UNIVERSITY OF CAPE TOWN

MASTER'S THESIS

Towards Estimating the Dynamics of Cheetahs in the Wild

Author:
Zuhayr Parkar

Supervisors:
Dr Fred NICOLLS,
Dr Amir PATEL

A thesis submitted in fulfilment of the requirements
for the degree of Master of Science in Electrical Engineering

in the

African Robotics Unit
Department of Electrical Engineering

2025-10-24

Declaration of Authorship

I, Zuhayr Parkar, declare that this thesis titled, "Towards Estimating the Dynamics of Cheetahs in the Wild" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Z. Parkar

Date:

UNIVERSITY OF CAPE TOWN

Abstract

Engineering and the Built Environment

Department of Electrical Engineering

Master of Science in Electrical Engineering

Towards Estimating the Dynamics of Cheetahs in the Wild

by Zuhayr Parkar

Gaining insight into animal biomechanics is critical for understanding legged manoeuvrability. However, studying animals in the wild remains challenging due to the lack of accurate and non-invasive measurement techniques. This thesis proposes a novel approach for studying cheetah motion in its natural environments for the purpose of investigating the biomechanics of the cheetah.

Previous attempts at 3D reconstruction and motion capture of cheetahs in the wild have been limited by the inaccuracies in conventional markerless motion capture techniques. These challenges are addressed by the use of a photogrammetry system to reinterpret the marker less motion capture problem posed for studying animals in the wild.

Videos of cheetahs performing running and walking motions were recorded using a multi-camera photogrammetry setup, enabling 3D reconstructions of the cheetah's outer surface. The unique fur pattern of the cheetah allowed for the placement of "virtual markers", that provided a non-invasive alternative to traditional physical markers. To further aid in kinematic estimation, a skeletal model based on anatomical data of a cheetah was developed to track internal joint positions. This approach achieved root-mean-square marker tracking errors below 4 cm.

Furthermore, a kinetic model of the cheetah was developed. This model, combined with trajectory optimisation techniques, was used to estimate ground reaction forces and joint moments from the motion capture data. The optimisation process refined the captured gaits by enforcing physical constraints, ensuring physically plausible motion reconstructions.

The methods presented in this thesis establish a new framework for studying cheetah biomechanics, combining non-invasive virtual markers with skeletal modelling to achieve centimetre-scale accuracy in joint tracking. This approach enables detailed analysis of cheetah locomotion in natural environments without the limitations of traditional marker-based systems.

Acknowledgements

I would like to express my heartfelt gratitude to my supervisors, Dr. Fred Nicholls and Dr. Amir Patel, whose guidance, expertise, and support have been invaluable throughout this project.

I am also deeply thankful to Naoya Muramatsu and Pasha van Bijlert for their consistent availability and insightful contributions whether over email or during our meetings.

My sincere appreciation goes to my parents, whose encouragement and belief in my pursuit of further education have been a constant source of motivation.

Lastly, I would like to extend my thanks to my friends and family for their unwavering support and understanding throughout the duration of this project.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	**T**wo-**d**imensional
3D	**T**hree-**d**imensional
fps	**f**rames **p**er **s**econd
DOF	**D**egrees **O**f **F**reedom
COM	**C**entre **O**f **M**ass
MVS	**M**ulti-**V**iew **S**tereo
SfM	**S**tructure **f**rom **M**otion
GRF	**G**round **R**eaction **F**orce
CT	**C**omputed **T**omography
MSK	**M**usculo**sk**eletal
ID	**I**nverse **D**ynamics
IK	**I**nverse **K**inematics

Chapter 1

Introduction

In recent years, the field of biomimetics and bio-inspired robotics has seen a significant increase in popularity. This can largely be attributed to advancements in computer vision systems and enhanced computational capabilities, all of which aid in better understanding and quantifying animal behaviour. The study of bio-inspired robots is inseparable from the study of animal locomotion. Thus, before robots can be built based on animals, these animals must be studied, and their movements and actions quantified. Animal studies are difficult; more often than not, their behaviour, as well as the environment in which they are found, cannot be easily controlled. As a result, in-the-wild capture of these animals' dynamics is needed. Understanding the movements of animals in their natural environment is crucial for advancing both the field of animal biomechanics and biomimetics or bio-inspired robotics.

The cheetah (Acinonyx jubatus) presents itself as a subject for studying animal locomotion due to its impressive speed and agility. However, conducting such studies cannot be done without overcoming significant challenges because of both the animal's behaviour and its natural habitat. These challenges are further complicated by the inability to use traditional marker-based motion capture (mocap) techniques. While research on cheetah motion capture is not entirely novel, prior efforts exhibit notable limitations.

AcinoSet [1] presented the first dataset on cheetah locomotion. It provides 3D reconstructions of the animal's motion based on multi-camera setups. Despite its contributions, the methodology and results reveal several shortcomings, including:

- Only three cameras on each side, leading to significant triangulation error.
- The triangulation method heavily dependent on in-field calibration and matching 2D keypoints.
- A frame rate of 120 frames per second (fps), which is low for capturing such a fast animal.
- Inability to quantify error in 3D due to the lack of ground truth.
- Motion data that includes running and manoeuvres but excludes walking.

These issues must be overcome to obtain better kinematics, without which meaningful biomechanical analysis of the animal cannot be conducted. Thus, before any further

biomechanical analysis can be conducted, motion capture that adheres to current biomechanical standards must be achieved.

At present, markerless motion capture relies heavily on computer vision techniques, often resulting in a significant disconnect from the underlying anatomical realities of the animals being studied [2]. To fully understand and quantify the animals' motion, the underlying biomechanical principles of the animal must also be considered.

As a solution to this problem, this project aims to bridge the gap between computer vision and biomechanics with the goal of evaluating the animal's motion. To this end, a photogrammetry system is proposed to reinterpret the markerless motion capture problem as a pseudo-marker-based motion capture approach. The cheetah, due to its unique spotted coat, is an ideal subject for markerless motion capture. The initial motion capture can then be tied to the cheetah through the use of an underlying skeletal model, enabling further kinematic and dynamic analyses of the animal.

Further complicating the problem is the lack of real-world data to serve as ground truth. Thus, rather than solely attempting to create a ground truth dataset, this project aims to propose a methodology and solution to the problem of estimating the dynamics of cheetahs in the wild. By addressing these challenges, the proposed solution aims to provide new insights into cheetah biomechanics while advancing the field of markerless motion capture.

1.1 Problem Statement

This project aims to develop a methodology for 3D motion capture and dynamic analysis of wild cheetahs. Given the in-the-wild nature of the project, the scope of motion capture is strictly limited to markerless motion capture. Previous endeavours addressing this problem have presented several shortcomings, as mentioned. This project seeks to overcome the limitations of traditional markerless motion capture and provide a pathway for estimating kinematics and dynamics in alignment with current biomechanical motion capture standards [3].

While existing multi-camera footage of cheetahs is available, these videos and reconstructions are outdated and subject to their own limitations. Consequently, the development of this project requires the collection of new footage of cheetahs to adequately address the problem. Additional constraints are imposed on the camera systems used due to the uncontrolled nature of the environment in which the footage is recorded. This makes 'laboratory'-type setups infeasible.

Although motion capture systems capable of providing millimeter-level accuracy are available, these systems are completely impractical for studying wild animals. As a result, motion capture methods for this application are inherently limited by the shortcomings of current markerless approaches. Therefore, the motion capture pipeline must deviate from traditional methods and instead connect more directly to the animal's actual biomechanics.

The problem is further complicated by the necessity of understanding the underlying biology of the animal. It is imperative that, before any dynamic analysis of the animal

can occur, the proposed method adheres to biomechanical standards in both motion capture and dynamic analysis. The goal of the dynamic analyses is not to produce exact results but rather to develop a method applicable for fully in-the-wild analyses of cheetahs.

1.2 Project Objectives

The main objectives of the project are as follows:

- **Investigation of Methodologies:** Explore methods to address the challenges of traditional markerless motion capture, along with techniques for connecting the motion capture approach to the animal's biology.
- **Development of a Motion Capture Pipeline:** Create a pipeline for estimating the joint angles and positions of cheetahs from multi-camera video footage.
- **Development of a Dynamic Analysis Pipeline:** Develop a method to estimate the dynamics of cheetahs in motion based on the derived kinematics.

1.3 Project Scope

The outcomes of this project are based on the design and implementation of a motion capture system for cheetahs, as well as the development of a method for estimating their dynamics. To achieve these outcomes, the project has specific requirements and limitations:

- The project will involve the collection of multi-camera video data of cheetahs in motion.
- Any method implemented must rely strictly on markerless motion capture.
- The cameras used are limited to unmodified GoPro Hero 12 Black models.
- Any proposed method must be verified through simulations or real-world testing before being applied to cheetahs.
- The final motion capture and dynamic analysis must be connected to the animal's underlying biomechanics.
- The motion capture is not restricted to any particular type of motion.
- A dynamic analysis based on the motion capture data will be conducted.
- The dynamic analysis must address the absence of ground reaction force data.

1.4 Project Outline

The progression of the project follows linearly. Each chapter builds on the work presented in the preceding chapters. The project outline is visually presented in Figure 1.1 and summarised below:

- **Chapter 2**: Provides an overview of the literature on the state of animal motion capture and skeletal models used for estimating the dynamics of animals.
- **Chapter 3**: Outlines the methodology used in this project.
- **Chapter 4**: Describes the data collection process and the methodology for photogrammetric reconstruction of cheetahs from the video data.
- **Chapter 5**: Details the development of a method for using a skeletal model to estimate the kinematics of cheetahs based on the photogrammetric reconstructions.
- **Chapter 6**: Explains a methodology for utilising the derived kinematics for dynamic analysis.

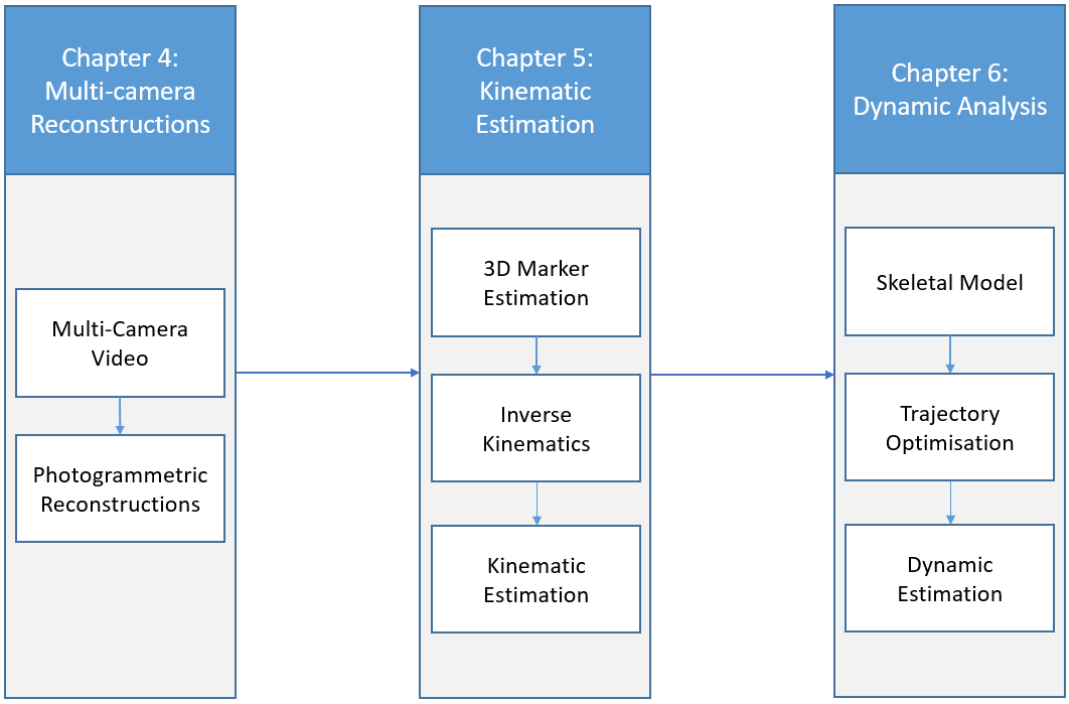

FIGURE 1.1: System overview of the project chapters

1.5 Project Overview

This thesis addresses the challenge of capturing and analysing the biomechanics of wild cheetahs in uncontrolled environments, where traditional motion capture methods are impractical. The primary goal was to develop a markerless, photogrammetry-based system capable of reconstructing 3D motion and estimating dynamics without relying on physical markers or ground reaction force measurements. By integrating computer vision techniques and biomechanical modelling, the project aimed to address the problem of studying animal locomotion in the wild.

Multi-camera arrays and photogrammetry techniques were used to reconstruct dense 3D point clouds from synchronised video footage. Simulations were used to optimise camera placement and coverage, while real-world trials refined the robustness of the

system. The resulting reconstructions achieve over 90% completeness as compared to the animals' silhouette in the original video frames, thus providing a near complete reconstruction of the animals' surface to serve as a solution space for marker projection.

To convert surface reconstructions into biomechanically meaningful data, spots on the cheetah's fur were tracked across frames. These "virtual markers" were reprojected onto 3D point clouds and linked to a skeletal model derived from CT scan data. Inverse kinematics in OpenSim translated marker trajectories into joint angles. Finally, trajectory optimisation enforced dynamic feasibility and allowed for estimating GRFs and joint moments without direct force measurements. This two-stage process, tracking the recorded motion and refining it through physics-based constraints, provides a novel framework for analysing locomotion in the absence of laboratory-grade instrumentation.

The outcomes of the project provide a framework for studying cheetah locomotion with implications for wildlife conservation, robotics, and biomechanics.

1.6 Project Outcomes

These are the three key outcomes of the project:

- **Development of a Markerless Motion Capture System:** A pipeline for capturing and reconstructing the motion of wild cheetahs from multi-camera video footage.
- **Integration with Biomechanics:** A methodology for linking photogrammetric reconstructions with skeletal models to estimate kinematics and dynamics aligned with biomechanical standards.
- **Dynamic Analysis Framework:** A system for analysing cheetah motion dynamics, addressing challenges such as the absence of ground reaction force data.

Chapter 2

Literature Review

This chapter discusses the underlying theory and literature relevant to subsequent chapters. In particular, this chapter covers three main topics: animal pose estimation, photogrammetry for 3D reconstruction, and biomechanical modelling.

It begins with an overview of the current state of animal pose estimation, discussing both keypoint-based approaches and animal mesh recovery techniques, and critically examines their limitations. Following this, photogrammetry-based motion capture is proposed as a potential solution. Previous implementations of such systems are reviewed to motivate the research presented in this thesis. To provide a solid foundation for these methods, the chapter also reviews the mathematical principles underpinning photogrammetry, thereby grounding the work in a rigorous theoretical framework. Additionally, existing software and implementation techniques used for photogrammetry are explored, highlighting the key methods and tools required for developing an effective solution.

The chapter then shifts focus to a biomechanical perspective by examining previous research on quadrupedal gait analysis. This involves an overview of the methods required to perform similar biomechanical analyses on cheetahs. This section not only outlines the methods used in similar analyses, but also justifies the integration of biomechanical modelling software to enable the use of motion capture data in dynamic analyses.

The chapter concludes by reviewing previous attempts to address these challenges while identifying gaps in the current literature, and placing the present study within the broader research context.

2.1 Animal Pose Estimation

Animal pose estimation involves identifying and determining the locations of an animal's joints and skeletal structure. This section focuses exclusively on markerless approaches, which have gained traction as viable alternatives to traditional marker-based systems for capturing animal motion.

In recent years, the field of animal pose estimation has witnessed remarkable growth, driven by advancements in computer vision and machine learning. These developments have enabled the creation and availability of an increasing number of animal pose datasets, enhancing the accessibility and accuracy of pose estimation techniques.

However, compared to human pose estimation, research on animal pose estimation presents unique challenges. The most notable difficulty lies in estimating the poses of wild animals in uncontrolled, natural environments without relying on physical markers.

To address these challenges, two prominent approaches have emerged:

1. **Keypoint-Based Methods:** These involve detecting and tracking specific anatomical landmarks, such as joints, from images or video sequences. Keypoint-based methods are computationally efficient and have shown promising results, especially when supported by robust training datasets.
2. **Animal Mesh Recovery:** This approach goes beyond keypoints by reconstructing a 3D mesh of the animal's body. Mesh recovery provides a more comprehensive representation of an animal's pose and shape, making it particularly useful for biomechanical analysis and motion studies.

2.1.1 Keypoint Methods

Keypoint detection is one of the most commonly used techniques for animal pose estimation, particularly for estimating joint positions from single images. These methods typically employ deep learning architectures, such as ResNets, to detect anatomical landmarks with high accuracy. One of the most widely used tools for this is DeepLabCut [4].

Pose estimation in 2D involves predicting the positions of joints directly from images. While effective for identifying animal poses from a single image, 2D keypoints lack depth information, thereby limiting their applicability for complex motion analyses.

3D pose estimation, on the other hand, predicts joint positions in three-dimensional space, providing poses that are grounded in reality and suitable for biomechanical or kinematic analysis. Multiview 3D pose estimation systems typically rely on an array of synchronised and calibrated cameras to capture images of an animal from multiple perspectives simultaneously. A common pipeline involves performing 2D keypoint detection on each image, followed by triangulating the keypoints across views to compute their 3D positions. The depth information is derived from the geometric relationships between cameras, as described in [5].

Previous works, such as OpenMonkeyStudio [6] and AcinoSet [1], have successfully implemented this pipeline to estimate 3D poses for various animals. These studies highlight the potential of multiview systems to improve the accuracy of pose estimation in natural settings.

AcinoSet

In the context of cheetah motion capture, AcinoSet represents the first dedicated motion capture dataset for cheetahs. Understanding its merits and shortcomings is crucial, as it serves as a baseline for markerless motion capture approaches. While AcinoSet has laid important groundwork, its limitations highlight the challenges associated with traditional keypoint-based methods.

One significant limitation of AcinoSet is its sparse camera coverage. The dataset employs only three cameras on each side of the animal, which leads to considerable triangulation errors due to insufficient viewpoint diversity. Furthermore, accurate 3D reconstruction in AcinoSet is heavily reliant on precise in-field camera calibration. Variations in calibration conditions introduce additional uncertainty, affecting the consistency of pose estimation. Another critical issue lies in the inconsistent matching of 2D keypoints across multiple images. Ambiguities and errors in keypoint detection often result in unreliable matches, compounding the inaccuracies in 3D triangulation.

The dataset's temporal resolution is also a concern. Capturing at a frame rate of 120 frames per second is insufficient to resolve the rapid dynamics of a fast-moving animal like the cheetah. This low temporal resolution, combined with the aforementioned challenges, leads to high errors in the dataset, which hinder its utility for detailed biomechanical analysis. Moreover, the absence of ground truth data for 3D poses prevents the quantification of these errors in 3D space, further limiting the dataset's applicability.

The shortcomings of AcinoSet underscore two major issues inherent to traditional keypoint-based motion capture methods. First, the inherent uncertainty in 2D keypoint detection across multiple images creates significant challenges, as exact matches often do not exist. This uncertainty propagates through subsequent stages of the reconstruction pipeline. Second, the heavy reliance on precise calibration data, particularly in setups with sparse camera arrays, means that 3D pose estimations are inconsistent across the animal's surface. These limitations emphasise the need for more robust and comprehensive approaches to motion capture, particularly for applications requiring high accuracy, such as biomechanical analysis.

Traditional keypoint-based methods are constrained by their evaluation metrics. Since these methods are typically evaluated using 2D keypoints, the analysis remains disconnected from the physical reality of 3D space. Without reliable ground truth data for validation, it becomes difficult to accurately quantify the performance and robustness of such approaches. The limitations of AcinoSet illustrate the broader challenges faced in advancing markerless motion capture.

