(\theta) = \vec{v} \cdot \vec{z}$$
$$\rightarrow \qquad (1\text{-}29)$$
$$\theta = arccos(\vec{v} \cdot \vec{z})$$

$$\vec{m} = \frac{\vec{v} \times \vec{z}}{\|\vec{v} \times \vec{z}\|} \qquad (1\text{-}30)$$

Then, we can build a rotation matrix using an axis and an angle [16], (Eq. 1-31).

$$c = cos(-2\theta)$$
$$s = sin(-2\theta)$$
$$q = -cos(-2\theta) \qquad (1\text{-}31)$$
$$R_{3\times3} = \begin{pmatrix} m_x^2 v + c & m_x m_y v - m_z s & m_x m_z v - m_y s \\ m_x m_y v + m_z s & m_y^2 v + c & m_y m_z v - m_x s \\ m_x m_z v - m_y s & m_y m_z v + m_x s & m_z^2 v + c \end{pmatrix}$$

Then, we obtain the final transformation matrix describing the dual pose of the object by a post-multiplication of this rotation matrix.

$$T_{dual} = T_{orig} * \begin{pmatrix} R_{3\times 3} & \vec{0}_{3\times 1} \\ \vec{0}_{1\times 3} & 1 \end{pmatrix} \tag{1-32}$$

Given these two matrices, further exploration can determine which is the correct pose, though this will have to be done using information not directly present in the image contours.

CHAPTER 2
JOINT TRACK MACHINE LEARNING: AN AUTONOMOUS METHOD OF MEASURING 6-DOF TKA KINEMATICS FROM SINGLE-PLANE FLUOROSCOPIC IMAGES

2.1 Introduction

Total Knee Arthroplasty (TKA) is a standard procedure for alleviating symptoms related to osteoarthritis in the knee. In 2018, orthopaedic surgeons performed more than 715,000 TKA operations in the United States [2]. This number is projected to increase to 3.48 million by 2030 [33] due to an aging population and increased obesity rates. While TKA largely relieves symptomatic osteoarthritis, roughly 20% of TKA patients express postoperative dissatisfaction, citing mechanical limitations, pain, and instability as the leading causes [3, 8, 52]. Standard methods of musculoskeletal diagnosis cannot quantify the dynamic state of the joint, either pre- or post-operatively; clinicians must rely on static imaging (radiography, MRI, CT) or qualitative mechanical tests to determine the condition of the affected joint, and these tests cannot easily be performed during weight-bearing or dynamic movement when most pain symptoms occur. Unfortunately, most of the tools used to quantify 3D dynamic motion are substantially affected by soft-tissue artifacts [20, 55, 35], are prohibitively time-consuming or expensive [17], or cannot be performed with equipment available at most hospitals.

Model-image registration is a process where a 3D model is aligned to match an object's projection in an image [9]. Researchers have performed model-image registration using single-plane fluoroscopic or flat-panel imaging since the 1990s. Early methods used pre-computed distance maps [34, 71], or shape libraries [4, 61, 62] to match the projection of a 3D implant model to its projection in a radiographic image. With increasing computational capabilities, methods that iteratively compared implant projections to images were possible [39, 19, 37]. Most model-image registration methods provide sufficient accuracy for clinical joint assessment applications, including natural and replaced knees [6, 5, 31, 10], natural and replaced shoulders [30, 38, 42, 56], and extremities [14, 13, 18]. One of the main benefits of this single-plane approach is that suitable images can be acquired with equipment found in most hospitals. The main impediment to implementing this approach into a standard clinical workflow is the time and expense of human operators to supervise the model-image registration process. These methods require either (1) an initial pose estimate [19, 37], (2) a pre-segmented contour of the implant in the image [9, 34], or (3) a human operator to assist the optimization routine out of local minima [39]. Each of these requirements makes model-image registration methods impractical for clinical use. Even state-of-the-art model-image registration techniques [19] require human initialization or segmentation to perform adequately.

Machine learning algorithms automate the process of analytical model building, utilizing specific algorithms to fit a series of inputs to their respective outputs. Neural networks are a subset of machine learning algorithms that utilize artificial neurons inspired by the human brain's connections [40]. These networks have shown a great deal of success in many computer vision tasks, such as segmentation [15, 63, 51], pose estimation [66, 28], and classification [32, 48, 49]. These capabilities might remove the need for human supervision from TKA model-image registration. Therefore, we propose a three-stage data analysis pipeline (Fig. 2-1) where a convolutional neural network (CNN) is used to segment, or identify, the pixels belonging to either a femoral or tibial component. Then, an initial pose estimate is generated comparing the segmented implant contour to a pre-computed shape library. Lastly, the initial pose estimate serves as the starting point for a Lipschitzian optimizer that aligns the contours of a 3D implant model to the contour of the CNN-segmented image.

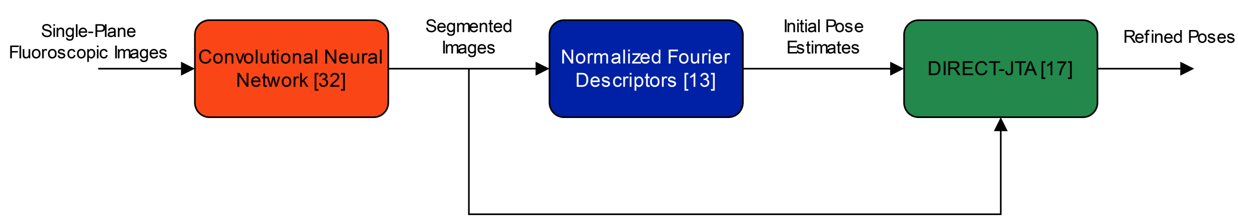

Figure 2-1. An overview of the pipeline for autonomous measurements of total knee arthroplasty kinematics. First, the data is processed through a convolutional neural network to locate the pixels belonging to the femoral and tibial implants [63], then, Normalized Fourier Descriptor shape libraries are used to determine and initial pose estimate [4], and lastly, DIRECT-JTA [19] is run on those segmented images using the NFD estimates as initializations for pose.

This paper seeks to answer the following three questions: (1) How well does a convolutional neural network segment the femoral and tibial implants from fluoroscopic and flat-panel images? (2) Can a Fourier descriptor-based pose estimation method produce useful initial guesses of 3D implant pose from the CNN-segmented images? (3) Can the Lipschitzian optimizer, given reasonable initial guesses, replicate human-supervised TKA kinematic measurements?