2.1.2 3D Mesh Recovery

3D mesh reconstruction has opened new possibilities for animal motion capture using image sets, with most methods building upon the SMAL model [7]. Unlike the human SMPL model [8], which is trained on thousands of human scans, SMAL was developed using scans of animal toys. Zuffi et al. created a template mesh, which could be deformed and fitted to various poses and shapes using a global/local stitched shape model. This parametric mesh enables reconstruction from a single image by fitting shape, pose, and camera parameters to 2D silhouettes and keypoints.

Later, Zuffi et al. [9] introduced SMALR, which improves upon SMAL by incorporating multi-view images and 2D silhouettes to reduce ambiguities in 3D pose estimation. However, these methods depend heavily on synthetic data and 2D-to-3D reconstruction, limiting their accuracy in providing reliable ground truth or joint position estimates for motion capture applications. Additionally, these methods

have not been evaluated against real-world measurements nor have they previously been used for any biomechanical analyses. While they provide a method for pose estimation, it does not seem entirely applicable for 3D kinematic estimation.

2.2 Photogrammetry for Motion Capture

Markerless motion capture poses significant challenges due to inherent limitations, including inaccuracies associated with multi-camera triangulation and the critical dependence on precise calibration [10]. These issues have long been recognised in the field, and alternative markerless methods have been proposed to address them. Among these, photogrammetry has emerged as a promising approach due to its adaptability and potential ease of use [11].

Unmarked photogrammetry using multiple synchronised video cameras has been applied for motion capture in birds [12]. Similarly, a photogrammetry-based approach with non-invasive markers drawn on bat wings was used to capture motion in a more natural setting [13]. Photogrammetry has also been successfully employed for human motion capture, using specially textured suits [14]. However, these methods generally require controlled environments and are challenging to implement in the wild.

In a notable application, Sellers et al. [15] demonstrated a photogrammetry system for motion capture of primates in a zoo. Their work represents a significant step toward markerless motion capture in uncontrolled environments, although their system was limited to single-sided capture. While this simplification reduces system complexity, it fails to capture non-periodic gaits effectively, restricting its applicability in dynamic scenarios.

Photogrammetry offers a unique solution to the challenges of markerless motion capture, particularly for textured subjects. The cheetah, with its distinctive skin pattern, is an ideal candidate for photogrammetric reconstruction, presenting an opportunity for advancing motion capture in outdoor settings.

2.3 Mathematical Theory for Photogrammetry [16]

Reconstructing 3D geometry from photographs is a classic computer vision problem. The goal of an image-based 3D reconstruction algorithm is to solve the following problem, "When presented with multiple photographs capturing an object or scene, determine the most probable three-dimensional form that would generate these images, assuming we have knowledge of the camera's parameters." While this definition traditionally applies to photogrammetry with known calibration parameters, the scope of this project extends beyond these constraints. Instead, we focus on Milti-View Stereo(MVS) using Structure from Motion(SFM) algorithms, where camera parameters are computed as part of the solution. This approach follows three main steps:

1. Image collection.

2. Computation of camera parameters for each image.
3. Reconstruction of 3D geometry from the image set and their corresponding cameras.

This section establishes the mathematical foundation necessary to understand and solve this problem.

2.3.1 Camera Projection Models

To understand the problem posed by MVS reconstruction, it is crucial to comprehend how a 3D point in the world is projected onto a pixel in an image. The most commonly used model for MVS is the pinhole camera model [16].

In this model, a 3D point $P_w = (X_w, Y_w, Z_w)^T$ in the world is projected onto the image plane using a perspective transformation. This process maps the 3D point from Euclidean space $\mathbb{R}^3$ to 2D space $\mathbb{R}^2$, which results in dimensionality loss.

Specifically, under the pinhole camera model, a point in space is mapped to the image plane at the point where a line connecting the 3D point to the centre of projection intersects the image plane. This projection is shown in Figure 2.1 and is mathematically represented by the equation below:

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \mathrm{P}_{3\times4} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{2.1}$$

where $(x, y, w)^T$ are homogeneous coordinates on the image plane, $\mathrm{P}_{3\times4}$ is the camera projection matrix, and $(X, Y, Z, 1)^T$ is the homogeneous representation of the 3D point. The final 2D pixel coordinates (x', y') are obtained by normalising the homogeneous coordinates:

$$x' = \frac{x}{w}, \quad y' = \frac{y}{w}. \tag{2.2}$$

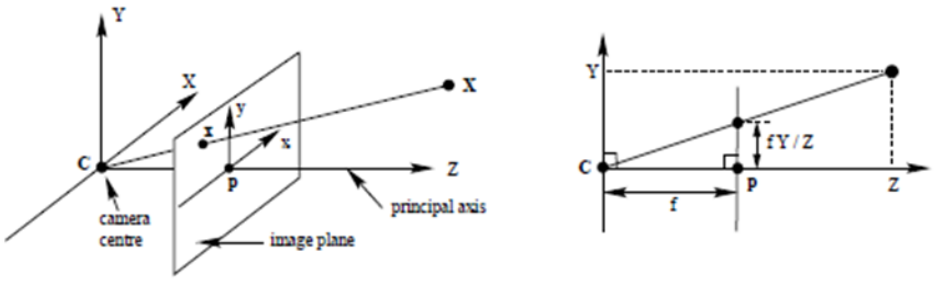

FIGURE 2.1: Pinhole camera geometry. Camera centre C and principal point p. Adapted from [16]

2.4 Projection Matrix

The projection matrix P, is a 3×4 matrix that encapsulates the complete mapping from 3D world coordinates to 2D image coordinates. It combines both the internal camera parameters (intrinsic parameters) and the camera's position and orientation in space (extrinsic parameters). This matrix can be decomposed as:

$$P = \underbrace{\begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}}_{K} \cdot \underbrace{\left(\begin{array}{ccc|c} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{array}\right)}_{R|t}. \tag{2.3}$$

The matrix K is called the camera calibration matrix, because it is composed of quantities intrinsic to the camera: vertical and horizontal focal lengths f_x, f_y, principal point (c_x, c_y), and skew s. However, certain assumptions regarding this matrix can be made as follows. Pixels are assumed to have no skew ($s = 0$), and to be square ($f_x = f_y$). Also, if an image has not been cropped, it is safe to assume the principal point is at the centre of the image. Thus, the matrix K is given as:

$$K = \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.4}$$

The matrix $[R|t]$ is commonly known as the extrinsic matrix, where R is the rotation of the camera and t is the translation of the camera.

To describe this, a coordinate transformation is applied to points in the global reference frame, converting them to the camera's coordinate system. This transformation uses a 3×3 rotation matrix R and a 3×1 translation vector t:

$$P_c = RP_w + t, \tag{2.5}$$

where $P_c = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}$ represents the 3D point in camera coordinates, $P_w = \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}$ is the point in world coordinates, R encodes the camera's orientation, and t specifies its position relative to the world origin.

In homogeneous coordinates, this rigid-body transformation becomes a single 4×4 linear transformation:

$$\begin{bmatrix} P_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^\top & 1 \end{bmatrix} \begin{bmatrix} P_w \\ 1 \end{bmatrix}, \tag{2.6}$$

where the transformation matrix combines rotation and translation. When expanded:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \tag{2.7}$$

This formulation enables efficient composition of transformations while preserving the geometric relationship. The complete projection of a 3D world point P_w to a 2D image point x is then achieved by combining this extrinsic transformation with the camera's intrinsic parameters through perspective projection. This mathematical framework underpins camera pose estimation and 3D reconstruction in multi-view stereo systems.

2.4.1 Epipolar Geometry

Beyond single camera projection, there arises the topic of multiple camera projections. Epipolar geometry is the intrinsic projective geometry between two views. It is essentially the geometry of the intersection of the image planes with the pencil of planes having the baseline as its axis, where the baseline is the line joining the camera centres.

Thus, given a point X in 3-space, which is imaged in two views, at x in the first, and x' in the second, the relationship between the two points is described by a 3×3 matrix F, known as the fundamental matrix. The fundamental matrix satisfies the condition that for any pair of corresponding points $x \leftrightarrow x'$ in the two images:

$$x'^T F x = 0. \tag{2.8}$$

This is true because if points x and x' correspond, then x' lies on the epipolar line $l' = Fx$ corresponding to the point x. In other words, $0 = x^T l' = x^T F x$. Conversely, if image points satisfy the relation $x^T F x' = 0$, then the rays defined by these points are coplanar. This is a necessary condition for points to correspond.

Suppose only x is known, and the goal is to determine how the corresponding point x' is constrained. The plane π is defined by the baseline and the ray associated with x. From the geometric relationship, the ray corresponding to the (unknown) point x' lies within π, which implies that x' must lie on the line of intersection l' of π with the second image plane. This line l' is the projection in the second view of the ray back-projected from x. It represents the epipolar line corresponding to x. This is demonstrated in Figure 2.2

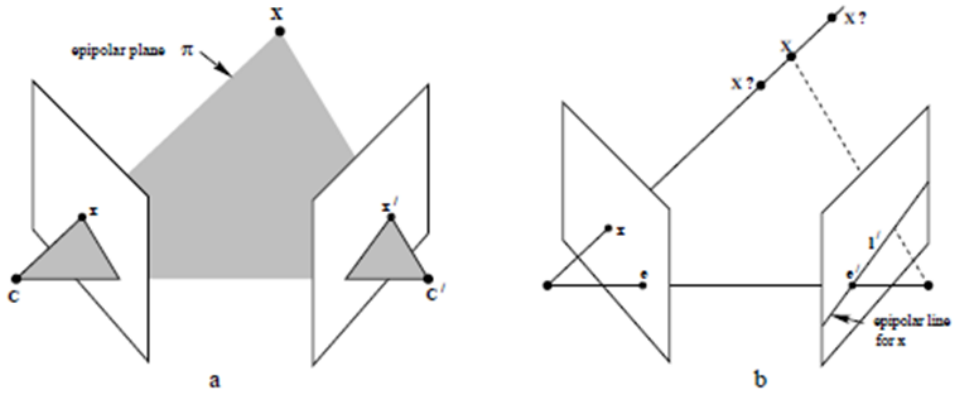

FIGURE 2.2: Demonstration of epipolar geometry. Adapted from [16]

2.4.2 3D Triangulation

The correspondence of two points in 3D space is defined by the constraints of epipolar geometry, assuming exact pixel matches. However, real-world measurements often involve discrepancies between pixel points, leading to rays that do not intersect. The goal is to estimate the 3D point that best fits the observed data. The triangulation process is shown in Figure 2.3.

This is commonly achieved through linear triangulation, where the 3D point $\mathbf{X}$ is estimated using projections from multiple camera views. The direct linear transformation (DLT) method is formulated as

$$A\mathbf{X} = 0, \tag{2.9}$$

where A is derived from camera projection matrices and image points. The system is solved using singular value decomposition (SVD), with the solution being the singular vector corresponding to the smallest singular value of A. While straightforward, the DLT method assumes perfect correspondences and may yield suboptimal results when noise is present.

An alternative is the Sampson approximation [16], which simplifies the non-linear minimisation of geometric error. Instead of minimising perpendicular distances between points and epipolar lines directly, it uses a linearised error metric based on a first-order Taylor expansion. This approach balances computational efficiency and accuracy, making it well-suited for real-time 3D reconstruction where exact optimisation is impractical.

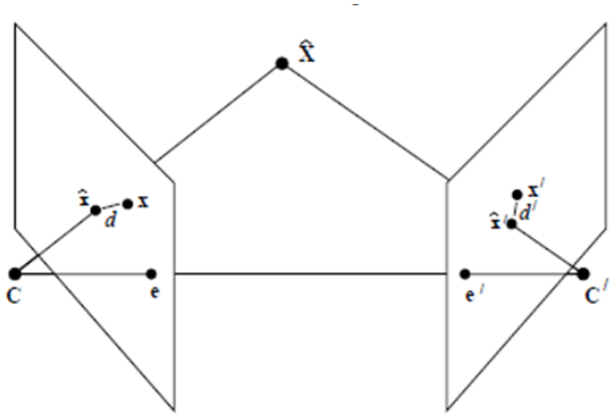

FIGURE 2.3: Demonstration of the triangulation process. Adapted from [16]

2.4.3 Bundle Adjustment

Bundle adjustment is a crucial optimisation technique commonly employed to refine SfM models by minimising reprojection errors in the camera parameters and 3D point coordinates. It involves minimising the following non-linear least-squares error function:

$$E(P, M) = \sum_{j} \sum_{i \in V(j)} \left| P_i(M_j) - m_{ij} \right|^2 . \tag{2.10}$$

Here, $P_i(M_j)$ denotes the projection of the 3D point M_j in the i-th camera, while m_{ij} represents the 2D image coordinate corresponding to the projection of M_j in the same camera. $V(j)$ is the set of cameras where point M_j is visible. The goal is to minimise the discrepancy between the observed 2D image coordinates m_{ij} and the 3D projections $P_i(M_j)$ across all cameras.

The accuracy of the bundle adjustment is often quantified by the root mean square error (RMSE), which provides a measure of the reprojection error in pixels:

$$\text{RMSE}(P, M) = \sqrt{\frac{E(P, M)}{N}}, \tag{2.11}$$

where N is the total number of residual terms summed in the error function. Before bundle adjustment, RMSE values are typically on the order of several pixels, whereas after optimisation, RMSE values often fall below the sub-pixel level, indicating a significant improvement in accuracy.

In practical applications, bundle adjustment can be extended to fuse additional sensor data into the optimisation process. This is done by adding penalty terms to the objective function that enforce deviations of the camera parameters from the predicted values based on GPS measurements or initial camera calibration. If the camera model is inaccurate, large reprojection errors can prevent effective matching of

2D image points to their corresponding 3D points, leading to significant degradation in reconstruction quality.

As such, bundle adjustment is often required to ensure sub-pixel accuracy in the camera parameters and 3D point locations, which is essential for high-quality MVS results. Since MVS is highly sensitive to reprojection errors, bundle adjustments are frequently a prerequisite to achieve sub-pixel reprojection accuracy.

It is worth noting that because the reprojection error is measured in pixels, one can downsample the input images and rescale the camera parameters until the reprojection error drops below a certain threshold. This approach is effective as long as the downsampled images retain sufficient texture and details for MVS to function properly.

2.4.4 SfM Implementations

Given the need to provide photogrammetric reconstructions from videos, choosing the SfM/MVS pipeline best suited for the task is an important consideration. This topic has been the subject of numerous studies [17–20], all of which evaluate previously implemented solutions to the SfM and MVS problems. This section briefly provides an overview of these results to aid in making an informed decision on which reconstruction pipeline to utilise.

Notably, three programs are frequently mentioned in papers regarding SfM: OpenMVG [21], VisualSFM [22], and COLMAP [23]. OpenMVG and COLMAP are open-source libraries and implementations of the SfM pipeline. VisualSFM, however, only provides an implementation, making it much less flexible compared to the other two. The required implementation of SfM/MVS is to be applied to batches of frames sequentially, thus the need for automation rules out VisualSFM due to its inflexibility. It should also be noted that due to the flexibility requirement, notable commercial solutions such as Agisoft[1] and Autodesk ReCap[2] are not considered.

As for COLMAP and OpenMVG, both are very similar in their implementations. Specifically, both use SIFT [24] for initial feature extraction, the same algorithms for four-point homography estimation and five-point relative pose estimation [16], and the same algorithms for incremental and global bundle adjustment. The primary difference between the two lies in their implementation of dense point cloud generation.

Across three papers evaluating the performance of the two against each other, all three found that COLMAP produced the best average results for 3D reconstruction. Additionally, COLMAP allows for the use of previously calculated calibrated intrinsic and extrinsic parameters during the dense reconstruction phase. Considering the use case and the need for applying photogrammetry to batches of video frames, COLMAP presents itself as the most suitable implementation of an SfM/MVS pipeline.

[1] www.agisoft.com
[2] www.autodesk.com/products/recap/overview

2.5 Gait Analysis

Following an exploration of the relevant literature on photogrammetry, it is similarly necessary to review the theory of biomechanical gait analysis given the scope of this project. A key observation from the literature is that human gait analyses and biomechanical studies far exceed research conducted on animals, particularly wild animals. However, it has been broadly established in the literature that the workflows used for analysing human gait can generally be applied to animals [25].

Similarly, despite the vast variety of quadrupedal mammals, their locomotion patterns, in terms of gait and kinematics, exhibit significant commonalities [26, 27]. For these reasons, the methodologies for gait analysis, kinematic, and kinetic estimations used in human studies are often applicable to animals as well. Furthermore, these similarities in gait and kinematics among quadrupeds provide additional insights and evidence that the study of animal biomechanics, particularly within the engineering domain, cannot be conducted in isolation. It must consider the fundamental biomechanical principles underlying these studies.

This section aims to provide a broad overview of previous research on quadrupedal gait analysis to offer insights into establishing a suitable and effective approach.

2.6 Quadrupedal Analyses

The subject of this study is the cheetah. Therefore, rather than focusing on research regarding bipeds, studies on quadrupeds are more informative. While the field of animal gait analysis is not as advanced as human gait analysis, there exists significant research that aids in understanding the dynamics of quadrupeds.

The first study of importance is by Zhang et al. [28], where the authors perform a mechanical analysis of a running cheetah. This study dramatically simplifies the cheetah's motions and neglects the significance of Ground Reaction Forces(GRF). Additionally, it provides joint torques for the analysis; however, due to the simplifications, these cannot be considered accurate. Notably, their use of a three-linked leg model for dynamic analysis, which excludes pelvis motion, is of particular interest.

Nielsen et al. [29] proposed a link segment model to study limb dynamics in dogs. Their findings suggest that linked limb models can provide reasonable insights but necessitate the inclusion of GRFs. However, the simplifications made in their study resulted in larger standard deviations for joint moment estimates across all results.

While 2D linked limb models can offer insights into kinematics, they fundamentally struggle to provide meaningful dynamic estimations. As an alternative, 3D skeletal models based on the animal's internal geometry offer a more accurate approach. For instance, Andrada et al. [30] used skeletal models derived from x-ray fluoroscopy to estimate dynamics. Similarly, skeletal models combined with marker-based motion capture have been employed to analyse the dynamics of greyhounds [31, 32]. These models provided a deeper understanding of animal motion by leveraging detailed anatomical geometry.

Notably, skeletal models do not necessarily require a full-body representation to provide accurate analyses. Becker et al. [33, 34] demonstrated this by using a single-limb model of a horse, consisting of five joints based on scanned skeletal geometry, to compute joint reaction forces. Despite the use of a partial model, they were able to estimate and validate joint loading with significant accuracy, illustrating that simplified skeletal models can yield detailed insights.

A common theme across the use of skeletal models is the reliance on biomechanical analysis software to assist in their computations. The aforementioned studies on skeletal models consistently used OpenSim [35] for their analyses. Based on these findings and recommendations, creating a skeletal model and employing biomechanical modelling software appears to provide substantially more freedom and accuracy than relying on simplified linked limb models.

2.7 OpenSim

Based on the clear advantages of using biomechanical models for analysing animal dynamics, it is evident that this approach is both suitable and effective. OpenSim[3] is an open-source software system for biomechanical modelling, simulation, and analysis. It is the most widely used biomechanical analysis tool and has garnered thousands of researchers due to its accessibility and broad applicability in biomechanical studies [36].

Simulations representing the dynamics of movement complement experimental data by calculating quantities that cannot be directly measured. These simulations are invaluable for understanding movements and conducting analyses that are either challenging to perform experimentally or impossible to achieve otherwise. By combining simulations with experimental data, a more comprehensive understanding of the dynamics at play can be achieved.

OpenSim offers two primary approaches for this purpose:

- **Forward Dynamics**: This approach is applicable when muscle or actuator activations are known. By using these activations as inputs, the resulting motion can be predicted. While forward dynamics is a powerful tool, it is less relevant to this project except in cases of direct validation.
- **Inverse Dynamics**: This process uses experimental measurements of a subject's motion to derive meaningful insights into the forces and actuation required to produce the observed movements. Inverse dynamics is a common strategy for integrating experimental data with musculoskeletal models. The workflow for this process is illustrated in Figure 2.4.

The first step in inverse dynamics involves using a biomechanical model of the body to convert marker position measurements into joint angles through a process called inverse kinematics. After the experimental data is mapped onto the model, the joint angles are differentiated with respect to time to estimate joint angular velocities and accelerations. These values, combined with measurements of external forces applied

[3] https://opensim.stanford.edu/

to the biomechanical model, are used to estimate joint moments. Furthermore, the model can be enhanced with optimisation algorithms to calculate muscle forces or, in cases where muscles are not modelled, actuator controls.

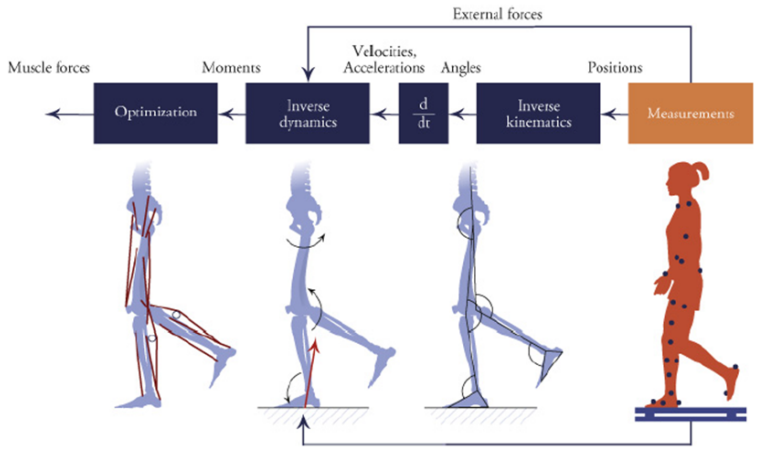

FIGURE 2.4: Elements of a typical inverse dynamics analysis. Adapted from [36]

2.8 OpenSim Moco

Musculoskeletal simulations in OpenSim are often categorised based on how the motion is derived. Motion can either be measured through experimental data, resulting in prescribed motion, or predicted by the simulation itself. A third category, which lies between prescribing and predicting motion, is tracking motion, where errors between model kinematics and reference data are part of the problem. In this case, both the underlying model and the data contribute to an optimal control problem.

Optimal control problems seek the parameters and time-varying controls of a system that minimises a cost (e.g. energy consumption) subject to the system dynamics, which are expressed as differential-algebraic equations [37]. These problems are often highly complex, and writing code to solve them can be time-consuming. As a result, many biologists lack the technical expertise required to implement such solutions effectively.

OpenSim Moco, introduced by Dembia et al. [38], is an open-source, customisable, and extensible software toolkit designed to solve optimal control problems with OpenSim musculoskeletal models. Figure 2.5 provides an overview of the problem that OpenSim Moco aims to address:

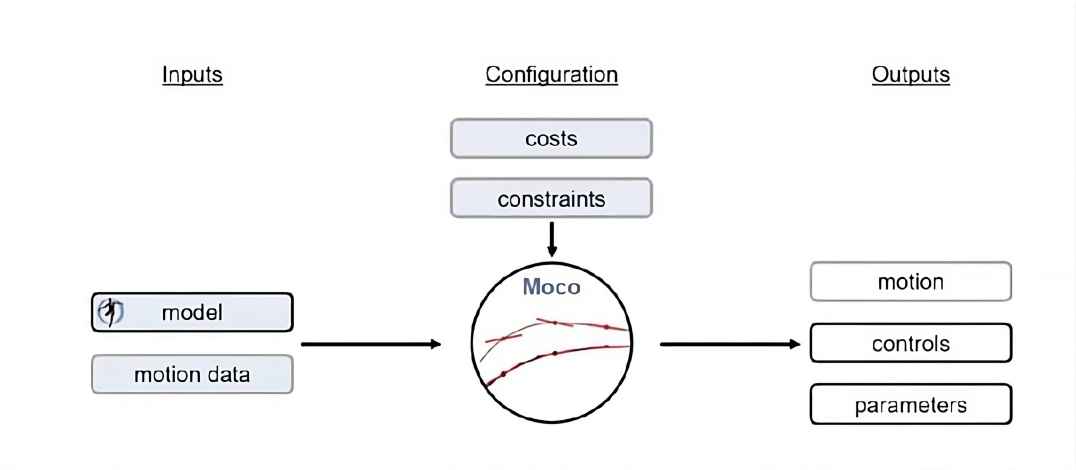

FIGURE 2.5: Problem definition used by Moco [38]

Typically, solving optimal control problems in their simplest form uses a form of trajectory optimisation called single shooting, which involves finding solutions to the initial value problem for different initial conditions until one finds the solution that also satisfies the boundary conditions of the boundary value problem [39].

A more popular method in recent times, however, is direct collocation, which is much faster. Direct collocation avoids the need for time-stepping integration and allows a more easily configurable trade-off between accuracy and computational cost compared to direct multiple shooting [37]. Direct collocation produces a nonlinear program in which the states are introduced as variables, and the system dynamics are enforced as constraints.

To solve optimal control problems with musculoskeletal models, which often lead to optimisation problems with thousands of variables, Moco uses direct collocation. The downside of implementing this manually would be the bookkeeping associated with solving a problem with that many variables. By using Moco, the downsides of the method itself are mitigated.

Consequently, the exact step-by-step process of implementing a direct collocation method to solve the optimal control problem falls outside the scope of this review. However, it is important to understand the formulation of the problem that Moco solves to maintain control over the simulation and ensure the solution aligns with the expected results. This is presented in Moco as a "MocoProblem", which allows for the creation of custom cost functions as well as custom bounds for the trajectory optimisation problem. Figure 2.6 provides a brief overview of a MocoProblem.

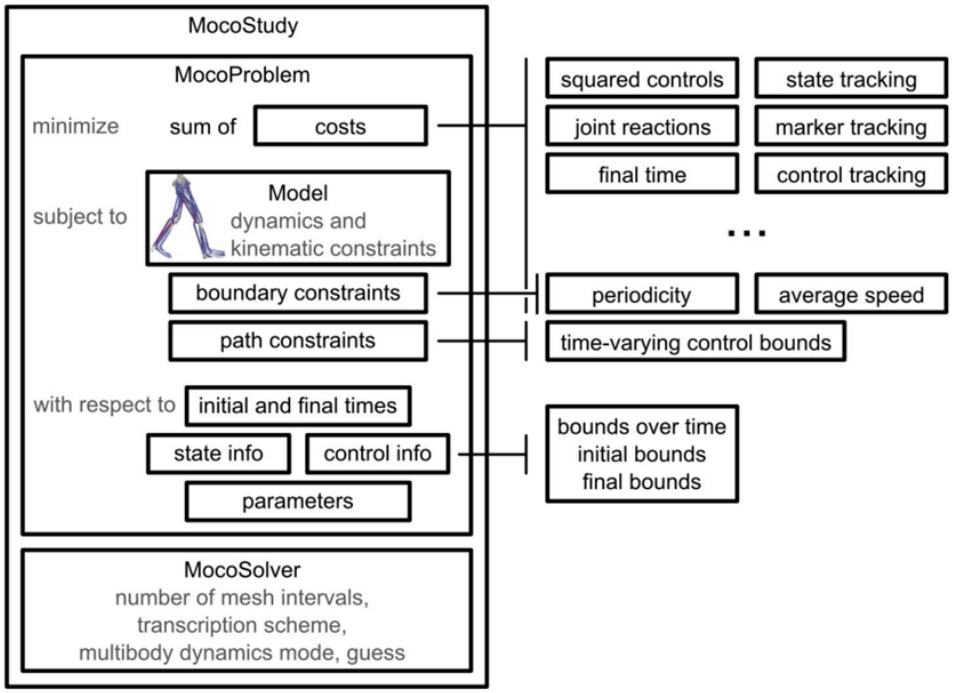

FIGURE 2.6: Overview of a MocoStudy [38]

2.9 Cost Terms, Dynamics, and Constraints

Computational modelling and optimisation of musculoskeletal motion require careful consideration of dynamic principles, anatomical constraints, and performance objectives. The MocoStudy framework addresses these challenges by formulating motion optimisation as an optimal control problem, enabling the customisation of cost functions, dynamic constraints, and boundary conditions to simulate physiologically plausible movements. This section outlines the foundational components of such optimisations.

2.9.1 Cost Terms

In a MocoStudy, a weighted sum of several cost terms can be minimised by appending terms to the initial optimisation problem. Notable examples include:

- **MocoControlGoal**: Minimises the sum of squared controls.
- **Deviation from Observed Motion**: Minimises the deviation of the model's motion from experimentally observed motion.
- **Joint Reaction Loads**: Minimises the joint reaction forces or torques.
- **Motion Duration**: Minimises the duration of a motion to meet specific timing constraints.
- **Other Costs**: Any other custom cost terms can be incorporated as needed.

2.9.2 Multibody Dynamics, Muscle Dynamics, and Kinematic Constraints

OpenSim models are a widely used format for modelling musculoskeletal systems, and Moco utilises these models to obtain the underlying differential-algebraic system of equations. These equations encompass:

- **Multibody Dynamics**: Governs the motion of bodies interconnected by joints.
- **Muscle Dynamics**: Includes muscle activation dynamics and tendon compliance.
- **Kinematic Constraints**: Useful for modelling anatomy that cannot easily be described with standard joints, or when precise modelling of ligaments and cartilage is needed. Kinematic constraints are also necessary for modelling closed kinematic loops.

2.9.3 Boundary Constraints

Users can enforce boundary constraints, such as:

- **Average Speed**: Enforcing a specific average speed during motion.
- **Symmetry**: Imposing symmetry constraints between the left and right sides of the model.
- **Periodicity**: For periodic motions, boundary constraints relate the initial and final states.

2.9.4 Bounds on Variables

Users can set bounds on various variables, including:

- **States**: Limits on the state variables, such as joint angles and velocities.
- **Controls**: Boundaries on control inputs, such as muscle activations or actuator forces.
- **Initial and Final Time**: Constraints on the time intervals for the motion, which are important for motion duration and synchronisation.

These terms are relevant to the study and provide the flexibility needed for controlling and optimising musculoskeletal simulations.

2.10 Previous Work on Musculoskeletal Models for Gait Analysis and Prediction

The problem of optimising gaits is not confined solely to inverse dynamics. As such, the relevant literature regarding the use of OpenSim and Moco for solving such problems may provide useful insights for addressing the challenges at hand.

Moco itself is relatively new, having been released within the last five years. Due to its recent development, many problems that may benefit from its capabilities have

not yet been fully explored, meaning that Moco presents potential for enabling novel work in biomechanical optimisation.

Despite being in its early stages, Moco has already seen application in various studies aimed at solving problems where collecting all the necessary data required for performing inverse dynamics is not feasible. This demonstrates promise in tackling complex biomechanical problems that may not be easily addressed using conventional methods.

Two standout papers have used OpenSim Moco for gait analysis in animals. The first paper [40] employed a musculoskeletal model for the predictive simulation of an emu walking. In the absence of reliable motion capture data, the model was able to predict GRFs and muscle forces that closely aligned with expected values. However, the emu is a bipedal animal, so while this study offers valuable insights into performing dynamic analysis using Moco, its approach is more broadly applicable to bipedal gaits.

The second paper by van Bijlert et al. [41] also utilised a musculoskeletal model for gait analysis. In this study, however, the analysis was performed on a horse. While the focus of the paper is on predictive simulation, the methodology is applicable to quadrupedal gait analysis and provides useful insights for addressing similar challenges in quadrupeds.

Finally, Bottini et al. [42] explored the use of Moco for predicting GRFs in humans based on IMU-based kinematics. The results demonstrated that despite inaccuracies in motion capture, GRFs could still be estimated using optimisation techniques. Although this study focused on bipedal motion, it is promising as it shows that dynamic analysis is possible even when GRF data is not directly available.

While these studies are promising, they do not directly address the specific problem at hand. Each of these papers provides valuable insights into how optimisation can be applied in Moco to perform dynamic analysis, but they all solve a problem different to the one at hand. The application of trajectory optimisation using a skeletal model, particularly based on kinematics, is therefore a novel approach, especially when using Moco and addressing the problem from a biomechanics standpoint rather than a strictly mathematical perspective.

2.10.1 Previously Work on Dynamic Analysis of Cheetahs

It is necessary to mention that dynamic analysis of cheetahs is not entirely novel either. The works covering it are however extremely limited.

Zhang et al. [43] discuss the biomechanical analysis of cheetahs based on digital reconstruction. . However, their analysis relies on a single video, making the motion capture strictly planar and limited to a single-camera perspective, which is insufficient for accurate 3D motion tracking. Additionally, their joint torque estimations are derived directly from manipulator equations, without accounting for GRFs. As such, their stance phase calculations are unreliable and cannot be used for comparison. Instead, their findings are only relevant for analysing the flight phase, where no ground contact occurs.

Da Silva et al. [44] however,considers torque estimation along with force estimation. While their approach to torque estimation is more robust than the method in [43], their method relies on initial GRF data for contact estimation. This GRF data is sourced from [45], meaning that their torque estimates are heavily dependent on pre-existing GRF measurements. Consequently, their force estimations are essentially reverse solutions constrained by the input data. For that reason, while their research aligns with the objective of torque estimation, their methodology and reliance on predefined GRF data differ significantly from the approach taken in this project.

Chapter 3

Method

This chapter outlines the overarching methodology of the project, which is organized into three main components: data collection and photogrammetric reconstruction, kinematic estimation, and dynamic estimation. Each of these components is explored in its own dedicated chapter, however this chapter provides a roadmap that not only previews the content of the subsequent chapters but also clarifies how these processes interrelate and support the overall analysis.

The discussion begins with the data collection process, which is essential for establishing a reliable foundation for the project. Initially, the configuration and verification of a multi-camera system are described, with preliminary 3D computer graphics simulations playing a crucial role in optimising the camera setup. The chapter further explains the pipeline used to transform these video frames into detailed point clouds through photogrammetric reconstruction techniques.

Following the data acquisition, the chapter covers the kinematic estimation process. This section discusses the use of "virtual markers" and a skeletal model to facilitate the inverse kinematics process in OpenSim. This section details the process of going from raw 3D point clouds to a detailed motion model of the cheetah.

The final section of the methodology focuses on dynamic estimation, where trajectory optimisation is employed to predict ground reaction forces and joint moments in the absence of direct force measurements. This outlines the optimisation strategy that ensures dynamic consistency of the initial motion capture, effectively bridging the gap between kinematic data and an estimation of the animal's dynamics. This section explains optimisation strategy employed to achieve dynamic consistency from the motion capture data in order to estimate the animals' dynamics. This final stage integrates the preceding processes and ties together data collection, kinematic analysis, and dynamic modelling into a coherent methodological framework.

3.1 Data Collection

The goal of this thesis is ultimately to develop a methodology for performing kinematic and dynamic estimation on cheetahs in-the-wild. Traditional motion capture systems, such as markered mocap are deemed to be too invasive and infeasible for use in uncontrolled environments, especially with untrained animals. Consequently, this requires some form of markerless motion capture.

Dynamic estimation, requires kinematics with very high levels of accuracy. This means that the standard markerless motion capture approaches, taken by previous attempts at cheetah motion capture, are often subpar or totally unusable for further analyses. For this reason, the use of a photogrammetric approach to transform the traditional markerless motion capture problem to a markered one is proposed. 3D photogrammetric reconstructions however, present their own unique challenges separate from motion capture or 3D key point estimation.

To this end, the development of a new method for estimating the joint positions and pose of cheetahs in the wild requires the collection of data that aligns with the method for 3D reconstructions of the animals' surface. Photogrammetry relies heavily on texture and as such cheetahs provide an ideal subject for performing photogrammetric reconstructions on due to their unique skin pattern. That said, performing 3D reconstructions of the animals' surface requires a very specific setup to obtain usable data.

The first step to this aim, before experimental data collection can be carried out, is to determine an appropriate camera setup. In this case, simulation using 3D computer graphics software is used to simulate an approximate representation of the presumed environment where the experiment is to take place. Following simulations, the camera system was tested using dogs as test subjects and the system was further adjusted and verified in simulation. Following the verification of the system in simulation and in testing, data collection took place at the Ann Van Dyk Cheetah Centre in North West, South Africa, a facility dedicated to cheetah conservation and research. The setup itself consisted of 14 Gopro hero 12 black cameras[1], each recording at 2.7k resolution at 240 frames per second. Alongside running trials, walking trials were also recorded as both motions lie within the scope of the project. The cheetahs were observed running on six days and walking trials were collected on three days. The result was 12 running motions captured alongside 15 walking motions captured. A photograph of the setup used for the running trials is shown in Figure **??**.

3.1.1 Point Cloud Synthesis

Following the data collection process, the video data was curated, processed and formatted for generating 3D point clouds. Figure 3.1 illustrates the processing pipeline for the generation of 3D point clouds.

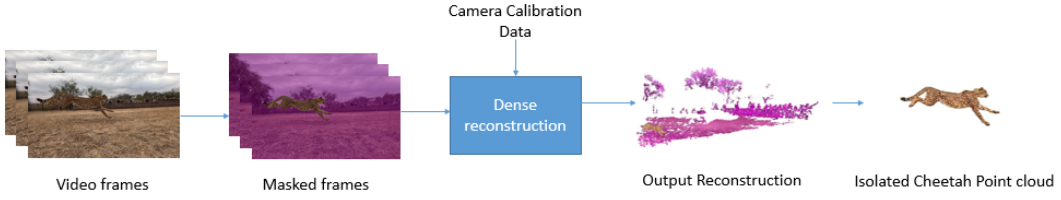

FIGURE 3.1: Point cloud synthesis pipeline

[1] https://gopro.com/en/us/shop/cameras/hero12-black/CHDHX-121-master.html

First, the synchronised video frames are extracted and separated depending on which side of the camera array they are on (i.e left or right side). The cheetah is isolated in each image and the background is masked. Notably, the entire background is not completely masked out of the images, as the 3D point clouds are additionally used for ground plane estimation. After the masking process, the images on each side are reconstructed. The camera system is calibrated initially and these parameters are further refined using COLMAP's bundle adjustment feature at the beginning of each trial to offset any deviation in camera parameters that might have occurred during the experiment due to uncontrolled conditions. Following reconstruction, the point clouds are fused based on the camera parameters. Figure 3.2 shows the point cloud reconstruction alongside the camera parameters.

Following the point cloud reconstruction phase, the masked cheetah was removed from the scene, and the final point cloud underwent a post-processing stage to clean it up. Additionally, the point clouds were used to define a ground plane. This ground plane was used to ensure the correct orientation of the scene relative to the z-axis (assuming z is upwards).

The 3D reconstructions were then evaluated on two metrics, namely silhouette completion and density, to ensure that the resulting point clouds could be used further and taken as a form of ground truth for marker approximation. A further qualitative analysis was performed as well.

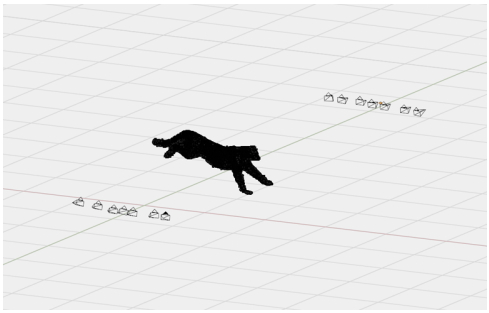

FIGURE 3.2: 3D reconstruction of entire scene

3.2 Kinematic Estimation

Following the 3D reconstruction of the cheetahs, kinematic estimation was performed using virtual markers generated from the 3D point clouds and video data. The purpose of inverse kinematics is to estimate the joint angles of a subject based on experimental data.