2.2 Methods

Data from seven previously reported TKA kinematics studies were used for this study [27, 46, 45, 64, 25, 65, 53]. These studies utilized single-plane fluoroscopy or flat-panel imaging to measure tibiofemoral implant kinematics during lunge, squat, kneel, and stair climbing movements from 8248 images in 71 patients with implants from 7 manufacturers, including 36 distinct implants. From each of these studies, the following information was collected: (1) deidentified radiographic images, (2) x-ray calibration files, (3) manufacturer-supplied tibial and femoral implant surface geometry files (STL format), and (4) human supervised kinematics for the tibial and femoral components in each of the images. CNNs were trained with images from six of the studies using a transfer-learning paradigm with an open-source network [63]. CNN performance was tested using two image collections: a standard test set including images from the six studies used for training and a wholly naïve test set using images from the seventh study, where the imaging equipment and implants were different from anything used in training (Fig. 2-2). We used both test image sets to compare human-supervised kinematics with autonomously measured kinematics. Separately, two independent groups utilized our software to assess the accuracy of TKA kinematics measurements compared to their previously reported reference standard systems using RSA [58] or motion capture [17].

2.2.1 Image Segmentation

Images were resized and padded to 1024x1024 pixels. Images containing bilateral implants had the contralateral knee cropped from the image. Segmentation labels were created by taking the human-supervised kinematics for each implant and generating a flat-shaded ground-truth projection image (Fig. 2-3). Two neural networks [63] were trained to segment the tibial and femoral implants, respectively, from the x-ray images. Each network was trained using a random

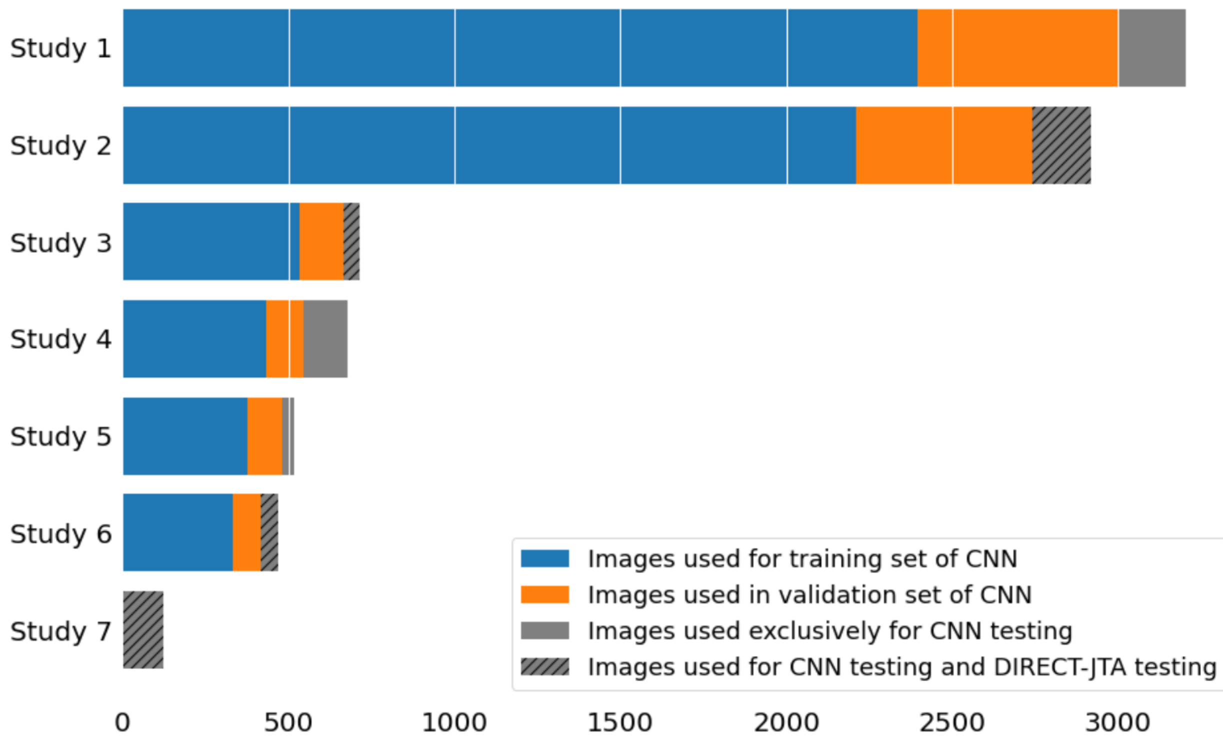

Figure 2-2. Data from seven studies were used to train and test the TKA kinematics measurement pipeline. Color coding in the figure identifies how many images were used for the training, validation, and testing functions. Images from the seventh study were used exclusively for testing the measurement pipeline that was trained using images from the other six studies.

6284/1572 (80/20) training/validation split. Augmentations were introduced in the training pipeline to improve the network's generalization to new implants and implant types [11]. Each neural network was trained on an NVIDIA A100 GPU for 30 epochs. The performance of the segmentation networks was measured using the Jaccard Index [24]. This calculates the intersection between the estimated and ground-truth pixels over the union of both sets of pixels. The ideal Jaccard index is 1.

2.2.2 Initial Pose Estimates

Initial pose estimates were generated from bounding contours of the CNN-segmented implant regions using Normalized Fourier Descriptor (NFD) shape libraries [4, 61, 62]. Shape libraries were created by projecting 3D implant models using the corresponding x-ray calibration parameters with ±30° ranges for the out-of-plane rotations at 3° increments (Fig. 2-4). Pose

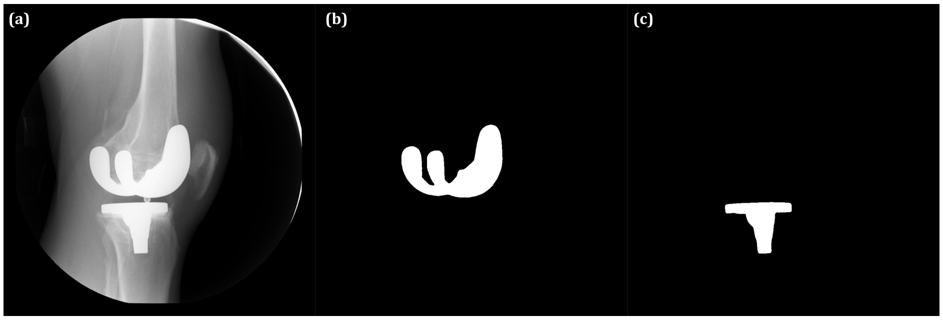

Figure 2-3. A representative fluoroscopic images is shown (a) with corresponding femoral (b) and tibial (c) ground-truth images created by flat-shaded projections of registered implant models.

estimates were determined as previously described [4] NFD-derived femoral and tibial implant poses were transformed to anatomic joint angles and translations [21] and compared to the human-supervised kinematics for the same images using RMS differences for each joint pose parameter. The performance of this method was also assessed using flat-shaded projection images with perfect segmentation as a ground-truth reference standard.