To create virtual markers, specific spots were strategically chosen on the cheetah's fur from the collected video data. These markers were initialised in the first frame and tracked automatically using Blenders built in tracker throughout the entire video segment. Subsequently, the markers were projected onto the 3D point cloud. This approach was preferred over the standard triangulation process, as triangulation is prone to errors if the keypoints are not accurately identified. Moreover, the marker locations for each marker are subject to varying levels of error and accuracy. By

confining the marker positions to a solution space using the point clouds, these errors were mitigated.

After the markers were projected, a marker set was created for each trial. A model for constraining the kinematics was then developed, based on the real cheetah skeleton. Joint constraints were derived both from observed motions and existing literature.

The motion capture process adhered to OpenSim's standard practices. Figure 3.3 illustrates the inputs and outputs of the inverse kinematics process in OpenSim.

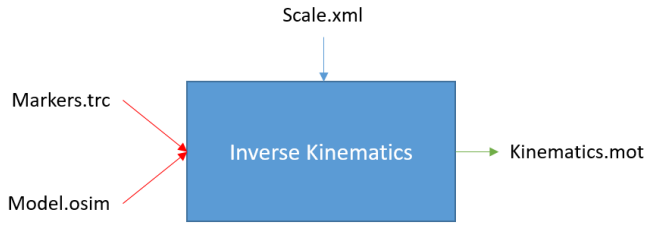

FIGURE 3.3: OpenSim Inverse Kinematics process

The primary inputs to the IK process are the following files:

- **Model.osim**: A subject-specific OpenSim model. OpenSim models are defined by an XML file that contains the model parameters. The model itself can be manually created, as done in Andrada et al. [30]. In this project, the model was constructed using a combination of OpenSim Creator [46], Muskemo [47], and manual editing of the XML file.
- **Markers.trc**: A file containing experimental marker trajectories for a trial obtained from a motion capture system, along with the time range of interest. The ".trc" file format is native to OpenSim. However, due to the widespread use of OpenSim in biomechanics research, the ".trc" format can be easily converted to industry-standard formats such as ".c3D" or VICON.
- **Scale.xml**: A file containing all the settings information for the IK tool, including the framerate, trial length, and marker weightings. Marker weightings allow certain markers related to specific body parts to be prioritised during the optimisation process.

The output is a single file:

- **Kinematics.mot**: A file stored in OpenSim's native format. The joint angles for each body part are described using Euler angles [48] Notably, the angles in the ".mot" file are relative, not absolute. The default unit is radians, as opposed to degrees.

Following the inverse kinematics process, joint angles derived from IK were evaluated for biomechanical plausibility using a periodicity and variance analysis. A sensitivity analysis was also performed to analyse the effect any scaling or manual adjustment might have on the final output. Given that there exists no true measurement of the

cheetahs' actual joint positions, these results are analysed stricly for the purpose of assessing the method's validity

3.3 Dynamic Estimation

The ultimate goal of this project is to provide a method to perform dynamic estimation of the cheetah based on some initial kinematics. Traditional inverse dynamics methods, which provide accurate solutions, rely on precise kinematics and external forces. In the absence of GRFs, however, the problem becomes significantly more challenging.

To address these challenges and the lack of GRFs associated with the kinematic data, trajectory optimisation is employed to generate physically realisable motions from the captured data. This approach enables a reasonable prediction of GRFs and, subsequently, joint moments.

For this purpose, OpenSim's Moco was used to formulate the trajectory optimisation problem. A kinetic model based on CT scan data of a cheetah was developed to associate rigid body parameters with the previously created skeletal model. Once the kinetic model was established, an optimisation problem was formulated to solve for optimal control.

Directly tracking the kinematics and using them to solve for joint moments and external forces through a single stage of optimisation would result in a dynamic solution requiring large residual forces to maintain system consistency. To address this issue, the optimisation process is divided into two stages, as illustrated in Figure 3.4.

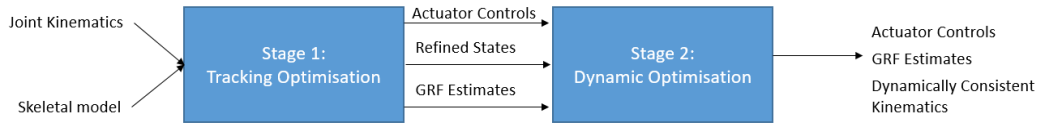

FIGURE 3.4: Optimisation Process

In the first stage, the optimisation is formulated as a tracking problem. Initial kinematics are used to solve for actuator controls, refine the kinematics, and determine system states, providing an initial guess for the second stage.

The second stage ensures the system maintains dynamic consistency. Trajectory optimisation, when formulated as a predictive solver, requires an initial guess for all states and controls. The results from the first stage serve as the initial guess for this second stage. Here, a modified cost function is used to calculate GRFs and control actuation, enabling the final computation of joint moments.

Due to the lack of ground truth data from measurements, the results for both GRFs and joint torques are analysed qualitatively based on previous literature to confirm the GRF profiles matched expected quadrupedal GRF patterns. This qualitative

analysis evaluates whether the outcomes align with expected real-world behaviour, providing insight into the method's validity.

3.4 Software

The development of this project involves a diverse range of software and libraries. This section provides an overview of the main software tools and libraries used throughout the project. The first half of the project is primarily developed using Python, while the sections requiring OpenSim and Moco are implemented using the MATLAB [49] API for OpenSim and Moco.

Software Used

- **Blender** [50] : An open source computer graphics software used for 3D modelling, animation, and rendering.
- **COLMAP** [23]: A popular implementation and library for SfM and MVS problems.
- **MeshLab** [51]: A widely used software system for processing and editing 3D meshes and point clouds.
- **OpenSim** [35]: an open source software for biomechanical analyses.

Libraries Used

- **OpenCV** [52]: A cross-platform library for computer vision tasks, such as image processing, feature detection, and video analysis.
- **Open3D** [53]: An open-source library for the rapid development of software that deals with 3D data, including point clouds and mesh processing.
- **PyMeshLab** [54]: Python bindings for MeshLab, allowing automation of MeshLab's features and integration with Python-based workflows.

Chapter 4

Multi-camera Reconstruction of Cheetahs in the Wild

Creating 3D reconstructions of wild animals is a complex task, particularly in uncontrolled environments. Previous efforts, such as AcinoSet [1], made significant strides toward addressing the challenge based solely on joint positions, but ultimately failed to provide a comprehensive solution. This chapter outlines the experimental design and methodology developed to address the problem of 3D reconstructions of animals in the wild.

The approach involves designing a setup capable of creating a full 3D reconstruction of cheetahs using photogrammetry. The cheetah's distinctive spotted coat provides an advantage, enabling the application of MVS methods to reconstruct the animal's surface by exploiting the fur pattern. The focus of this chapter is on the techniques used to capture video footage suitable for photogrammetry and on synthesising point clouds from the video data.

The chapter is organised as follows: it begins with a discussion of the methodology behind the experimental design, which uses simulated data to determine the necessary setup for video capture. The process of reconstructing point clouds from the captured videos is then described. Finally, both a quantitative and qualitative analysis of the point clouds is presented to evaluate the effectiveness of the proposed method.

4.1 Experimental Design

Designing experiments to collect video data in uncontrolled environments is challenging due to factors related to performing such an experiment in the wild. To overcome the difficulties of designing an experiment with this uncertainty in mind, simulations were employed as a foundational step before deployment in real-world scenarios.

This section outlines the experimental design, beginning with simulations used to iterate on the setup and validate the feasibility of the approach. These simulations provided a controlled framework to refine camera placements, relative positioning, and data capture strategies. Following the simulations, the setup was tested in a semi-controlled environment using dogs. This phase allowed for practical validation of the methodology, consideration of unforeseen factors, not accounted for in the

simulations, and further refinement of the experimental procedure before applying it to wild cheetahs.

4.1.1 Simulations

Simulations serve as an essential first step in validating the use of photogrammetry for 3D reconstructions of cheetahs. Key factors to be determined include the number of cameras, their placements, and the feasibility of achieving results that can extend beyond simulated scenarios to function effectively in real-world environments.

Blender[50], a comprehensive computer graphics (CG) software, was employed for this purpose. Its capabilities include creating CG scenes and automated rendering capabilities, making it an ideal tool for simulating complex setups. Furthermore, Blender supports various camera models, allowing the simulations to closely mirror real-world conditions. To facilitate the simulation process, a CG model of a cheetah was used[1]. Figure 4.1 illustrates the simulated setup in Blender.

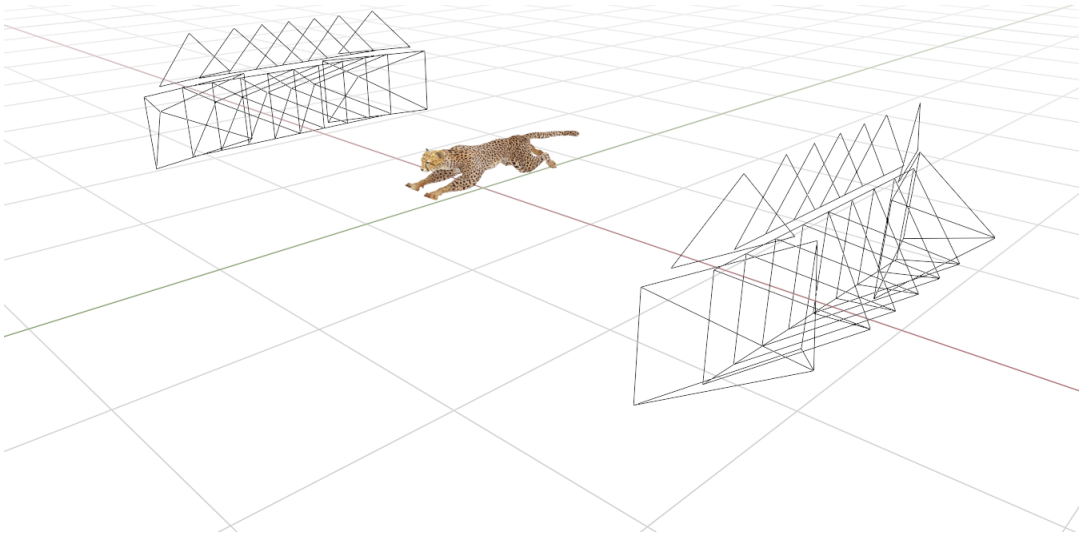

FIGURE 4.1: Pre-render of the simulated scene in Blender

4.1.2 Optimal Camera Number

Using the simulated scene in Blender, the optimal number of cameras was determined by varying the number of cameras positioned on a single side of the setup. At each step, the scene was rendered for every camera in the array. The rendered images were then processed using COLMAPs' automatic reconstruction feature and a point cloud was synthesised. The highest quality setting defined by COLMAP was used for this step.

The quality of the point cloud was assessed at each instance and compared across setups. Quality was quantified by the number of points in the reconstructed point cloud, serving as a proxy for the area covered in the reconstruction. This analysis is visualised in Figure 4.2, which shows the relationship between the number of cameras and the point cloud quality as represented by vertex count. This provides a

[1] https://www.cgtrader.com/3d-models/animals/mammal/cheetah-f862e1d0-4c6c-4cd6-907d-87da89eb9f11

basis for justifying the purchase of cameras up to the point of diminishing returns. From this analysis, it was determined that six cameras is the optimal number for the setup. This decision is based on the observation that vertex improvements plateau but the cost increases. However, to account for a safety factor, an extra camera was added, bringing the total to seven cameras per side.

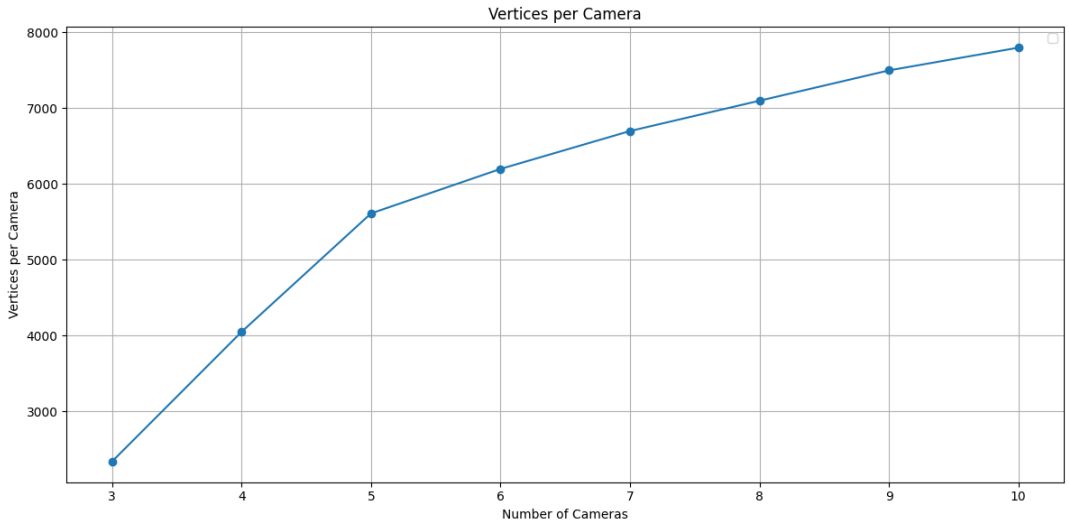

FIGURE 4.2: Vertex count per camera

4.1.3 Optimal Camera Placement

Determining the optimal camera placement was a crucial step in ensuring the feasibility of the experiment. Notably, there was inherent uncertainty in the real-world implementation as the exact conditions were uncertain. Simulations were used once again to establish an approximate layout that could be adapted to the real-world scenario.

The GoPro cameras utilised in this setup have an 80-degree horizontal field of view (FOV) when filming in linear mode. Additionally, according to Hudson et al. [55], the stride length of a cheetah at high speed is approximately 6 meters. Based on this, the cameras were arranged to accommodate this stride length. Using the determined number of cameras, the cameras were positioned 3 meters away from the cheetah's central running line, with an average baseline of 15 cm between adjacent cameras.

A critical constraint in the setup was that the cameras could only be positioned side-on as the cheetah passed through the centre of the capture area. This restriction prevented the placement of cameras in front of or above the cheetah, limiting the ability to capture a complete 3D point cloud directly. However, cheetahs have a slim profile, and by calibrating the two side-view camera arrays, the separate point clouds from each side could be fused to generate a comprehensive 3D reconstruction. Figures 4.3 and 4.4 illustrates the camera setup and the resulting fused point cloud.

FIGURE 4.3: Simulation renders

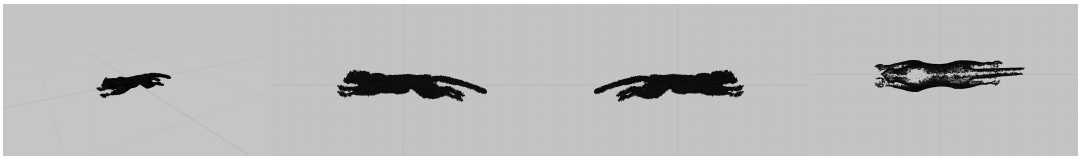

FIGURE 4.4: Point cloud reconstruction of simulated images

4.2 Real World testing

After finalising the decisions regarding the experimental setup, the next step was to test it in a slightly less controlled environment and to validate its performance under real-world conditions. The setup was adapted for use with German Shepherd dogs, which, while less textured than cheetahs, provided an opportunity to assess the system's ability to capture motion and generate point cloud reconstructions.

These tests highlighted several challenges. The cameras exhibited a tendency to overheat during outdoor use, primarily due to the high computational demand and strain on the internal batteries. Additionally, lens flare was observed in bright outdoor conditions, which negatively impacted the quality of the reconstructions. To mitigate overheating, the camera batteries were removed and external power banks were used to power the cameras via the cameras USB-C connector. Custom lens hoods were 3D printed for the cameras, effectively reducing glare and improving the consistency of the captured video data.

FIGURE 4.5: Sample frame from four cameras captured during the dog runs

4.3 Cheetah Data Collection

Building on the insights gained from simulations and real-world testing, the validated system was deployed to capture motion data of cheetahs. The primary objective of this phase was to acquire high-quality video footage of cheetahs both running and walking, providing the foundational raw data for 3D reconstructions.

This section details the setup used for data collection, addressing the logistical and technical problems encountered and the strategies employed to overcome them. All

data collection was conducted at the Ann van Dyk Cheetah Centre in North West, South Africa, a facility dedicated to cheetah conservation and research.

4.3.1 Data Collection Setup

The cameras were arranged based on the simulated design, with modifications to accommodate the specific constraints of the cheetah track. The main track, which contained a lure to motivate the cheetahs to run, was approximately 2 meters wide. For safety reasons, the camera arrays were positioned at least 2.5 meters away from the track boundaries, resulting in a total distance of 7 meters between the two camera setups. This was slightly greater than the originally planned 6-meter spacing which was determined based on the cheetahs' stride length as well as the camera sensor quality.

For walking trials these safety restrictions were relaxed, enabling the cameras to be positioned closer to the track for higher-detail captures. In this configuration the opposing cameras were spaced 5 meters apart, optimising coverage and resolution for slower, more controlled movements.

Video footage for both walking and running trials was recorded at a resolution of 2.7K at 240 FPS, the maximum capability of the GoPro cameras used. Given that consumer-grade cameras were used in this study, hardware synchronisation was not possible. To ensure frame synchronisation across the camera arrays, a combination of built-in time-code sync and a flashing LED was used at the start of each recording session and before each trial. In all cases, syncronisation was performed within a frame. Further subframe synchronisation is a limitation of the cameras themselves. While this presents a potential issue, the large number of cameras mitigates large synchronisation errors, making them less significant for photogrammetry.

The impact of synchronisation on photogrammetry has been extensively studied in previous works [56–58]. These studies conclude that at sufficiently high frame rates, synchronisation discrepancies become negligible, ensuring reliable reconstructions.

The layout of the track presented several challenges during the data collection process. One of the most significant issues was the position of the cameras relative to the sun. The orientation of the track caused one side of the setup to face directly into the early morning sun, leading to glare and lens flare despite the use of custom-designed lens covers. This negatively impacted image quality for certain cameras.

Furthermore,the setup time was limited to approximately 45 minutes before each trial. This time restriction often prevented the precise alignment of the cameras to match the simulated setup exactly. As a result, deviations in camera spacing, angles, and overlap occurred. Figure 4.6 illustrates the camera setup for the running trials, highlighting the layout and orientation of the arrays.

FIGURE 4.6: Setup used for running trials

4.3.2 Calibration

Given the positioning of the two opposing camera arrays, 3D reconstruction required performing two separate reconstructions, which were then aligned based on the relative positions of the cameras. This posed a challenge for calibrating the opposing cameras, as there was no direct overlap between their fields of view, making the use of a standard calibration board impractical.

To address this, extrinsic calibration of all 14 cameras was performed in a single pass using a 0.5m $\times$ 0.5m $\times$ 0.5m calibration cube equipped with ChArUco boards[59] on each face. Figure 4.7 shows an image of the calibration object used for scene calibration. The object itself is made from an aluminium frame with the boards attached to it to prevent the boards from deforming or warping during the calibration process. The ChArUco boards used have the advantages of both marker-based and checkerboard-based calibration methods, providing robust feature detection at different camera angles. To perform the calibration procedure, the Object was moved across the capture area. For calibration, 500 frames were selected based on visibility and used as the input. This process was performed iteratively, and calibration results were only accepted when the total reprojection error was less than 0.25 pixels.

The calibration process was carried out using MC-Calib [60], a specialised calibration toolbox for multi-camera systems that facilitates the calibration of converging vision systems. This toolkit was crucial in ensuring accurate alignment of camera arrays and achieving precise 3D reconstructions. However, given that the setup is meant for photogrammetry, calibration can also be done using a calibration object without the need for frame-by-frame calibration. Given an object of known dimensions, the two-sided reconstruction can be aligned, thus providing a solution for the relative camera positions.

FIGURE 4.7: Calibration cube

4.3.3 Motion Capture

Data scarcity emerged as a significant challenge during both running and walking trials. For running trials, only three cheetah runs were conducted per week, with a maximum of two cheetahs run in a single day. This imposed a strict limit on the amount of usable footage that could be captured. Additionally, some cheetahs did not run effectively during their trials, resulting in suboptimal motion capture on certain days. This variability in performance further reduced the availability of high-quality running data.

For walking trials, the constraints were even more pronounced. Only cooperative cheetahs could be used, thus further limiting the dataset. While most cheetahs were accustomed to running along the track, far fewer were willing to participate in controlled walking trials. During these trials, each cheetah was accompanied by a handler to guide its movement and ensure it stayed within view of the camera array. Despite this guidance, the animals' natural behaviour meant that precise control over their motion was often difficult. Consequently, only a limited number of complete and usable walking cycles were recorded.

These limitations highlight the inherent difficulties of working with non-domesticated animals, particularly in wild environments. The unpredictability of animal behaviour, combined with previously mentioned constraints, highlights the complexity of obtaining large amounts of high-quality data for biomechanical studies.

In total 3 walking sessions were recorded along with 9 running sessions. Figure 4.8 shows a subset of the collected video frames.

FIGURE 4.8: Sample frames from running trials (top three rows) walking trials (bottom three rows)

4.3.4 Data Curation

The quality of the reconstructed point cloud is heavily dependent on the input video frame data. Before reconstruction, the collected video data must be analysed and curated to remove frames deemed unusable. This ensures that only data suitable for capturing the cheetah's full surface is utilised. Given the uncontrolled outdoor environment in which the experiments were conducted, several challenges compromised the quality of the video data.

Environmental factors accounted for most of the issues encountered during the experiments. One significant problem was sand kicked up by the cheetah during its sprints, which often obscured the visibility of the paws in the video footage. This issue was especially prevalent in running trials, where sand clouds rendered many frames unusable for full-body reconstruction. Another environmental challenge was lens flare, caused by the fixed setup position and variable weather conditions. Lens flare affected entire frames, making them overexposed and unsuitable for reconstruction. Additionally, grass in the area occasionally occluded parts of the cheetah, further reducing the quality of the data.

The cheetah's behaviour introduced further complications. During running trials, the lure system allowed the cheetah to run either along the center or outside the designated track. While the system was designed for the cheetah to remain in the center, instances where it ran closer to one side resulted in incomplete capture, with only a portion of the gait cycle visible. In walking trials, the handler often walked ahead of the cheetah to guide it and prevent occlusion. However, the cheetah sometimes refused to follow, leading to situations where the handler inadvertently blocked one side of the camera array, compromising the data from those frames.

These challenges highlight the inherent difficulties of conducting experiments with wild animals in uncontrolled environments. Despite these limitations, a subset of the collected data was suitable for reconstruction. Figure 4.9 illustrate examples of discarded frames due to issues such as sand obscuration, lens flare, grass occlusion, and handler interference. Out of the nine observed running trials, only 6 runs were deemed usable. Similarly, 15 walking trials were conducted, of which 11 were successfully curated for use in the reconstruction process. These limitations underscore the complexity of working with non-domesticated animals and the unpredictability of outdoor experimental setups.

FIGURE 4.9: Example frames from discarded sessions, highlighting lens flare, sand occlusion, excessive proximity to the camera, and handler obstruction

4.4 Point cloud synthesis

Following data collection, the next step was to synthesise point clouds from the recorded data by transforming synchronised footage into detailed 3D reconstructions of the cheetah.

The frames are processed using Sfm and MVS techniques to generate dense point clouds. The cheetah's textured coat helps match the initial features, thus improving the density of the reconstruction. Post processing after the point clouds are synthesised filters out noise and artifacts, resulting in a point cloud that captures only the cheetah's geometry. This section details the aforementioned steps used to generate and refine the point cloud data.

4.4.1 Overview of Point Cloud Generation

Given the complexity of generating point clouds based through photogrammetry, the process was based on previously implemented solutions addressing the MVS problem. Specifically, 3D point cloud data was generated from the synchronised frames using COLMAP, which allowed for the automation of dense point cloud reconstruction across all frames in a sequence using custom command-line interface code. The point clouds were reconstructed and aligned using the calibration object.

Due to the uncontrolled nature of the data collection environment, potential changes in camera parameters must be considered, such as those caused by temperature fluctuations, wind or changes in camera positions for any other reason. Since the point cloud synthesis relied on known camera poses, these external factors could introduce misalignments or changes in calibration. To mitigate the resulting error, periodic calibration of the entire setup was avoided by re-calibrating the one sided camera arrays at the beginning of each motion capture cycle using SfM. Calibration points were extracted from the scene using SIFT algorithms [61]. By matching feature points across the images, the camera extrinsic parameters were solved for using a bundle adjustment algorithm, ensuring accurate alignment throughout the capture process. A similar method of using SfM for calibration been used previously in multi-camera setups, such as in the OpenMonkeyStudio dataset [6], where a multi-camera setup was used for animal kinematic estimation.

To ensure the correct relative positioning between opposing cameras, the extrinsic parameters obtained via SfM were robustly aligned to the extrinsic parameters calculated from the initial calibration run. This approach ensures consistent global scale despite the inherently scale-ambiguous nature of SfM/MVS reconstructions. This method also generalises to setups where in-field calibration is impractical after the initial setup or where calibrating a large multi-camera setup is infeasible. In such cases, an object of known size, such as the calibration cube used in this project, can serve as a reliable reference for calibration after the initial setup.

4.4.2 Preprocessing: Cheetah Masking

It is important to note that the MVS methodology used to generate point cloud data from images reconstructs the entire scene. This differs significantly from the simulated

reconstructions, where backgrounds are already segmented out. The presence of the entire scene in the reconstruction process poses a challenge, as MVS algorithms reconstructs every matching pixel. This means that for any given image set, all matching pixels, whether from the background or the foreground, are reconstructed.

Isolating the subject (i.e. the cheetah) from the background after reconstruction thus is needed to separate the subject in the final point cloud reconstruction. While this step is not absolutely necessary for purely reprojecting image points onto point clouds, it becomes essential when dealing with sparse point clouds, as meshing the entire scene can present significant challenges from a topological standpoint.

To address this, the cheetah was masked using Segment-anything-2 (SAM2) [62]. SAM2 offers an advantage over other segmentation models of being able to perform consistent segmentation across video frames. This means that only the initial frame needs to be annotated. The remaining frames are automatically segmented, thus streamlining the masking process. Additionally, SAM2 allows for further refinement of the segmentation at discrete frames to ensure consistency over the entire sequence. Figure 4.10 illustrates masked frames from both walking and running trials. Notably, the background is still intact in the final image. While it is typical to remove the non-subject area, in this case, the background provides vital information about the ground plane, and is thus kept in the image for reconstruction.

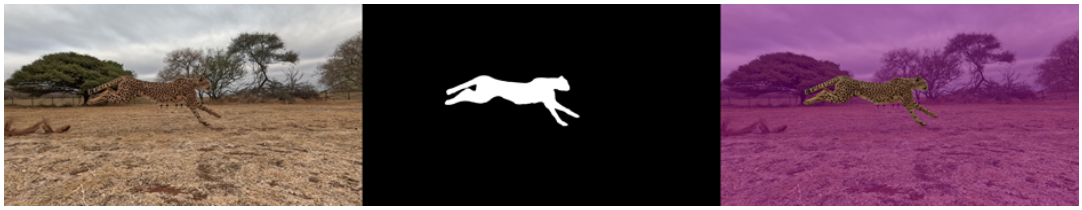

FIGURE 4.10: Image masking process

4.4.3 Point Cloud Synthesis

Following the preprocessing and data curation steps, the final stage is point cloud generation. Typically, photogrammetry pipelines, such as COLMAP, rely on unknown calibration priors for each camera, meaning that each iterative set of frames would have its own unique set of calibration parameters. However, since the goal is to derive the animal's pose from video sequences, maintaining scale consistency across the entire sequence is essential. To achieve this, the calibration parameters are kept consistent across all frames in a given run or walk. This ensures that scale is preserved throughout the sequence, and the camera parameters are not iterated upon during the course of a run. Scale drift is thus eliminated by keeping the camera intrinsics and extrinsics constant throughout each trial.

The 3D point cloud data is generated from the synchronised frames using the algorithms described by Schönberger [63]. COLMAP employs a patch match stereo algorithm to compute depth maps for each image by analysing overlapping views and estimating depth hypotheses for every pixel. These depth maps are refined iteratively using photo-consistency metrics and multi-view constraints to ensure accurate correspondences across views. The process relies heavily on matching features across

multiple images and, by extension, on the texture of the animal's skin. Texture plays a critical role in accurately capturing the animal's body curvature, as it provides distinctive points for reliable feature matching across views.

Finally, the depth maps are fused into a unified dense point cloud through the stereo fusion process. This process projects the depth points into 3D space, combining them while removing redundant or points with very high error. The result is a detailed, colourised 3D point cloud that captures the geometry and texture of the cheetah in high detail.

The cheetah itself is separated from the point cloud based on the colourised mask. Since the cheetah is a different colour, the masking process is straightforward as the resultant point cloud is colourised. The points that match the Red-Green-Blue values of the mask are separated from the cheetah. The purple colour of the mask was thus specifically chosen for this process.

4.4.4 Point Cloud Post-Processing

As a final step, the generated point clouds undergo filtering to remove noise from the point cloud surface. This process occurs in two stages. The first stage involves reprojection error filtering, where points with a reprojection error greater than 2 pixels are removed. The reprojection error is calculated automatically during the dense reconstruction process and the error metrics for all points are stored. Thus the point clouds were inspected and based on the quality of the points, the specific threshold was chosen that balances noise removal with data retention. This aligns as well with COLMAP's default filtering range which is within 1-4 pixels. This helps ensure higher accuracy in the point cloud data by eliminating points that do not match well across views or result from incorrect pixel matches.

The second stage of post-processing involves the application of a LooP algorithm [64], which is used to remove any outliers that may remain outside of the desired reconstruction area. This algorithm further refines the point cloud by ensuring that only relevant points remain, which contributes to a cleaner and more accurate final model.

Once these steps are complete, the point clouds are considered processed and ready for further analysis or integration into the motion capture pipeline. The initial and final point cloud is shown in Figure 4.11

FIGURE 4.11: Initial point cloud and processed point cloud

4.4.5 Ground Plane Approximation

An additional problem with data collection in the wild is the orientation of the world frame. Camera calibration is typically done with either an object as a reference or a camera as a reference. This means that for experiments where the camera setup cannot be entirely controlled, as in the case of this project, the reconstructions are often misaligned with the actual ground. This problem is however, solved by using photogrammetry.

Due to the texture of the grass and sand in front of the cameras, the entire scene can be reliably reconstructed. This means that assuming a planar floor surface, an approximation of the ground plane can be derived from the point cloud. It must be noted that the assumption made is that the ground surface itself is relatively uniform. During the trials, the ground was observed to be flat with minimal long grass. Therefore, both experimentally and theoretically, the ground can be considered a plane.

By segmenting the area immediately in front of the cameras, the points in the point cloud relating to the ground can be isolated. These points thus form a rough 3D ground surface. To formally define this plane mathematically, a plane can be fit to the points using the following procedure.

The ground surface is formed by a set of 3D points $\{\mathbf{p}_i\}_{i=1}^{N}$, the best-fitting plane can be found as follows:

First, the centroid $\mathbf{c} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{p}_i$ is computed and subtracted from each point to centre the data at the origin via $\mathbf{p}_i' = \mathbf{p}_i - \mathbf{c}$.

The centred coordinates are arranged into a $3 \times N$ matrix, $\mathbf{X}$, where each column represents a point and N is the total number of points. The SVD of $\mathbf{X}$ is then computed:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}. \tag{4.1}$$

The columns of $\mathbf{U}$ are the left singular vectors, which form an orthonormal basis. The normal vector of the best-fitting plane is given by the third column of $\mathbf{U}$.

This SVD-based approach is particularly effective for noisy point clouds that deviate from a perfect plane, as it minimises the sum of squared orthogonal distances from the points to the fitted plane.

The final step involves reorienting the entire scene to align the best-fit plane with the reference plane $z = 0$. This alignment is achieved by constructing a rotation matrix from to plane normals, providing a canonical orientation for further analysis of the reconstructed point cloud.

Over a trial, the plane fitting always produces a correlation coefficient of over 0.98. This means that despite the initial noise on the ground plane point cloud itself, the method remains robust to noise. Conclusively, the fitted ground plane is taken as the true ground plane.

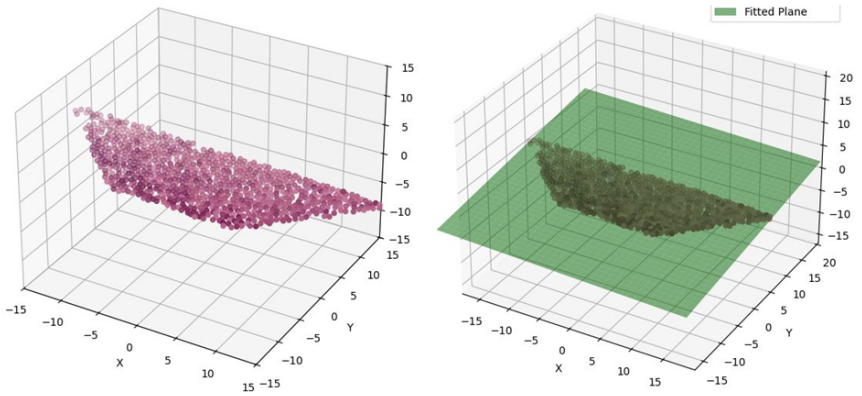

FIGURE 4.12: Ground plane fitting for a point cloud

4.5 Results and Analysis

Two methods are proposed to analyse the results of point cloud synthesis. In the absence of ground truth to directly evaluate the 3D accuracy of the point clouds, alternative approaches are employed. Given that the point cloud reconstructions' purpose is to serve as a solution space for any projected markers or points, evaluating the overall quality of the reconstructed surface is vital in determining whether the method and results are sufficient for further use.

Firstly, a quantitative analysis is conducted, focusing on metrics such as the completeness and density of the synthesised point clouds. Completeness measures the extent to which the reconstructed point cloud represents the subject as compared to the silhouetted masks, while density evaluates the distribution and spacing of points across the cloud.

Secondly, a qualitative analysis is performed, involving a visual inspection of the point clouds.his subjective evaluation assesses the fidelity and realism of the reconstruction, providing insights into areas that a qualitative analysis might overlook

This section presents a detailed exploration of these analysis techniques and their application to assess the effectiveness of the proposed methods.

4.5.1 Completeness

The completeness metric evaluates the coverage of reprojected points within the valid regions defined by the mask silhouette. This measure is considered only for one-sided reprojection only and serves as a heuristic to assess the success or failure of the photogrammetry implementation. The mask delineates the region of interest (ROI) boundaries, which the reprojection aims to cover. In this case, the mask from SAM2 is taken as the ground truth measurement. SAM2's segmentation exact pixel accuracy was not manually validated as this was not possible. However, its consistency across frames was visually verified, and errors (e.g. missing paws) were mitigated during the initial masking phase.

Since the reprojection provides discrete pixel points projected onto the image plane, an alpha-shape algorithm [65] is employed to create a continuous boundary around

the reprojected points. This facilitates a direct and quantitative comparison of the reconstruction area with the original boundary, independent of point density.

The mask is extracted as a binary image where white pixels define the valid ROI. The projected boundary is overlaid onto the mask. Binary masks are then generated from the original mask and the reprojected boundary.

The intersection between the reprojection mask and the segmented mask is then computed to determine the reprojected area that lie within the valid ROI. Completeness is then calculated as the ratio of pixels in the reprojected area to the total number of valid pixels in the mask as follows:

$$\text{Completeness} = \frac{\text{Area of Reprojected boundary within mask}}{\text{Total Area of Mask}} \times 100. \tag{4.2}$$

Figure 4.13 illustrates the mask and reprojected boundary for two images, with results for walking and running trials compared. This separation is crucial because the high-speed motion in running versus the slower motion in walking could lead to different results due to differences in the camera setup, data capture process or reconstruction process. The completeness results for 550 frames are summarised in Table 4.1.

TABLE 4.1: Completeness and standard deviation results for running and walking trials

Trial Type	**Completeness (%)**	**Standard Deviation (%)**
Runs	89.52	1.83
Walks	93.81	1.20

From this analysis, it is evident that the point cloud coverage compared to the silhouette covers nearly the entire surface. The difference in completeness between running and walking trials is minimal, suggesting comparable performance across these conditions. However, while this analysis provides insight into coverage, it does not account for all factors affecting the quality of the reconstruction, particularly as it is done independent of point cloud density.

4.5.2 Density

While the completeness measurement is independent of density, it is essential to evaluate the density of the point clouds to provide a more complete analysis. The density can then be estimated as a ratio. Given a silhouette previously defined as a mask, instead of wrapping the points to measure completeness, the actual number of reprojected pixels within the mask can be measured. This is expressed as:

$$\text{Density} = \frac{\text{Total Number of Reprojected Pixels}}{\text{Number of Pixels within Mask}} \times 100. \tag{4.3}$$

The results for density in the same walking and running trials used for completeness are summarised in Table 4.2.

When compared to the completeness scores, the density results reveal a more significant difference between the walking and running trials. For the walking trials

TABLE 4.2: Density results for running and walking trials with standard deviation

Trial Type	Density (%)	Standard deviation
Runs	18.42	4.21
Walks	23.56	0.73

the density is closer to one in four pixels being reconstructed while for running it is closer to one in five. While the average density is comparable, what stands out is the standard deviation. The density scores for the running reconstructions are much more likely to be lower quality than the walking data.

To tie these results more closely to the actual scale of the cheetah and to gain an idea of density irrespective of pixel size, the reconstructed point clouds were sampled into segments in 3D space. These segments represent surface area segmentations of the point cloud itself. Since the point clouds are scaled to match the real size of the cheetah, this sampling allows for a measure of the actual real-world density. For walking trials, the density of the point clouds is approximately 11 vertices per square cm over the whole body. However, for running trials, the density drops significantly to just 5 vertices per square cm.

Comparatively, the resolution of the point clouds and the high frame rate of the cameras make photogrammetry a significantly more effective solution than standard depth cameras, such as the Kinect Azure or Intel RealSense, as based strictly on their point cloud densities [66, 67]. Given the number of cameras, their resolution, and the specific settings used, this outcome is expected, as low-cost depth cameras are not ideally suited for outdoor use.

These density measurements highlight a more pronounced difference between the two trial types. Despite having very similar completeness scores and pixel reprojection density, the walking trials exhibit nearly double the density of the running trials when measured from the real world scale, providing a clearer picture of the disparity between the two trial types. Notably, the pixel density itself did not differ too much, thus pointing to the distance impacting the measurement as well. Figure 4.13 shows the original image, the reprojection, and the masked reprojection.

FIGURE 4.13: Frames, reprojections and mask

4.5.3 Qualitative Results

In addition to the quantitative metrics discussed above, the resulting point clouds can also be evaluated qualitatively to provide a rough assessment of their accuracy. While the silhouette-based metrics offer quantitative evaluation methods, they may not always fully capture the nuances of the reconstruction and may be misleading. For motion capture, the legs are more important than the body. Since the previous metrics assess the body as a whole, a separate evaluation of the legs is not included. However, qualitatively, point clouds can be assessed to determine whether they are sufficient for further motion capture.