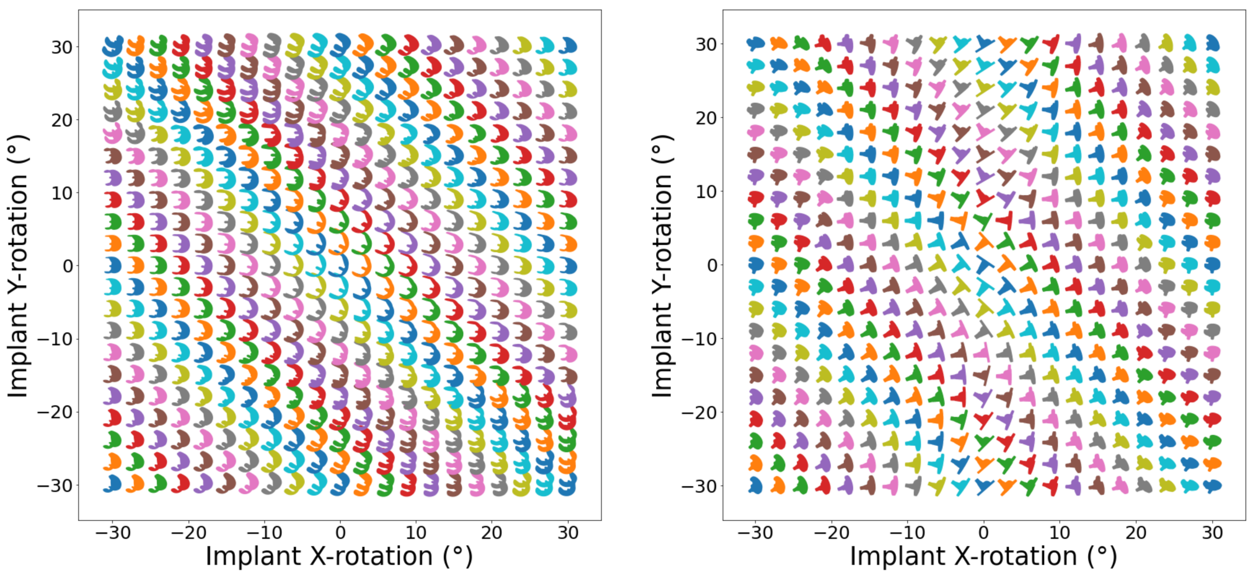

Figure 2-4. Femoral (left) and tibial (right) NFD shape libraries were generated to capture the variation in projection silhouette geometry with out-of-plane rotation [4]. Initial pose estimates were generated by comparing the NFD contour from the x-ray image to the shape library.

2.2.3 Pose Refinement

A modified Dividing Rectangles (DIRECT) algorithm called DIRECT-JTA [19] generated the final pose estimates. This method of Lipschitzian optimization divides the search into three stages, the “trunk,” “branch,” and “leaf.” Each of the three stages was assigned distinct cost function parameters and search regions. The cost function used a computationally efficient L1-norm between the dilated contour from the segmentation label and the projected implant. Successively decreasing the dilation coefficient allowed the optimization routine to escape local minima, and the leaf branch served to find the optimal out-of-plane translation. Transversely symmetric tibial implants posed problems during registration because two distinct poses produced roughly identical projections [29]. Because of this pose ambiguity, the tibial implant was always optimized after the non-symmetric femoral implant. In addition to the dilation metric, the tibial mediolateral translation and varus/valgus rotations relative to the femur were penalized. Final implant poses were transformed into knee joint rotations and translations [21] and compared to the human-supervised kinematics for the same images using RMS differences for each joint pose parameter. Squared differences between data sets were compared using one-way MANOVA with post-hoc multiple pair-wise comparisons using the Games-Howell test (R v4.2.0 using R Studio, rstatix, and stats).

2.2.4 Pose Ambiguities and Registration Blunders

A blunder was defined as an image frame with the squared sum of rotation differences greater than 5° between autonomous and human-supervised measures. These blunder frames contain errors considerably larger than would be clinically acceptable and warrant further exploration. Blunders were analyzed with respect to the tibial implant’s apparent varus/valgus rotation relative to the viewing ray (Fig. 2-5). A probability density function and cumulative density function were calculated for the blunder likelihood. Due to the high likelihood of blunders in this region, an ambiguous zone was defined for all apparent tibial varus/valgus-rotation less than 3.6 degrees, which is the mean + 1std of the blunder distribution (Fig. 2-5). Squared measurement differences between images inside and outside the ambiguous zone were also compared using one-way MANOVA with post-hoc multiple pair-wise comparisons using the Games-Howell test.

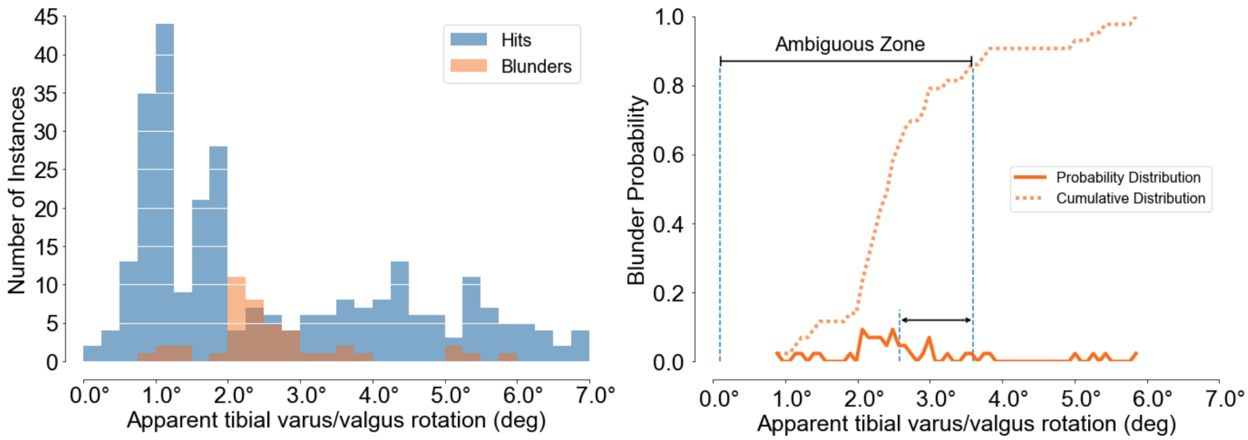

Figure 2-5. The histogram (left) shows the correctly registered frames (Hits, blue) and incorrectly registered frames (Blunders, orange) plotted as a function of the apparent tibial varus/valgus angle relative to the viewing raw. The probability plot (right) shows the distribution of blunders (solid orange) and the cumulative probability of blunders (dotted orange). The Ambiguous Zone is defined as apparent tibial varus/valgus rotations less than the mean + one standard deviation of the blunder probability distribution, capturing approximately 85 % of the blunders.

2.3 Results

CNN segmentation of standard test set images produced Jaccard indices of 0.936 for the femoral and 0.883 for the tibial components. CNN segmentation performance on the completely naïve test set was lower, 0.715 and 0.753, respectively.

The initial pose estimates were within the range of convergence for the DIRECT-JTA optimizer and offered a robust initialization for optimization (Table 1). The RMS differences for initial pose estimates on ground-truth images were smaller (better) than for CNN-segmented images, but the differences were mostly within a few millimeters or degrees. Due to poor sensitivity for measuring out-of-plane translation with monocular vision, the mediolateral translation had the largest RMS differences for both image types.

RMS differences between DIRECT-JTA optimized kinematics and human-supervised kinematics were sub-millimeters for all in-plane translations (Table II). Mediolateral translations and out-of-plane rotation differences were smaller when the pose of the tibia was outside the

Table I

RMS Differences Between NFD Initial Estimates and Human-Supervised Kinematics

Implant	Images	Translation (mm) x	y	z	Rotation (deg) z	x	y
Femoral	CNN-Segmented Images	2.37	0.71	17.59	2.54	2.45	4.75
	Ground-Truth Projections	2.06	0.57	13.53	0.85	1.42	4.00
Tibial	CNN-Segmented Images	2.06	1.49	29.93	0.94	5.59	9.47
	Ground-Truth Projections	2.05	0.87	14.60	0.55	4.73	6.23

ambiguous zone. The RMS differences for the completely naïve test set were within 0.5 mm or 0.5 deg compared to the standard test set, indicating similar performance on the entirely novel dataset.

Table II

RMS Differences Between DIRECT-JTA Optimized and Human-Supervised Kinematics

Test Set	Image Group	Number of Images	A/P (mm)	S/I (mm)	M/L (mm)	Flx/Ext (°)	I/E (°)	V/V (°)
Standard	Inside AZ	187	0.694	0.523[b]	1.752[a]	0.730[a]	3.380	1.938[a]
	Outside AZ	83	0.685	0.466[c]	0.917	1.029	1.811	0.605
Naïve	Inside AZ	47	0.802	0.739	1.715[d]	1.388	4.044	2.480[d]
	Outside AZ	75	0.692	0.644	0.691	1.031	1.154	0.846

AZ = Ambiguous Zone
Superscripts denote pairwise differences ($p<0.05$) in squared errors for:

a. Standard Inside AZ vs Standard Outside AZ
b. Standard Inside AZ vs Naïve Inside AZ
c. Standard Outside AZ vs Naïve Outside AZ
d. Naïve Inside AZ vs Naïve Outside AZ

There was one femoral blunder and 43 tibial blunders out of 392 test images. Using the definition of the ambiguous zone as apparent tibial varus/valgus rotation less than 3.6 deg, 11% of images have a tibial blunder within this zone, compared to 3.2% outside. Sixty-six percent of tibial blunders were due to symmetry ambiguities (Fig 2-6).

One-hundred thirteen image pairs from an RSA study of TKA were used to independently assess the accuracy of the autonomous kinematics measurement for single-plane lateral TKA images. RMS errors were 0.8mm for AP translation, 0.5mm for SI translation, 2.6mm for ML translation, 1.0° for flexion-extension, 1.2° for abduction-adduction, and 1.7° for internal-external rotation. At a different institution, 45 single-plane radiographic images were acquired with an instrumented sawbones phantom that was independently tracked using motion capture.