This subjective evaluation focuses on the visual inspection of the point clouds to determine whether they align with expected characteristics, such as surface detail, consistency and uniformity of the points. By examining the reconstructed surfaces, further insights can be gained into the success or failure of the photogrammetry implementation beyond the numerical results. Instances of the original image, reprojection and point cloud are shown in Figure 4.14.

The previously discussed density measurements were performed for the entire body. However, qualitative observations reveal that the density around the body is notably higher than that of the legs, particularly the leg in motion. This discrepancy is expected, as the body generally exhibits more texture than the legs, resulting in a more complete and detailed point cloud in these regions. Furthermore, in frames where the paw is in contact with the ground or obscured by dust, the point cloud density is significantly lower, with some segments of the paw and foot missing entirely. While this poses a challenge, the legs, till the ankle, are generally well-preserved in the point cloud.

The reduced density in the legs and paws may also be attributed to factors such as synchronisation issues or motion blur in the cameras. During running, the legs and paws exhibit the fastest motion, making these areas more challenging to reconstruct accurately. Despite these challenges, the point clouds are qualitatively observed to exhibit a high level of detail, with the overall reconstruction successfully capturing key surface features. Notably, curvatures in the body, such that around the pelvis and midriff of the animal are well defined in the point clouds. This level of detail can be attributed to the animals skin texture.

From the qualitative evaluation, the reconstructions demonstrate a near-complete representation of the cheetah. Notably, regions with little to no texture, such as large black blotches on the tail, ears, or face, exhibit patches where the reconstruction is less dense. These observations align with the initial premise of aiming to leverage the cheetah's distinctive texture to achieve more accurate reconstructions compared to an untextured animal. The results highlight the importance of surface texture in the photogrammetry process, as areas with richer texture demonstrate significant differences in completeness and density of the reconstructed point clouds.

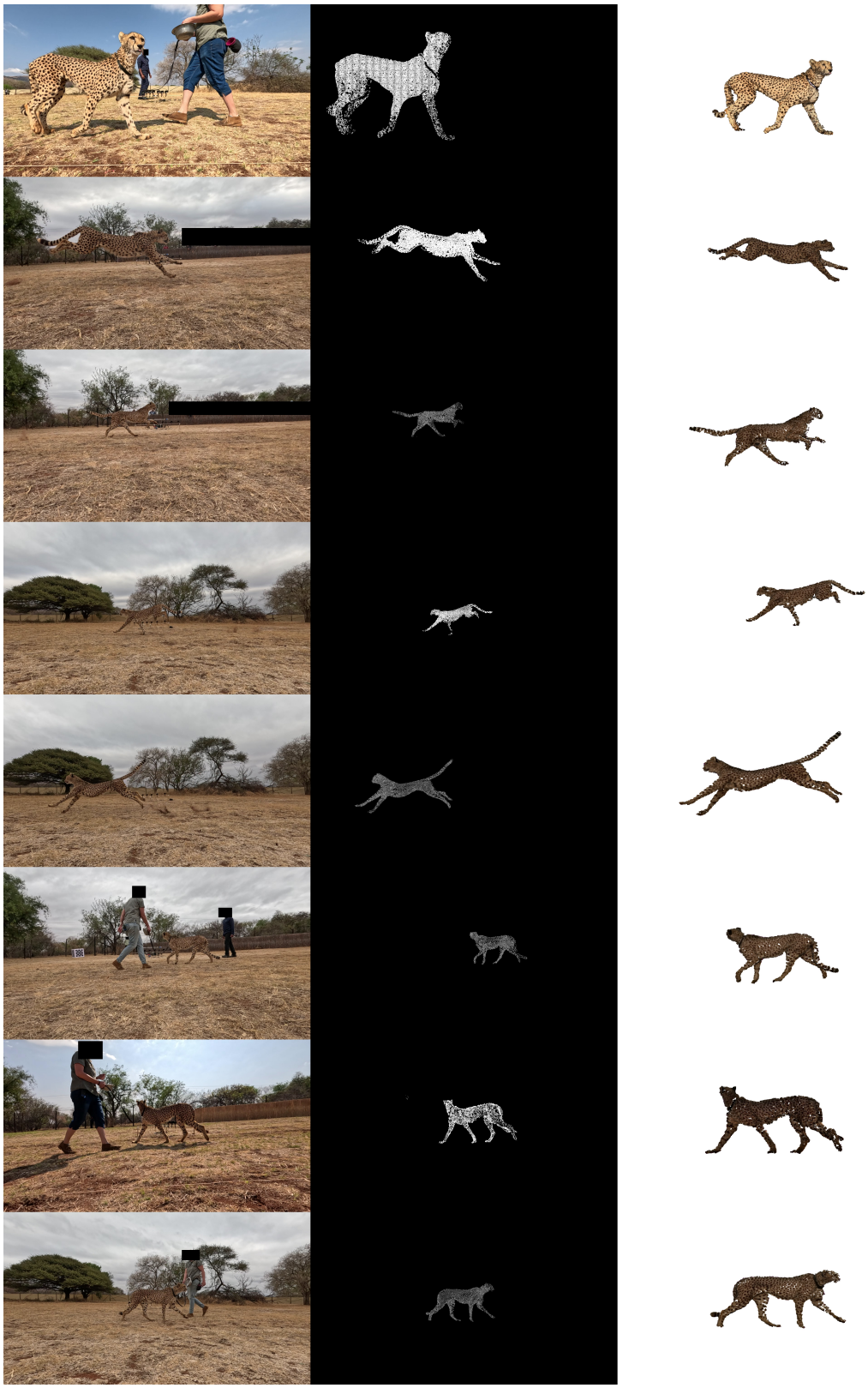

FIGURE 4.14: Video frame, point cloud reprojection and 3D point cloud

4.6 Discussion

The lack of conclusive measurements across the entire surface of the cheetah has been identified as a significant limitation for accurate pose estimation and reconstructions of the animal in general. The photogrammetry system and the resulting point clouds presented in this chapter represent a substantial step towards addressing the absence of a ground truth for 3D keypoint estimation.

The data collection process and point cloud synthesis described in this chapter have yielded promising results. Considering that photogrammetry has not previously been applied to high-speed animals, especially in natural environments and with low-cost camera setups. These findings suggest that photogrammetry can be a viable method for capturing highly detailed 3D representations of fast-moving subjects.

The density and completeness metrics of the reconstructed point clouds suggest that the results are reliable and can be accepted. In the absence of an actual 3D ground truth—which is not available in this context—the reconstructed point clouds, particularly those from walking trials, provide a highly accurate estimate of the cheetah's outer surface. These reconstructions can be regarded as a form of potential pseudo-ground truth for 3D measurements.This could be experimentally verified using more expensive depth cameras. However, given the low-cost approach of this project, such an investigation is considered beyond its scope.

The findings also highlight areas for improvement. Higher-resolution cameras and better synchronisation, especially during high-speed movements, could enhance the accuracy and quality of the reconstructions. Such improvements may lead to an even more precise approximation of the cheetahs' outer surface further advancing the utility of photogrammetry for biomechanical analyses.

Given that this approach was designed to address the shortcomings of AcinoSet, the point clouds and associated depth maps generated from them provide a solution for reconstructing the cheetah's actual surface—something not possible with AcinoSet's method or data. Unlike AcinoSet, which reconstructs sparse joint positions, our method captures the cheetah's full surface geometry. Thus, the ability to recreate the cheetah's full surface in motion pre-emptively addresses concerns raised about AcinoSet regarding the accuracy of its 3D reconstruction.

Nonetheless, the point clouds presented in this chapter address the challenge of reconstructing a complete 3D surface of the cheetah. This surface serves as a solution space for projecting markers, forming the foundation for a solution to the markerless motion capture problem.

Chapter 5

Kinematic Estimation

The state of the art for motion capture involves the use of marker clusters on limb segments [68]. While this approach works well for human subjects, applying markers to animals is often impractical. Furthermore, the absence of markers compounds the difficulty of achieving accurate 3D joint estimates, as markerless approaches generally result in lower accuracy.

Markerless motion capture, which has become a popular method for animal kinematic estimation, is heavily reliant on single-marker estimation for each limb segment. This reliance on joint markers, combined with triangulation methods for estimating 3D positions, introduces inherent limitations to the method that undermine the goal of achieving precise kinematic measurements.

To address these limitations, this chapter proposes an alternative solution to traditional markerless motion capture approach. By leveraging the initial point cloud estimates and further exploiting the unique skinned pattern of the cheetah, the markerless motion capture problem is reformulated to align more closely with state-of-the-art techniques used for human motion capture and kinematic estimation.

The chapter is structured as follows: an overview of the proposed approach is presented first, followed by a discussion on marker tracking based on the cheetah's skin pattern. The methodology for 3D reprojection of markers is then described, leading into the derivation of a 3D kinematic model. Using the identified markers, an inverse kinematics problem is formulated, and the chapter concludes with an analysis of the kinematic estimation results.

5.1 3D Marker Estimation

This method aims to enhance the accuracy of markerless motion capture for animal kinematic estimation such that it aligns with the requirements for biomechanical analysis. The method leverages the unique skinned pattern of the cheetah and initial point cloud estimates to overcome the limitations of markerless motion capture techniques.

The point clouds used in this chapter are derived from the multi-camera setup described in Chapter 4. This provides a complete surface mapping of the cheetahs' skin and thus provides a solution manifold for marker projection.

Instead of relying on physical markers, the distinctive fur pattern of the cheetah is exploited to track movement across its body and allows for spots to be tracked throughout the video frames. This is achieved through optical flow techniques enabling automation after initial marker registration. These markers are then reprojected back to the point cloud surface using known camera parameters. Following this, the initial problem has been transformed into a traditional markered motion capture approach.

Following marker estimation, a 3D kinematic model of the cheetah is constructed to represent the joint positions of the underlying skeleton. This model forms the basis for solving an inverse kinematics problem, where the joint angles and body positions that best explain the observed marker positions are estimated. By aligning the measured data with the kinematic model, a more accurate reconstruction of the cheetah's movements can be achieved. Furthermore, this step allows the motions to be constrained depending on the underlying degrees of freedom afforded to the kinematic model.

5.1.1 2D Marker Estimation

For the point clouds to be used further, discrete marked locations along the cheetah's spots need to be placed for the entire trajectory. Unlike traditional markerless motion capture methods, these virtual markers are not placed at joint centers but are instead located at spots on the cheetah's body.

Tracking these markers over an entire trajectory presents its own challenge. In traditional markered setups, infrared markers can be isolated to determine their exact positions. To apply a similar technique, the video frames are processed to identify regions with the highest levels of contrast. Thus, the distinctive spots of the cheetahs, which provide high contrast regions, are used as virtual markers. These spots are tracked using Blender's VFX framework, which employs a robust variant of the Kanade-Lucas-Tomasi algorithm [69]. Tracking is done frame by frame, using the pattern from the previous frame to improve robustness and account for skin deformations. Figure 5.1 illustrates the deformation of a single marker over time as the cheetah walks.

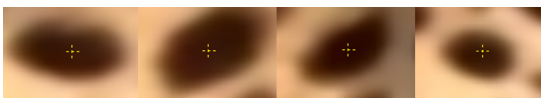

FIGURE 5.1: Marker variation of a single spot over a trial

This automated approach is significantly faster than manual marking, which takes approximately 10 minutes per 100 frames. By comparison, automated tracking is nearly instantaneous after initial marker placement. Additionally, this method provides more consistent marker placement over time than deep-learning-based approaches such as DeepLabCut.

In theory, the number of trackable spots is only limited by the number of visible spots on the cheetah. While skin displacement remains an unresolved issue, even

in traditional markered motion capture, increasing the number of markers beyond joint centers and careful marker placement helps mitigate its impact [70, 71]. Thus, to minimize the effects of skin displacement, markers are strategically placed near bony regions. Figure 5.2 shows sample marker locations for a walking and running trial along with their trajectories.

FIGURE 5.2: Video frames showing marker tracking

5.1.2 Marker projection

After obtaining the 2D marker-set, the next step is to get the 3D locations of the markers. This serves as the final step in the problem reformulation of markerless mocap to a markered mocap problem.

Given a set of marker locations in pixel coordinates, first a projection from the image plane to the 3D point location is needed. The mapping for projection from 3D world coordinates to camera coordinates to pixel coordinates is given below.

$$[X_w; Y_w; Z_w] \rightarrow [X_c; Y_c; Z_c] \rightarrow [U; V; W]. \tag{5.1}$$

Here $W = 1$. The reduction in dimensionality of the mapping becomes apparent, namely the depth loss associated with the mapping $\mathbb{R}^3 \rightarrow \mathbb{R}^2$. Thus, the projection from the image plane back to some 3D point contains a degree of ambiguity that for a given pixel coordinate $[u'; w']$ there exists a one-to-many mapping for the corresponding 3D point $[X'_w; Y'_w; Z'_w]$ along the projection ray corresponding to the same image coordinates.

Only with additional knowledge of the depth $d(x, y)$ is it possible to uniquely invert the projection process. The depth refers to the distance of the 3D point on this ray.

Given a pixel coordinate x, y, we can reconstruct its corresponding 3D ray through

1) **Normalized Image Coordinates:**

$$\begin{bmatrix} \bar{x} \\ \bar{y} \\ 1 \end{bmatrix} = K^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{5.2}$$

where K is the camera intrinsic matrix.

2) **Ray Formation:** The ray in world coordinates is defined by an origin C and direction d:

$$C = -R^\top t. \tag{5.3}$$

$$d = R^\top \begin{bmatrix} \bar{x} \\ \bar{y} \\ 1 \end{bmatrix}. \tag{5.4}$$

3) **Ray Equation:** The ray in world coordinates is parametrised as:

$$X(\lambda) = C + \lambda d, \quad \lambda \geq 0, \tag{5.5}$$

where λ is a non-negative scalar representing the distance along the ray.

To resolve depth ambiguity, the closest point on the point cloud, $\mathcal{P}$, to the ray is identified. The point cloud $\mathcal{P}$ is obtained from a depth sensor or a 3D reconstruction algorithm. Since point clouds are discrete and irregular, meshing via the alpha shape algorithm [65] creates a surface representation, which is resampled for uniform density.

The closest point, $p^* \in \mathcal{P}$, to the ray minimizes the perpendicular distance computed as:

$$p^* = \underset{p \in \mathcal{P}}{\operatorname{argmin}} \left\| (p - C) - \left((p - C) \cdot \hat{d} \right) \hat{d} \right\|, \tag{5.6}$$

where $\hat{d} = d_{\text{world}} / \|d_{\text{world}}\|$ is the unit direction vector, and $\| \cdot \|$ is the Euclidean norm. The depth $d(x, y)$ is then derived from the position of p^* along the ray as:

$$d(x, y) = \|p^* - C\|. \tag{5.7}$$

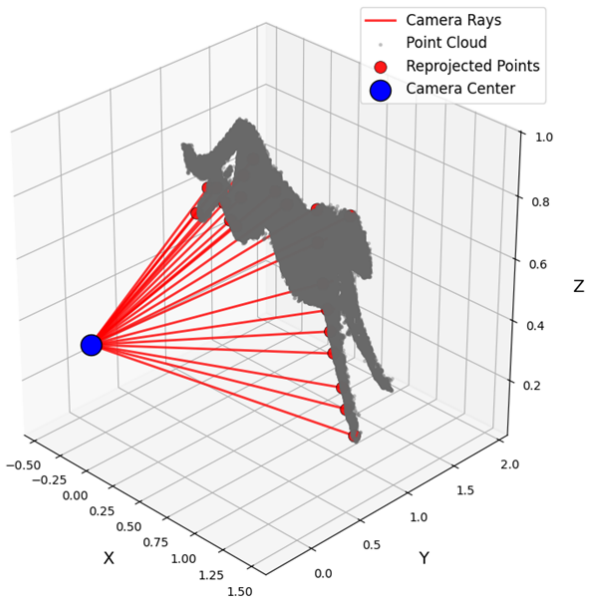

FIGURE 5.3: Image showing points reprojected from camera to 3D point cloud

5.1.3 Foot/Paw Prediction

Based on the results presented in Chapter 4, there are cases where the generated point cloud does not include the foot or paw. In such instances, directly projecting and estimating the depth of the markers leads to incorrect outcomes.

To address this issue and improve the accuracy of foot and paw marker predictions, an alternative approach is proposed. When the perpendicular distance to the closest point, as defined by Equation 5.6, exceeds a threshold, it indicates that the foot was not accurately reconstructed in the point cloud. In such cases, the nearest marker with sufficiently low error as defined by the threshold is used instead. Since the ankle or wrist marker are consistently projected accurately—due to the inclusion of the ankle region in the point cloud data—its depth value is substituted to provide an initial prediction of the 3D position of the foot or paw marker. Based on the average distance between the markers, the depth is then scaled to get a more accurate approximation than directly using another markers' depth. This method assumes that the foot and paw are approximately coplanar and that their motion relative to the ankle is not off axis. Given that the ankle joint can be approximated as a hinge joint, this assumption is reasonable.

To validate this method, trials were conducted on point clouds where the foot region was initially included. The foot region was then removed, and the proposed method was applied to calculate the 3D position of the missing markers. Over 300 frames, the average error was found to be less than 0.5 cm, confirming the method's efficacy. For all cameras where foot or paw marker prediction was utilised, the animal's motion was restricted to a straight-line trajectory.

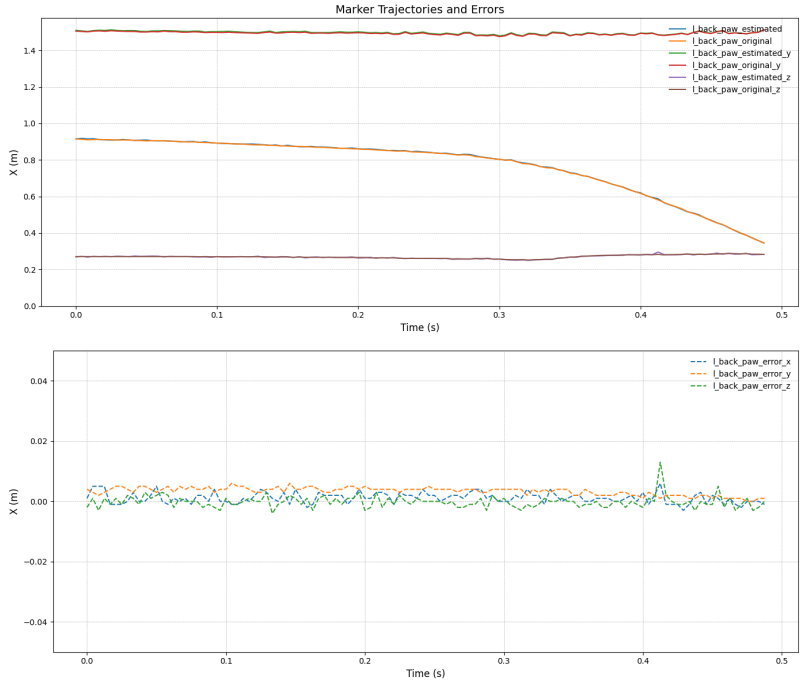

FIGURE 5.4: Ankle marker projection and prediction along with error over time

5.1.4 Marker Trajectories

Analysing the trajectories of the raw marker data provides insight into the overall movement of the markers over time, including an estimation of noise levels and variations in marker positions between frames. Figure 5.5 illustrates the trajectories of five markers during a walk spanning two gait cycles. As shown, the markers exhibit varying degrees of noise.

To mitigate this noise and prevent it from affecting the kinematic calculations, the trajectories are then filtered using a second-order Butterworth low-pass filter with a cutoff frequency of 6 Hz.