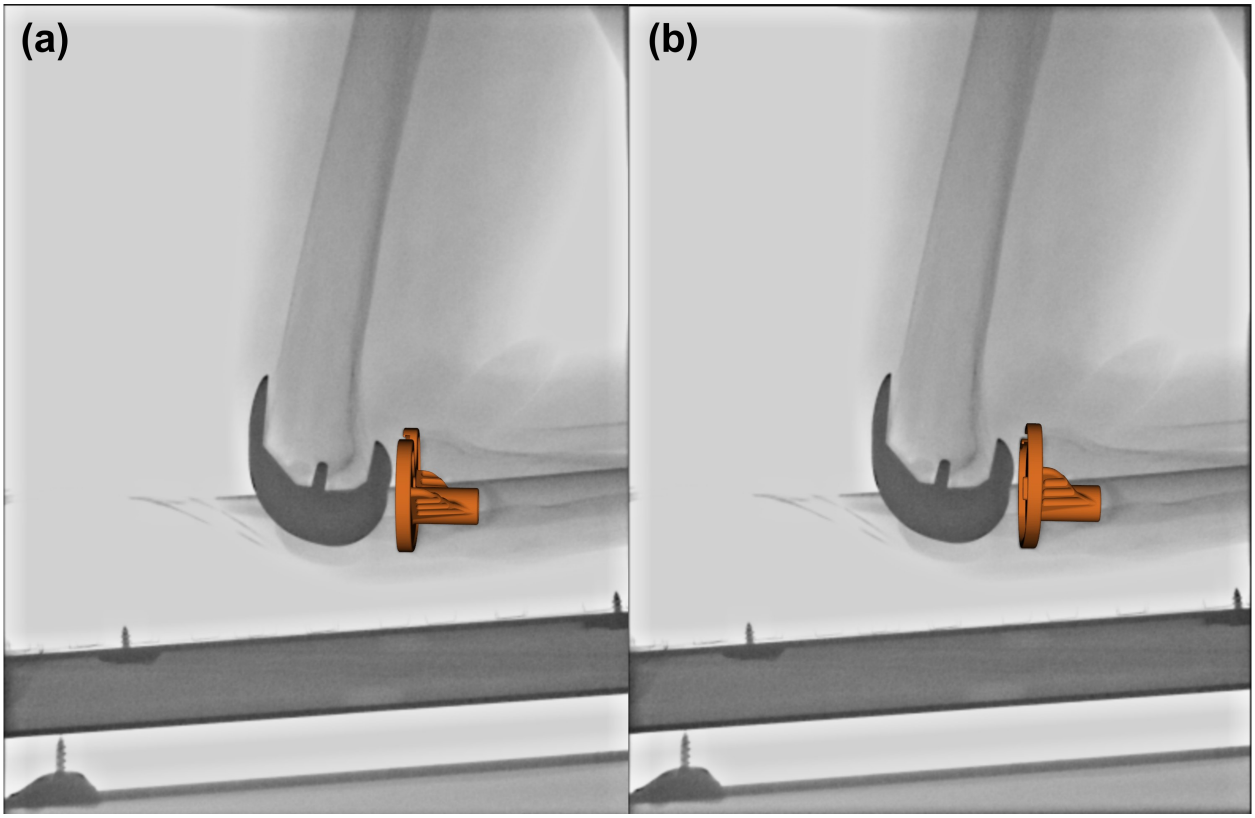

Figure 2-6. The figure shows the same radiographic image with two registered tibial implant poses: (a) shows a correctly registered tibial implant, while (b) shows an implant caught in a local cost function minimum corresponding to a nearly symmetric pose.

Comparing the motion capture and autonomously measured radiographic kinematics, the RMS errors were 0.72mm for AP translation, 0.31mm for SI translation, 1.82mm for ML translation, 0.56° for flexion-extension, 0.63° for abduction-adduction, and 0.84° for internal-external rotation.

2.4 Discussion

Dynamic radiographic measurement of 3D TKA kinematics has provided important information for implant design and surgical technique for over 30 years. Many surgeons have expressed an interest in utilizing this type of measurement in their clinical practices; however, current methods are impractical. We developed a completely autonomous TKA kinematics measurement pipeline that can potentially provide a practical method for clinical implementation. This study sought to answer three questions, (1) How well does a neural network segment TKA implants from fluoroscopic and flat-panel images? (2) How well can an NFD shape library estimate the pose of a TKA implant given a CNN-segmented image? And (3) How well does a Lipschitzian optimization routine replicate human-supervised kinematics for TKA implants given an approximate initial guess?

CNN image segmentation of TKA implants worked well, with Jaccard indices greater than 0.88 for the standard test set, and greater than 0.71 for the naïve test set. Segmentation performance for the standard test set outperformed published examples by 0.05-0.1 Jaccard points [69, 50], with the naïve test set on par with other segmentation examples. The most notable decrease in segmentation performance occurred along the perimeter of the segmented pixel region, especially in areas where implant projections occluded each other. These imperfectly segmented perimeter regions likely affect the initial pose estimate and the DIRECT-JTA optimization solution since both methods rely heavily on the segmented implant boundary. Further improvements can be made for the perimeter segmentation results by introducing intelligent augmentations during training using generative models [23] and performing neural network bolstered contour improvement strategies [68].

Our initial pose estimates were satisfactory as an initialization for the DIRECT-JTA optimization, falling within the convergence region of ±30° [19]. However, the performance for the ground-truth projections was not as good as the cited method [4], which achieved errors of less than 1mm for in-plane translation and 2° for rotation. The cited method utilized an additional refinement step for the NFD estimation, interpolating the apparent out-of-plane angles between nearest shapes in the library. This extra step was not done because only approximate initial pose estimates were needed. In addition, the current study incorporated a vastly larger set of implant shapes (36 vs. 2) and image quality and calibration variations. Distinct implant shapes manifest unique normalization maps, where there can be discontinuities or jumps in normalization angles which affect the best-fitting library entry (Fig. 2-4) [61, 62]. These details are easily upgraded with additional code using previously reported methods but were not pursued because the initial pose results were well within the DIRECT-JTA convergence region. The initial pose estimates for the CNN-segmented images were not as good as for the ground-truth projections. This follows directly from the fact that the perimeter of the segmented implants was not as accurately rendered, leading to poorer results with the edge-based NFD method. Finally, the out-of-plane translation estimates were relatively poor for both ground-truth projects and CNN-segmented images. This translation estimate is extremely sensitive to model projection and edge detection details and can be adjusted for better results if required.

RMS differences between human-supervised and DIRECT-JTA optimized kinematics demonstrate the two methods provide similar results. In-plane translation differences of less than 0.8mm and out-of-plane less than 1.8 mm, indicate good consistency in determining the relative locations of TKA implants. Rotation differences of 4° or less for frames within the ambiguous zone, and less than 1.7° for frames outside the ambiguous zone, indicate joint rotation measures with sufficient resolution to be clinically useful. We observed two important characteristics in the measurement comparisons that will affect future implementations and use. First, we identified an ambiguous zone of apparent tibial rotations wherein there is a higher incidence of registration errors. These errors resulted in significant differences in measurement performance for the out-of-plane translations and rotations. This phenomenon, resulting from the nearly symmetric nature of most tibial implants [34, 71, 4, 39, 19] prompts either practical modification to imaging protocols to bias the tibial view outside the ambiguous zone or modifications of the model-image registration code to enforce smooth kinematic continuity across image frames and/or to impose joint penetration/separation penalties [44]. Second, we observed similar measurement performance for the standard and naïve test sets, which differed only in the superior/inferior joint translation. This suggests that the autonomous kinematic processing pipeline can provide reliable measures for implants and imaging systems that were not part of the training set, which will be important for application in novel clinical environments.

Two independent research teams utilized our software to evaluate the accuracy of our autonomous measurement pipeline compared to their reference standard methods using implants and image detectors that were not part of our training sets. In both cases, the accuracy results were comparable to results reported for contemporary human-supervised single-plane model-image registration methods for TKA kinematics [4, 19, 6, 5, 31]. Interestingly, the independent accuracy results appeared superior to our assessment of differences between autonomous and human-supervised measures of TKA kinematics. In both cases, the independent centers used high-resolution flat-panel detectors that provided better spatial resolution and grayscale contrast than most of the imaging systems included in our datasets. With images of similar quality, it is reasonable to expect similar measurement accuracy.

This work has several limitations. First, the image data sets resulted from previous studies in our labs, so there was no prospective design of which implant systems and image detectors should be included for a pipeline that generalizes well to other implants and detectors. Nevertheless, the naïve data set and the independent assessments, all involving implants and detectors not used for training, performed well and suggest that the method can usefully generalize to measurements of traditionally configured TKA implants. Future work is required to evaluate measurement performance with partial knee arthroplasty or revision implants. Second, many methodologic and configuration options and alternatives remain to be explored, and the current pipeline implementation should not be considered optimal. How best to disambiguate tibial poses and determine the most effective and robust optimization cost functions are areas of current effort.

We present an autonomous pipeline for measuring 3D TKA kinematics from single-plane radiographic images. Measurement reproducibility and accuracy are comparable to contemporary human-supervised methods. We believe capabilities like this will soon make it practical to perform dynamic TKA kinematic analysis in a clinical workflow, where these measures can help surgeons objectively determine the best course of treatment for their patients.

LIST OF REFERENCES

[1] *BMUS: The Burden of Musculoskeletal Diseases in the United States*, https://www.boneandjointburden.org/.

[2] Agency for Healthcare Research and Quality, *HCUP Fast Stats*, https://hcup-us.ahrq.gov/faststats/NationalProcedures.

[3] P. N. Baker, J. H. van der Meulen, J. Lewsey, and P. J. Gregg, *The Role of Pain and Function in Determining Patient Satisfaction After Total Knee Replacement: Data From the National Joint Registry for England and Wales*, The Journal of Bone and Joint Surgery. British volume **89-B** (2007), no. 7, 893–900.

[4] S.A. Banks and W.A. Hodge, *Accurate measurement of three-dimensional knee replacement kinematics using single-plane fluoroscopy*, IEEE Transactions on Biomedical Engineering **43** (1996), no. 6, 638–649.

[5] Scott A. Banks and W.Andrew Hodge, *2003 Hap Paul Award paper of the International Society for Technology in Arthroplasty*, The Journal of Arthroplasty **19** (2004), no. 7, 809–816.

[6] Scott A. Banks, George D. Markovich, and W.Andrew Hodge, *In vivo kinematics of cruciate-retaining and -substituting knee arthroplasties*, The Journal of Arthroplasty **12** (1997), no. 3, 297–304.

[7] Julius S. Bendat and Allan G. Piersol, *Random data: Analysis and measurement procedures*, 4th ed ed., Wiley Series in Probability and Statistics, Wiley, Hoboken, N.J, 2010.

[8] Robert B. Bourne, Bert M. Chesworth, Aileen M. Davis, Nizar N. Mahomed, and Kory D. J. Charron, *Patient Satisfaction after Total Knee Arthroplasty: Who is Satisfied and Who is Not?*, Clinical Orthopaedics & Related Research **468** (2010), no. 1, 57–63.

[9] Lisa Gottesfeld Brown, *A survey of image registration techniques*, ACM Computing Surveys **24** (1992), no. 4, 325–376.

[10] William Burton, Andrew Jensen, Casey A. Myers, Landon Hamilton, Kevin B. Shelburne, Scott A. Banks, and Paul J. Rullkoetter, *Automatic tracking of healthy joint kinematics from stereo-radiography sequences.*, Computers in Biology and Medicine (2021).

[11] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin, *Albumentations: Fast and Flexible Image Augmentations*, Information **11** (2020), no. 2, 125.

[12] John Canny, *A Computational Approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence (1986).

[13] F. Cenni, A. Leardini, M. Pieri, L. Berti, C. Belvedere, M. Romagnoli, and S. Giannini, *Functional performance of a total ankle replacement: Thorough assessment by combining gait and fluoroscopic analyses*, Clinical Biomechanics **28** (2013), no. 1, 79–87.

[14] Francesco Cenni, Alberto Leardini, Claudio Belvedere, Francesca Buganè, Karin Cremonini, Maria T. Miscione, and Sandro Giannini, *Kinematics of the Three Components of a Total Ankle Replacement:* In Vivo *Fluoroscopic Analysis*, Foot & Ankle International **33** (2012), no. 4, 290–300.

[15] Lyndon Chan, Mahdi Hosseini, Corwyn Rowsell, Konstantinos Plataniotis, and Savvas Damaskinos, *HistoSegNet: Semantic Segmentation of Histological Tissue Type in Whole Slide Images*, 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (Seoul, Korea (South)), IEEE, October 2019, pp. 10661–10670.

[16] Carl D. Crane and Joseph Duffy, *Kinematic analysis of robot manipulators*, digitally printed version ed., Cambridge University Press, Cambridge, U.K., 2008.

[17] R. Daems, Jan Victor, Patrick De Baets, S. Van Onsem, and Matthias Verstraete, *Validation of three-dimensional total knee replacement kinematics measurement using single-plane fluoroscopy*, International Journal Sustainable Construction & Design **7** (2016), no. 1, 14.

[18] Richard J. de Asla, Lu Wan, Harry E. Rubash, and Guoan Li, *Six DOF in vivo kinematics of the ankle joint complex: Application of a combined dual-orthogonal fluoroscopic and magnetic resonance imaging technique*, Journal of Orthopaedic Research **24** (2006), no. 5, 1019–1027.

[19] P. D. L. Flood and Scott A. Banks, *Automated registration of 3-D knee implant models to fluoroscopic images using lipschitzian optimization*, IEEE Transactions on Medical Imaging **37** (2018), no. 1, 326–335.

[20] Bo Gao and Naiquan (Nigel) Zheng, *Investigation of soft tissue movement during level walking: Translations and rotations of skin markers*, Journal of Biomechanics **41** (2008), no. 15, 3189–3195.

[21] Edward S. Grood and W. J. Suntay, *A Joint Coordinate System for the Clinical Description of Three-Dimensional Motions: Application to the Knee*, Journal of Biomechanical Engineering-transactions of The Asme (1983).

[22] Marsha Jo Hannah, *Computer Matching of Areas in Stereo IMages*, Ph.D. thesis, Stanford University, 1977.

[23] Ryuichiro Hataya, Jan Zdenek, Kazuki Yoshizoe, and Hideki Nakayama, *Faster AutoAugment: Learning Augmentation Strategies using Backpropagation*, arXiv:1911.06987 [cs] (2019).

[24] Paul Jaccard, *The Distribution of the Flora in the Alpine Zone*, New Phytologist **11** (1912), no. 2, 37–50.

[25] Jean-Yves Jenny, Scott Banks, and Florent Baldairon, *Registration of Knee Kinematics With a Navigation System: A Validation Study*, 2015.

[26] T. Kanade and M. Okutomi, *A stereo matching algorithm with an adaptive window: Theory and experiment*, IEEE Transactions on Pattern Analysis and Machine Intelligence **16** (Sept./1994), no. 9, 920–932.

[27] Vasiliki Kefala, Adam J. Cyr, Michael D. Harris, Donald R. Hume, Bradley S. Davidson, Raymond H. Kim, and Kevin B. Shelburne, *Assessment of Knee Kinematics in Older Adults Using High-Speed Stereo Radiography*, Medicine & Science in Sports & Exercise **49** (2017), no. 11, 2260–2267.