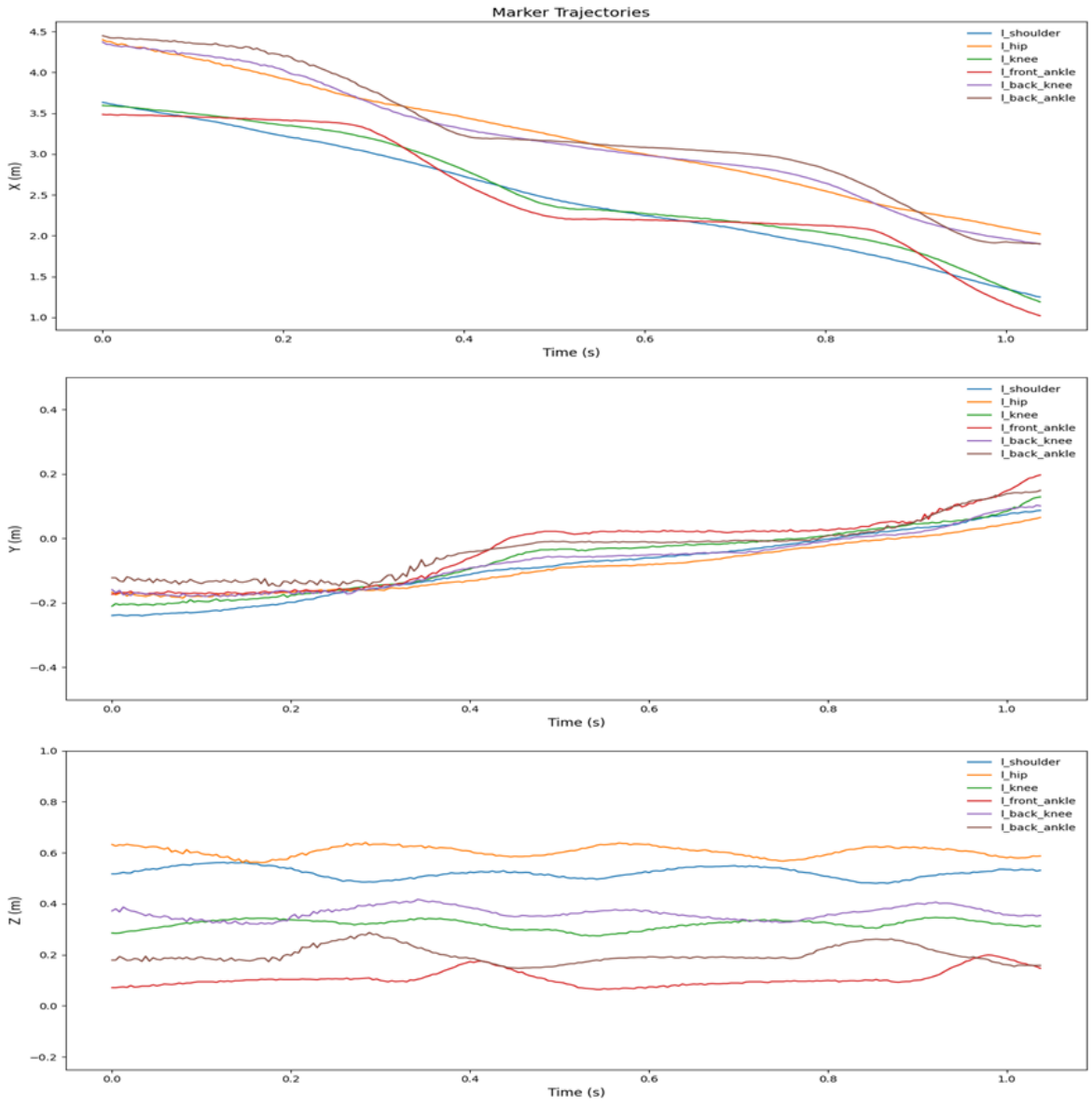

FIGURE 5.5: Marker trajectories in the x, y and z axis(z upwards)

5.2 Kinematic Model

Tying the marker set to a realistic model is necessary for further biomechanical analysis. To ensure that the model accurately reflects the internal skeletal geometry of a cheetah, a skeletal model based on a real cheetah is used as the basis for kinematic estimation. This section outlines the process of developing the cheetah's kinematic model.

5.2.1 Skeletal Model

The cheetah skeletal model was constructed using computed tomography (CT) data from two adult cheetahs. The first CT dataset [72] consisted of 1982 slices (spacing 0.625 mm), with a resolution of 514 × 510 pixels (spacing 0.841 × 0.841 mm), covering the entire body of the cheetah. The extraction of the skeleton was performed using the Amira3D software package, and both the original CT data and the extracted skeleton are shown in Figure 5.6.

Due to the poor quality of the full body CT scan, the extracted bones were often noisy and fused with nearby bones. To address this issue, a second, higher-resolution CT scan was used[1]. This scan focused exclusively on the cheetah's bones, with a higher resolution, which allowed for better separation of bones without interference from surrounding muscle or fat. The bones extracted from the second scan were then aligned with the original, full-body CT scan, resulting in a more accurate and clearer skeletal model.

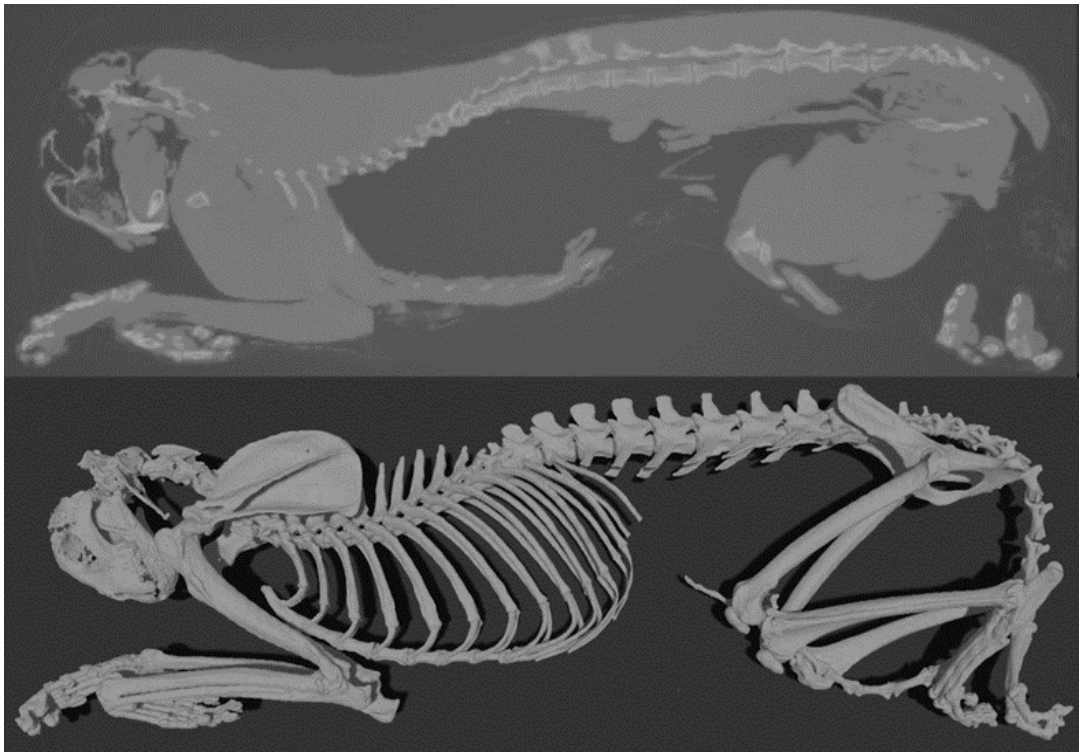

FIGURE 5.6: CT scan along with the segmented skeleton

[1]University of Wisconsin Zoological Museum | Catalog Number:UWZS 23961

5.2.2 Joint Centers

Accurately defining joint parameters is a challenging task, primarily due to the inherent variability and complexity of biological structures in skeletal models. The joint centers, serving as the center of rotation for the defined degrees of freedom, must be precisely located. Small errors in its position can propagate through the kinematic chain, amplifying inaccuracies in joint angle calculations and leading to significant deviations when performing further biomechanical analyses, such as force estimations and inverse dynamics.

To address this, the high-resolution skeletal geometry was used as the foundation for defining the joint centers. This detailed geometry provided an anatomically accurate framework for identifying and refining joint parameters. The process entails segmenting the bones of interest from the skeleton, followed by fitting geometric primitives—such as spheres and cylinders—to the segmented bone sections.

For ball-and-socket joints, like the hip, spheres were used to model the joint center, ensuring precise alignment with the curvature of the femoral head. In contrast, hinge joints like the elbow, which rotate along a single axis, were modelled using cylinders to define the axis of rotation. Figure 5.7 illustrates this process, showing both the original bone geometry and the fitted geometric primitives. The fitting process employed a least-squares method to minimise error between the skeletal mesh data and the modeled shapes.

By leveraging high-resolution skeletal geometry, joint centers and rotation axes were defined with a level of accuracy far surpassing what could be achieved through manual placement. This systematic approach not only improved the fidelity of the kinematic model but also minimises error propagation throughout the kinematic chain.

FIGURE 5.7: Joint centers for the femur and arm/scapula

5.2.3 Joint Constraints

The primary goal of developing a model for kinematic estimation is to ensure that it accurately reflects reality. For this reason, the model's joint angles are based on anatomical measurements of cats and are not arbitrarily chosen. Additionally, by

analysing the motions and initial joint angles, these joint constraints were further aligned with the actual joint range of motion found in the cheetah. Properly defined joint constraints are crucial for ensuring realistic motion and preventing implausible movements. These constraints are further adjusted according to the specific application of the model, in this case biomechanical analysis. For example, some joints are constrained to motion along the sagittal plane only, reducing their degrees of freedom from three to one.

The cheetah kinematic model consists of 24 joints and 27 degrees of freedom. Table A.1, given in Appendix A, shows the model's joints and their associated limits. Each joint is restricted within a specific range, reflecting typical joint motion found in other cats, as referenced in [73–77] as well as the joint constraints observed from the colected video data. These ranges of joints and degrees of freedom are crucial for ensuring that the model's movements are physically realistic and anatomically accurate. The complete skeletal model is shown in Figure 5.8.

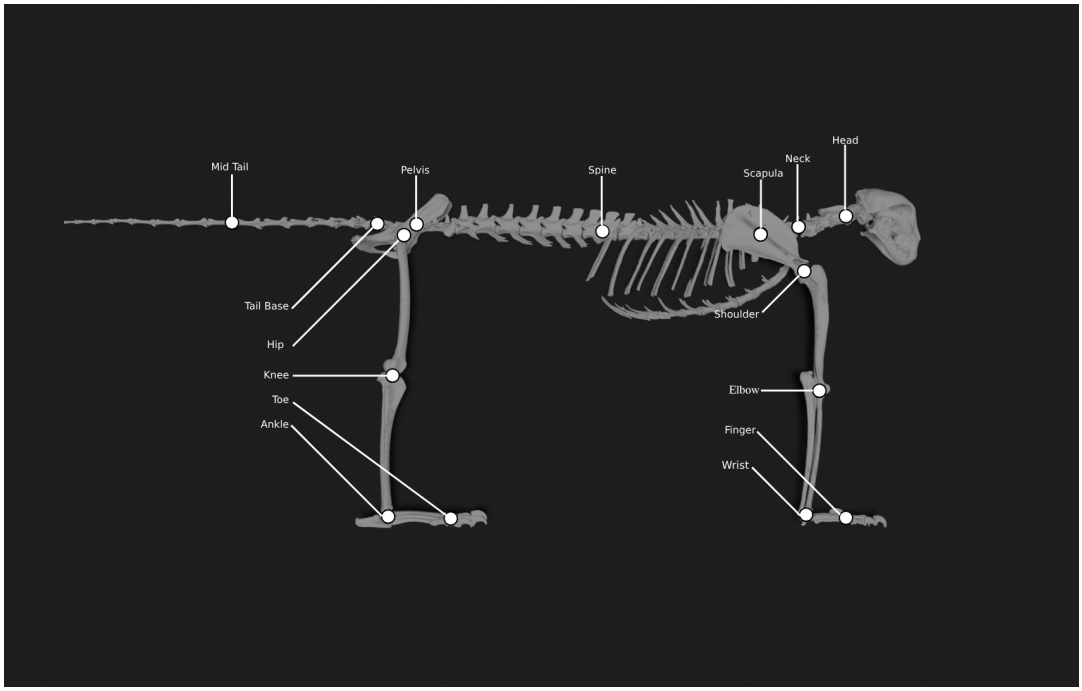

FIGURE 5.8: Kinematic model of the cheetah

5.3 Kinematic Estimation

The final step in preparing the mocap data for further analysis is the conversion of 3D marker data into kinematics, i.e joint angles. This conversion requires a kinematic model that includes the internal joints of the cheetah as well as 3D marker locations, both of which were obtained as previously described. This section outlines the process of converting marker trajectories into kinematic estimates using a body model in OpenSim, through inverse kinematics.

5.3.1 Inverse Kinematics

For traditional markered motion capture problems, converting markers to joint angles and body positions is a standard problem. Using the previously defined model, the constrained kinematics can be calculated from the marker data using OpenSim. Given that OpenSim is a biomechanical analysis software, the inverse kinematics step is usually the first before any further analysis can be done [78].

OpenSim solves the IK problem by solving a weighted least squares optimisation problem. At each time step of the recorded motion data, the inverse kinematics algorithm computes a set of joint angles that positions the model in a configuration that "best matches" the experimental kinematics. This "best match" is determined by minimising the marker error, defined as the distance between an experimental marker and its corresponding model marker.

Each marker is assigned a weight, which specifies the importance of minimising that marker's error term in the least squares optimisation. At each time step, the inverse kinematics tool solves for a vector of generalised coordinates (i.e. joint angles), that minimises the weighted sum of marker errors, as expressed by:

$$\min_{q} = \sum_{i=1}^{N} w_i \left(x_i^{\text{exp}} - x_i^{\text{mod}}(q) \right)^2, \tag{5.8}$$

where q is the vector of generalised coordinates (e.g., joint angles), x_i^{exp} is the position of experimental marker i, $x_i^{\text{mod}}(q)$ is the position of the corresponding model marker i (which depends on q), and w_i is the weight associated with marker i [78]

The result of this minimisation process is a set of coordinates that, over time, describe the positions of the virtual markers and the joint angles of the model, allowing it to closely follow the real movement of the subject.

A common practice in OpenSim is to ensure that the average RMS marker error over a run remains below 2-4 cm [78]. To achieve this, the markers and model needs to be scaled per trial.

5.3.2 Marker Placement

As mentioned, a common challenge in OpenSim, and in marker-based mocap in general, is aligning the marker placement during trial runs with the corresponding marker placement on the kinematic model. This issue has been the subject of several research papers [79, 80], which aim to quantify the uncertainty of marker placements. Typically, marker locations are estimated from static images and reconstructed in 3D. The method used here is an adaptation of the approach suggested by Dunn et al. [79]. The key advantage over 2D image-based marker placement is that the previously computed point clouds provide a more accurate representation of marker placements in 3D space. This challenge can be addressed by using 3D point cloud data.

Initially, a rough estimate of the skeleton in a single pose is fitted to the marker data. This pose can then be refined manually to ensure correct alignment with the

skeleton. Following this, the markers are scaled over the entire run to minimise the RMS marker error. This approach guarantees that the error at the scaled frame is zero, effectively minimising the error throughout the entire run based on an initial placement. Figure 5.9 illustrates the initial point cloud data, along with the final marker locations on the skeleton.

It is important to note that the model included certain joints constrained strictly to planar motion, particularly in the lower legs and arms. As a result, these motions were not exact representations of the animal's actual 3D movements but were instead simplified for biomechanical analysis. This simplification is justified given that the trials selected for walking and running involved strictly linear motion, making the approximation valid.

Despite these constraints, all trials underwent the previously described process to minimise marker tracking errors, as recommended in the OpenSim documentation.

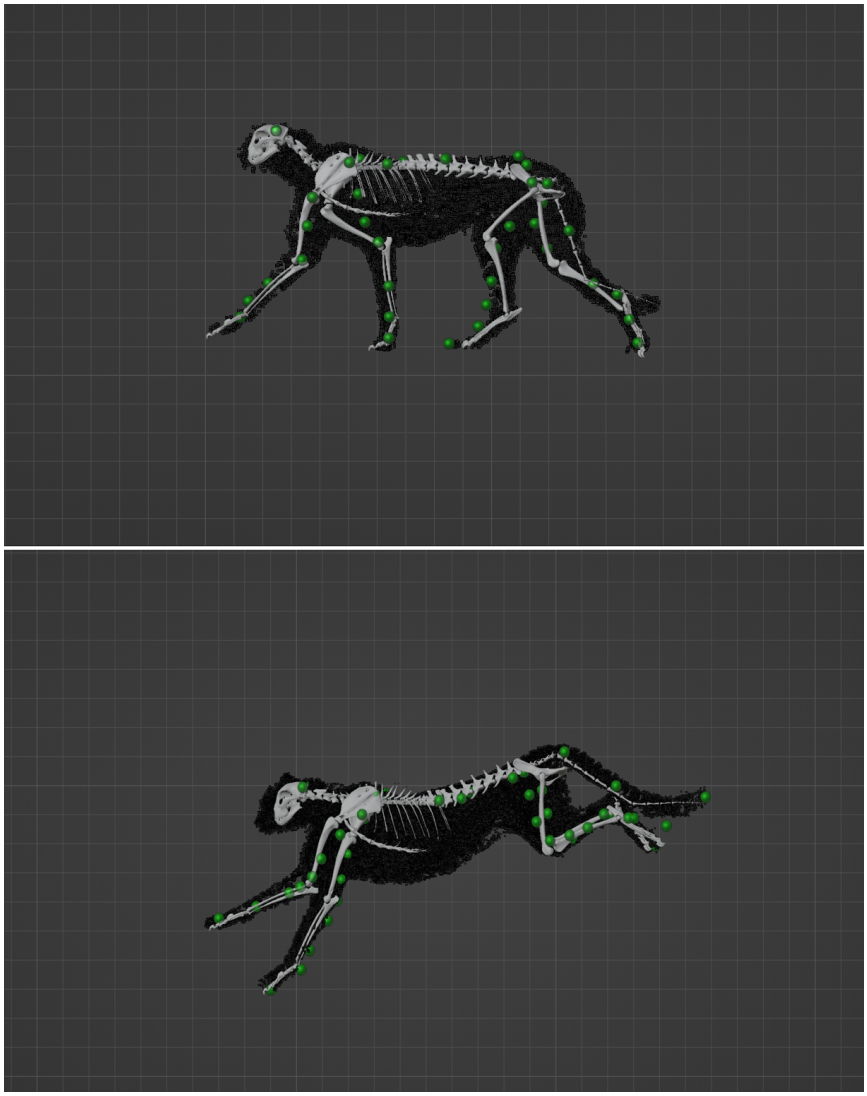

FIGURE 5.9: Resulting skeletal position along with point cloud and markers for a walking and running motion

5.4 Results and Analysis

Following the inverse kinematics process, kinematic estimation is complete. The next step involved analysing the resulting kinematics. In the absence of ground truth data, results from multiple trials were presented to provide a basis for evaluation.

A sensitivity analysis is conducted first. Since marker placement significantly impacts the final results, the marker placement process was independently repeated five times on the same run to estimate the error introduced by the placement process itself. These results are further qualitatively analysed under the assumption of periodicity. For walking trials the motion is cyclic, meaning that the left side repeats the exact motion of the right side. Assuming perfect symmetry, the left-hand side of the body kinematics matches the right-hand side exactly. This further gives an idea of the accuracy of the kinematics.