[28] Alex Kendall and Roberto Cipolla, *Geometric Loss Functions for Camera Pose Regression with Deep Learning*, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI), IEEE, July 2017, pp. 6555–6564.

[29] David G. Kendall, *Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces*, Bulletin of the London Mathematical Society **16** (1984), no. 2, 81–121.

[30] Takehiro Kijima, Keisuke Matsuki, Nobuyasu Ochiai, Takeshi Yamaguchi, Yu Sasaki, Eiko Hashimoto, Yasuhito Sasaki, Hironori Yamazaki, Tomonori Kenmoku, Satoshi Yamaguchi, Yoshitada Masuda, Hideo Umekita, Scott A. Banks, and Kazuhisa Takahashi, *In vivo 3-dimensional analysis of scapular and glenohumeral kinematics: Comparison of symptomatic or asymptomatic shoulders with rotator cuff tears and healthy shoulders*, Journal of Shoulder and Elbow Surgery **24** (2015), no. 11, 1817–1826.

[31] Richard D. Komistek, Douglas A. Dennis, and Mohamed Mahfouz, *In Vivo Fluoroscopic Analysis of the Normal Human Knee*, Clinical Orthopaedics & Related Research **410** (2003), 69–81.

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, Communications of the ACM **60** (2017), no. 6, 84–90.

[33] Steven Kurtz, Kevin Ong, Edmund Lau, Fionna Mowat, and Michael Halpern, *Projections of Primary and Revision Hip and Knee Arthroplasty in the United States from 2005 to 2030:*, The Journal of Bone & Joint Surgery **89** (2007), no. 4, 780–785.

[34] S. Lavallee and R. Szeliski, *Recovering the position and orientation of free-form objects from image contours using 3D distance maps*, IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995), no. 4, 378–390.

[35] Cheng-Chung Lin, Tung-Wu Lu, Hsuan-Lun Lu, Mei-Ying Kuo, and Horng-Chaung Hsu, *Effects of soft tissue artifacts on differentiating kinematic differences between natural and replaced knee joints during functional activity*, Gait & Posture **46** (2016), 154–160.

[36] Renate List, Mauro Foresti, Hans Gerber, Jörg Goldhahn, Pascal Rippstein, and Edgar Stüssi, *Three-Dimensional Kinematics of an Unconstrained Ankle Arthroplasty: A Preliminary* In Vivo *Videofluoroscopic Feasibility Study*, Foot & Ankle International **33** (2012), no. 10, 883–892.

[37] David G. Lowe, *Fitting parameterized three-dimensional models to images*, IEEE Transactions on Pattern Analysis and Machine Intelligence (1991).

[38] Mohamed Mahfouz, Gregory Nicholson, Richard Komistek, David Hovis, and Matthew Kubo, *In Vivo Determination of the Dynamics of Normal, Rotator Cuff-Deficient, Total, and Reverse Replacement Shoulders*, VO LU M E, 8.

[39] M.R. Mahfouz, W.A. Hoff, R.D. Komistek, and D.A. Dennis, *A robust method for registration of three-dimensional knee implant models to two-dimensional fluoroscopy images*, IEEE Transactions on Medical Imaging **22** (2003), no. 12, 1561–1574.

[40] D. Marr, *Early processing of visual information*, (1976).

[41] Keisuke Matsuki, Kei O. Matsuki, Shang Mu, Tomonori Kenmoku, Satoshi Yamaguchi, Nobuyasu Ochiai, Takahisa Sasho, Hiroyuki Sugaya, Tomoaki Toyone, Yuichi Wada, Kazuhisa Takahashi, and Scott A. Banks, *In vivo 3D analysis of clavicular kinematics during scapular plane abduction: Comparison of dominant and non-dominant shoulders*, Gait & Posture **39** (2014), no. 1, 625–627.

[42] Keisuke Matsuki, Kei O. Matsuki, Shang Mu, Satoshi Yamaguchi, Nobuyasu Ochiai, Takahisa Sasho, Hiroyuki Sugaya, Tomoaki Toyone, Yuichi Wada, Kazuhisa Takahashi, and Scott A. Banks, *In vivo 3-dimensional analysis of scapular kinematics: Comparison of dominant and nondominant shoulders*, Journal of Shoulder and Elbow Surgery **20** (2011), no. 4, 659–665.

[43] Keisuke Matsuki, Kei O. Matsuki, Satoshi Yamaguchi, Nobuyasu Ochiai, Takahisa Sasho, Hiroyuki Sugaya, Tomoaki Toyone, Yuichi Wada, Kazuhisa Takahashi, and Scott A. Banks, *Dynamic In Vivo Glenohumeral Kinematics During Scapular Plane Abduction in Healthy Shoulders*, Journal of Orthopaedic & Sports Physical Therapy **42** (2012), no. 2, 96–104.

[44] Shang Mu, *JointTrack: An Open-Source, Easily Expandable Program for Skeletal Kinematic Measurement Using Model-Image Registration*, (2007), 27.

[45] Nobukazu Okamoto, Leigh Breslauer, Anthony K. Hedley, Hiroshi Mizuta, and Scott A. Banks, *In Vivo Knee Kinematics in Patients With Bilateral Total Knee Arthroplasty of 2 Designs*, The Journal of Arthroplasty **26** (2011), no. 6, 914–918.

[46] Lindsey Palm-Vlasak, R Leitz, H Parvateneni, L Pulido, Mary Beth Horodyski, and Scott Banks, *Minimal Variation in Top Level and Decline Walking Speeds Between Pivoting TKA Subjects and Healthy Controls*, 2022.

[47] Barbara Postolka, Renate List, Benedikt Thelen, Pascal Schütz, William R. Taylor, and Guoyan Zheng, *Evaluation of an intensity-based algorithm for 2D/3D registration of natural knee videofluoroscopy data*, Medical Engineering & Physics **77** (2020), 107–113.

[48] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas, *PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space*, (2017).

[49] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas, *PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation*, arXiv:1612.00593 [cs] (2017).

[50] Pedro Rodrigues, Michel Antunes, Carolina Raposo, Pedro Marques, Fernando Fonseca, and Joao P. Barreto, *Deep segmentation leverages geometric pose estimation in computer-aided total knee arthroplasty*, Healthcare Technology Letters **6** (2019), no. 6, 226–230.

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, (2015).

[52] C. E. H. Scott, C. R. Howie, D. MacDonald, and L. C. Biant, *Predicting Dissatisfaction Following Total Knee Replacement: A Prospective Study of 1217 Patients*, The Journal of Bone and Joint Surgery. British volume **92-B** (2010), no. 9, 1253–1258.

[53] G. Scott, M. A. Imam, A. Eifert, M. A. R. Freeman, V. Pinskerova, R. E. Field, J. Skinner, and S. A. Banks, *Can a total knee arthroplasty be both rotationally unconstrained and anteroposteriorly stabilised?: A pulsed fluoroscopic investigation*, Bone & Joint Research **5** (2016), no. 3, 80–86.

[54] Leena Sharma and Francis Berenbaum, *Osteoarthritis: A companion to Rheumatology*, Mosby, Philadelphia, 2007.

[55] Rita Stagni, Silvia Fantozzi, Angelo Cappello, and Alberto Leardini, *Quantification of soft tissue artefact in motion analysis by combining 3D fluoroscopy and stereophotogrammetry: A study on two subjects*, Clinical Biomechanics **20** (2005), no. 3, 320–329.

[56] Akira Sugi, Keisuke Matsuki, Ryunosuke Fukushi, Takeshi Shimoto, Toshiaki Hirose, Yuji Shibayama, Naoya Nishinaka, Kousuke Iba, Toshihiko Yamashita, and Scott A. Banks, *Comparing in vivo three-dimensional shoulder elevation kinematics between standing and supine postures*, JSES International **5** (2021), no. 6, 1001–1007.

[57] Richard Szeliski, *Computer vision: Algorithms and applications*, second edition ed., Texts in Computer Science, Springer, Cham, 2022.

[58] Matthew G Teeter, Petar Seslija, Jaques S Milner, Hristo N Nikolov, Xunhua Yuan, Douglas D R Naudie, and David W Holdsworth, *Quantification of in vivo implant wear in total knee replacement from dynamic single plane radiography*, Physics in Medicine and Biology **58** (2013), no. 9, 2751–2767.

[59] L. Tersi, S. Fantozzi, and R. Stagni, *3D Elbow Kinematics with Monoplanar Fluoroscopy: In Silico Evaluation*, EURASIP Journal on Advances in Signal Processing **2010** (2009), no. 1, 142989.

[60] Tsung-Yuan Tsai, Tung-Wu Lu, Chung-Ming Chen, Mei-Ying Kuo, and Horng-Chaung Hsu, *A volumetric model-based 2D to 3D registration method for measuring kinematics of natural knees with single-plane fluoroscopy: 2D/3D registration method for measuring natural knee kinematics*, Medical Physics **37** (2010), no. 3, 1273–1284.

[61] Timothy P. Wallace and Owen R. Mitchell, *Analysis of three-dimensional movement using Fourier descriptors*, IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-2** (1980), no. 6, 583–588.

[62] Timothy P. Wallace and Paul A. Wintz, *An efficient three-dimensional aircraft recognition algorithm using normalized fourier descriptors*, Computer Graphics and Image Processing **13** (1980), no. 2, 99–126.

[63] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao, *Deep High-Resolution Representation Learning for Visual Recognition*, arXiv:1908.07919 [cs] (2020).

[64] Toshifumi Watanabe, Masafumi Ishizuki, Takeshi Muneta, and Scott A. Banks, *Knee Kinematics in Anterior Cruciate Ligament-Substituting Arthroplasty With or Without the Posterior Cruciate Ligament*, The Journal of Arthroplasty **28** (2013), no. 4, 548–552.

[65] Toshifumi Watanabe, Takeshi Muneta, Hideyuki Koga, Masafumi Horie, Tomomasa Nakamura, Koji Otabe, Yusuke Nakagawa, Mai Katakura, and Ichiro Sekiya, *In-vivo kinematics of high-flex posterior-stabilized total knee prosthesis designed for Asian populations*, International Orthopaedics **40** (2016), no. 11, 2295–2302.

[66] Anqi Wu, E. Kelly Buchanan, Matthew Whiteway, Michael Schartner, Guido Meijer, Jean-Paul Noel, Erica Rodriguez, Claire Everett, Amy Norovich, Evan Schaffer, Neeli Mishra, C. Daniel Salzman, Dora Angelaki, Andrés Bendesky, The International Brain Laboratory, John Cunningham, and Liam Paninski, *Deep Graph Pose: A semi-supervised deep graphical model for improved animal pose tracking*, Preprint, Animal Behavior and Cognition, August 2020.

[67] Satoshi Yamaguchi, Takahisa Sasho, Hideyuki Kato, Yuji Kuroyanagi, and Scott A. Banks, *Ankle and Subtalar Kinematics during Dorsiflexion-Plantarflexion Activities*, Foot & Ankle International **30** (2009), no. 4, 361–366.

[68] Yuhui Yuan, Jingyi Xie, Xilin Chen, and Jingdong Wang, *SegFix: Model-Agnostic Boundary Refinement for Segmentation*, Computer Vision – ECCV 2020 (Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, eds.), vol. 12357, Springer International Publishing, Cham, 2020, pp. 489–506.

[69] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, *UNet++: A Nested U-Net Architecture for Medical Image Segmentation*, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel R.S. Tavares, Andrew Bradley, João Paulo Papa, Vasileios Belagiannis, Jacinto C. Nascimento, Zhi Lu, Sailesh Conjeti, Mehdi Moradi, Hayit Greenspan, and Anant Madabhushi, eds.), vol. 11045, Springer International Publishing, Cham, 2018, pp. 3–11.

[70] Zhonglin Zhu, Daniel F. Massimini, Guangzhi Wang, Jon J.P. Warner, and Guoan Li, *The accuracy and repeatability of an automatic 2D–3D fluoroscopic image-model registration technique for determining shoulder joint kinematics*, Medical Engineering & Physics **34** (2012), no. 9, 1303–1309.

[71] S. Zuffi, A. Leardini, F. Catani, S. Fantozzi, and A. Cappello, *A model-based method for the reconstruction of total knee replacement kinematics*, IEEE Transactions on Medical Imaging **18** (Oct./1999), no. 10, 981–991.

BIOGRAPHICAL SKETCH

Andrew Jensen is a Florida native from Sarasota, Florida. He attended the University of Florida for his undergraduate degree in Mechanical Engineer, for which he received high honors. He took a brief hiatus from school to work at an orthopaedic solutions company, Exactech. The COVID-19 pandemic cut his time at Exactech short, so he joined the Gary J Miller Orthopaedic Biomechanics Laboratory as a part-time researcher during the summer leading up to his first official semester of graduate school.

Andrew enjoys being outdoors, hiking, reading, and doing different things.