Lastly, intra-trial variance is analysed, for motions where the cheetah performed multiple gaits, the results are briefly discussed. This gives an idea of what variation is expected under the exact same measurement conditions. And thus tells us what is within reason. The results presented correspond to the same motion performed at approximately the same speed. Assuming the motions are identical, the results are analysed based on motion variance across trials, providing a preliminary estimate of the kinematic accuracy.

5.5 Sensitivity Analyses

To assess the impact of manual intervention on the accuracy of motion capture, a sensitivity analysis was conducted. This step is critical as the initial marker placement relative to joint positions can introduce variability in joint position estimates, thereby influencing the final kinematic results.

A single walking trial was selected for this analysis. This trial comprised two full gaits, ensuring a complete representation of motion and providing insight into variance over the full movement pattern, as opposed to shorter trials that might not capture the full range of motion. Thus, to isolate the variability introduced by manual marker placement, the following steps were undertaken:

1. The marker set on the 2D images was kept consistent across all repetitions to eliminate additional sources of variability.
2. The initial marker placement and scaling process were independently repeated five times for the same trial.
3. For each iteration, the inverse kinematics procedure was executed to compute joint angles.

The resulting joint angles for key joints, including the knees, elbows, shoulders, and hips, were analysed. These results are visualised in Figure 5.10, illustrating the variability introduced by the marker placement process.

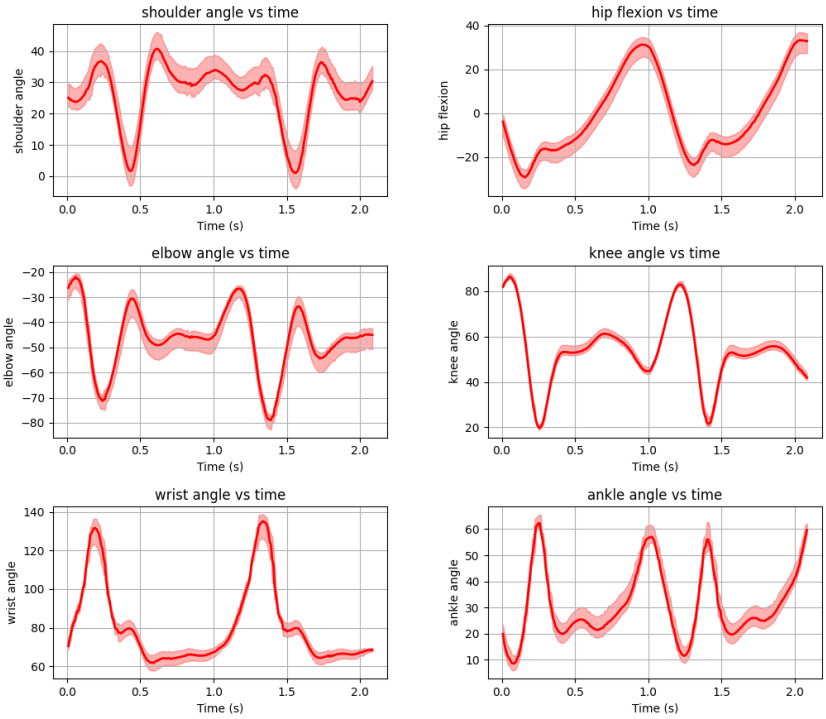

FIGURE 5.10: Sensitivity analyses results for six joint angles (mean, red line; range, shaded)

The sensitivity analysis reveals that the results are fairly uniform, as well as exhibiting periodicity, strongly indicating that the manual input required to initially align and scale the model is relatively invariant to changes. This consistency suggests that the process is robust and reliable for marker placement and scaling. The standard deviation for each joint was calculated and is summarised in Table 5.1:

TABLE 5.1: Summary of sensitivity analysis results for joint Angles

Joint	**Standard Deviation (°)**
Hip flexion	4.7592
Knee angle	2.0777
Ankle angle	3.9693
Shoulder angle	4.6593
Elbow angle	4.3375
Wrist angle	3.9509

These results highlight that, despite the inherent subjectivity of manual interventions, the variability in joint angle estimation remains within acceptable limits, further validating the robustness of our methodology.

5.6 Kinematic Results Across Trials

The behaviour of animals in in the wild experiments often presents challenges when analysing motion across multiple trials. Performing controlled movement of animals is a challenging task, which complicates the analysis further, especially when consistent motion across trials is required. With this in mind, the kinematic results are presented with only a few trials selected for the purpose of analysis.

5.6.1 Inter-trial results

Given the number of available trials and the need for consistency across them, a total of 5 recorded gaits were selected for analysis over the stance phase. It is important to note that locomotion speed has a significant impact on joint kinematics in quadrupeds, as demonstrated in several studies on kinematic estimation in dogs [29, 30]. For this reason, all trials cannot be analysed at once. The results are presented in Figure 5.11.

The figure demonstrates that the proposed method produces consistent kinematic results across multiple trials. Overall, the standard deviation of the joint angles is relatively low, indicating strong consistency. The primary source of variation occurs at the peaks of the joint trajectories, which is expected given the differences in locomotion speeds across trials.

Notably, the ankle joint exhibits the highest variance among the three joints. This is likely due to its position as the most distal joint in the kinematic chain, where variance in previous joints may propagate through. Despite this, the results are encouraging and highlight the robustness of the method in capturing the correct motion patterns across the stance phase.

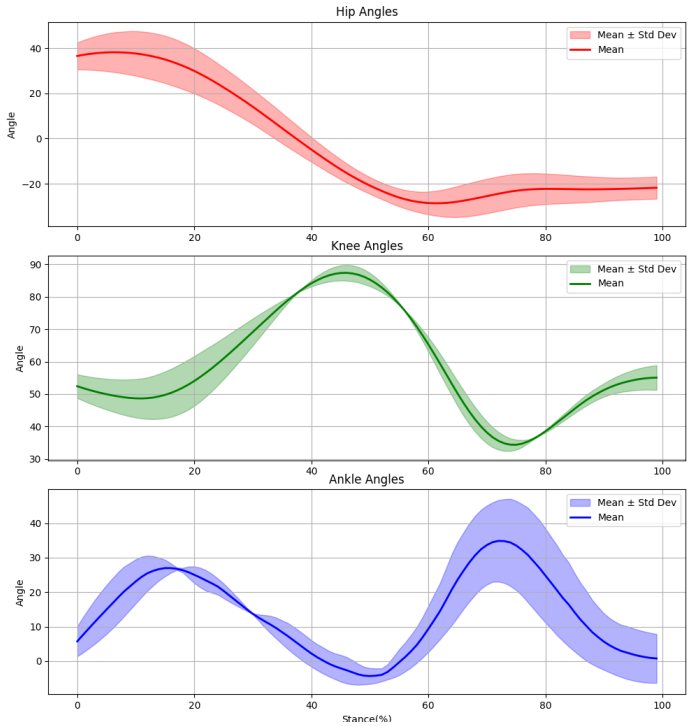

FIGURE 5.11: Trial results for hindlimb joint angles over the stance phase

5.7 Discussion

Traditional markerless motion capture is inherently prone to error. However, the method presented in this chapter demonstrates effectiveness in reducing these errors to levels that can be considered reasonable from a biomechanics standpoint. Particularly the marker projection method, which compared to keypoint triangulation presents an errorless solution for converting the 2D pixel to 3D position.

Notably, the method relies on the cheetah's spots and depends on markers being clearly visible. This contrasts sharply with the previously mentioned keypoint methods [4], which estimate joint positions directly from images using deep learning techniques. Joint estimates using these deep learning methods are highly variable and cannot alwasy be reliable used. However, the pseudo-markered approach is not a major limitation for replicating the method in animals with less textured coats, as previous studies have opted to paint markers directly onto the animals; an approach that is far less invasive than physical markers or motion capture suits [13]. That said, the method for 3D marker estimation is successfully implemented, effectively redefining the problem and making it possible to align markerless motion capture with current motion capture standards.

The sensitivity analysis highlights the method's ability to converge on joint angles that closely approximate what might be considered the true solution. Additionally, the analysis indicates that manual adjustments based on the point cloud and markers ensure that any initial errors can be corrected, leading to a conclusive estimate of the motion.

Although there is no definitive "true solution" for the motion capture performed, the proposed method provides a stable and reliable solution for joint angles and body positions. Furthermore, compared to standard triangulation techniques, the derived marker locations have sub-pixel reprojection errors. This allows the method to mitigate errors inherent in other markerless approaches, resulting in a consistent estimate.

Ultimately, the method aimed to address the shortcomings of previous approaches. The errors associated with the keypoint method used by Acinoset, as well as the issue off the motion capture not being tied to the animal's internal skeleton, have been addressed. The shortcomings of the current method, lie not in its formulation as a solution, but rather in the absence of a true benchmark for comparison, which prevents a comprehensive analysis. However, based on the presented analysis, the results are conclusive enough to consider the joint kinematics accurate for further use.

Chapter 6

Dynamic Analysis

Understanding the dynamics of wild animals remains a challenge. The limiting factor for dynamic estimation particularly in the case of analysing wild animals is firstly a lack of accurate and dynamically consistent kinematics and second, the lack of any further experimental data, particularly GRF data.

This chapter focusses on the design of a framework for analysing the dynamics of the cheetah. Rather than attempting to create a dataset of force estimates, this chapter presents a method for force and dynamic estimation in the absence of external data. This involved designing a kinetic model and using trajectory optimisation to derive consistent kinematics from the initial kinematics obtained in the previous chapter. The first section describes the development of the kinematic model, which integrates anatomical details such as inertial and CT data to provide a baseline kinetic model. This model provides a foundation for analysing the internal dynamics of the cheetah, enabling estimation of joint moments and external forces during locomotion.

A key focus of this chapter is the application of trajectory optimisation to refine the recorded gaits, addressing inaccuracies in the kinematics and ensuring that the reconstructed trajectories are physically realistic. Through optimisation, dynamic consistency is enforced, providing a more reliable predictions of GRFs and joint moments than using kinematic estimates directly.

6.1 Kinematic Model Design

The kinematic model is the foundation for subsequent dynamic analysis. It provides a detailed digital representation of a cheetah's body, encompassing its body segments, joints, and rigid body parameters. This model enables the estimation of dynamic parameters such as joint moments, joint forces, and ground reaction forces, while grounding this dynamic analysis in the animal's anatomical reality. This section details the creation of the kinematic model.

6.1.1 Rigid Body Parameters

The defining feature of the model, compared to the kinematic model described in Chapter 5, is its characterisation as a rigid body model. Rather than being purely kinematic, the model incorporates rigid bodies, each with their own masses and

inertial parameters. Furthermore, the model is able to interact physically with the ground using contact spheres added to the surface of each foot.

To calculate these rigid body parameters, the CT-scanned body of the cheetah was segmented into parts corresponding to the body components of the model, forming a template. Based on the point cloud data, these body parameters were scaled relative to one another. While it is impossible to capture these parameters with absolute precision, combining data from both the CT scan and point cloud provides a highly accurate approximation.

To determine the inertia tensors, the body segments were converted to watertight meshes. This was achieved using a Poisson surface reconstruction algorithm to close any holes in the segmented data. The inertia tensors for each body part were then calculated using the divergence theorem. This method yields the exact inertia tensor of the closed surface mesh. Similarly, the centers of mass for each segment were calculated using their mean centers as defined by the outer mesh surface. The mass of each segment was determined by scaling the segments proportionally to the cheetah's total body mass.

To position the mass segments relative to the underlying skeleton, the initial CT scan was reused to map the body segments to the skeletal structure. Using this data, the scaled model accurately aligned the bones and mass segments, allowing the rigid body parameters to be determined for each individual subject. Figure 6.1 shows the skeletal model with the COM for each body segment overlayed onto the skeleton.

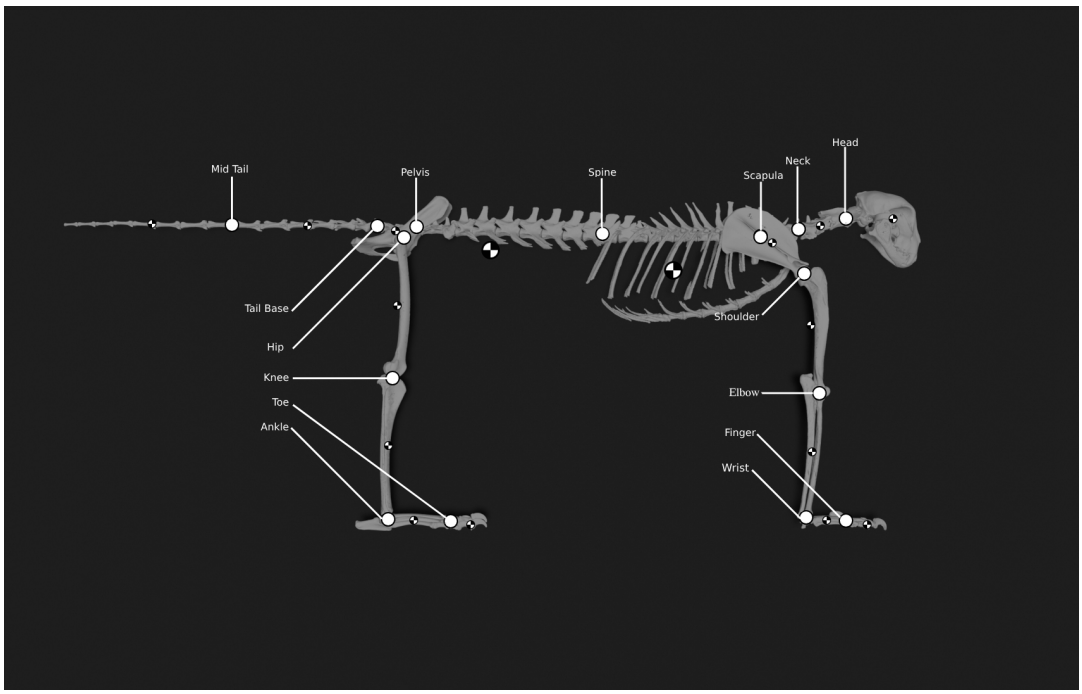

FIGURE 6.1: Kinetic model showing joint positions as well as COM positions

6.2 Dynamic Optimisation

The motion capture performed in the previous chapter provides full-body kinematics. However, these kinematics are not tied to any physical quantities, nor can the motion captured be guaranteed to be physically realisable. Additionally, due to the lack of external data, such as ground reaction forces, the motions themselves cannot be easily analysed. The goal of this section is to detail the process of enforcing dynamic consistency on the captured motions, achieving physically realisable gaits, and enabling further analysis.

A popular approach for acquiring gait cycles from a musculoskeletal model involves generating predictive solutions for locomotion based on multi-body constraints without prior knowledge of the animal's actual motion. However, since gaits have already been captured for the animal, a purely predictive solution would be counterproductive.

Instead, a trajectory optimisation method is employed. This method enables a more controlled approach for generating motions based on an initial motion which is used as an input. In addition, the initial kinematics based on the animals can be used to define an initial guess for the final solution as required by any trajectory optimisation method.

To generate this initial guess, a tracking problem is solved as the first step. This process generates the system's states and controls for the motion-tracked gaits. The resulting tracked trajectory is then used as the initial guess for a second optimisation process. This second optimisation stage refines the solution further, resolves any issues arising from strictly tracking the gaits, and minimises any residuals remaining in the system.

6.2.1 Stage 1 Optimisation

The first stage of the optimisation process is framed as a tracking problem. This involves solving for both the motion states and the underlying actuated controls. While this provides an initial estimate, it is not a complete solution. Due to the formulation of the tracking problem, errors in the kinematics are not easily filtered out, which can result in incorrect motions persisting while the solved controls can still yield a feasible solution. For this reason, the tracking problem forms the first stage of the total optimisation process.

The tracking problem is further complicated by the absence of residual actuators, specifically the pelvis. Typically, what might be used to offset any residual forces which are necessary to offset motions where joint actuators are not sufficient for generating such motions, "hand of god" forces, usually at the pelvis are used to ensure that Newtons' second law, F=ma, is satisfied throughout the motion [78, 81]. The kinetic model used for optimisation has none of these actuators to reduce residuals, instead relying completely on joint actuation for motion generation. This means that for kinematics where the body slides or shifts, there is no actuator able to compensate for such movements. Thus, any residuals which might exist in a static optimisation problem are removed. Lastly, the solution is constrained within

joint limits to ensure that the resulting motions remain realistic and do not produce unnatural gaits.

Cost Function

The exact formulation of the trajectory optimisation problem, as implemented in Moco, is detailed in Chapter 2. However, the specific cost function is not predefined, as it is entirely up to the user to determine.

For solving the tracking problem, the cost function is defined as the sum of three components

$$J = J_s + J_c + J_e, \tag{6.1}$$

where:

- J_s: State tracking cost function.
- J_c: Contact tracking cost function .
- J_e: Effort cost function.

State Tracking

The state tracking cost function is defined as

$$J_s = \int_{t_0}^{t_f} \sum_{i=1}^{N} w_i \left(s_i(t) - s_{\text{ref},i}(t)\right)^2 \, dt, \tag{6.2}$$

where:

- N: The number of tracked frames.
- $s_i(t)$: The state of the i-th variable at time t.
- $s_{\text{ref},i}(t)$: The reference state of the i-th variable at time t.
- w_i: The weighting factor for the i-th state variable.
- $[t_0, t_f]$: The time interval over which the cost is evaluated.

This cost function penalises the squared error over the trajectory, with each state being assigned an individual weight. This is a key aspect that differentiates a standard predictive problem from a tracking problem.

The states are further divided into two categories: positions and their derivatives. Since position errors are more likely, the speed tracking weight is set to $0.1w_i$ for the corresponding position state.

The weighting factors for positions are determined by their expected variance. As the kinematic chain is derived from the pelvis and spine downward, the weights decrease progressively. The spine is weighted as w, and each subsequent segment in the chain is weighted as $0.75\ w$, with the ankle receiving the smallest weight.

Control Signal/Effort

The cost function related to the control signals is given as

$$J_e = \frac{1}{d} \int_{t_0}^{t_f} \sum_{c \in C} w_c |x_c(t)|^p \, dt, \tag{6.3}$$

where:

- d: Displacement of the system.
- C: The set of control signals.
- w_c: The weight associated with control signal c (default: $w_c = 1.0$).
- $x_c(t)$: Control signal c at time t.
- $[t_0, t_f]$: The time interval over which the cost is evaluated.

This cost function, which penalises the cubed actuator excitation (p=3), is often referred to as effort or fatigue, as discussed in Ackermanm et al. [82]. Additionally, the cost is normalised by the system's displacement, d, yielding a value known as the "cost of transport." Additionally, the weights for all control signals are kept constant.

Contact Tracking

The cost function associated with contact tracking is given below.

$$J_c = \frac{1}{mg} \int_{t_0}^{t_f} \sum_{j \in G} \left\| \text{proj}_{\mathbf{n}} \left(\mathbf{F}_{m,j} - \mathbf{F}_{e,j} \right) \right\|^2 \, dt, \tag{6.4}$$

where:

- $[t_0, t_f]$: The initial and final times of the phase.
- m: The total mass of the system.
- g: Gravitational acceleration (set to 0 if not applicable; use m instead of mg in this case).
- G: The set of contact force groups.
- $\mathbf{n}$: A unit vector used for projecting the force error.
- $\text{proj}_{\mathbf{n}}(\cdot)$: The projection of the force error, either:
  - Onto $\mathbf{n}$, tracking components along $\mathbf{n}$.
  - Onto the plane perpendicular to $\mathbf{n}$, tracking components orthogonal to $\mathbf{n}$.
- $\mathbf{F}_{m,j}$: The sum of the model's contact forces for group j, expressed in the ground reference frame.
- $\mathbf{F}_{e,j}$: The experimentally measured contact force for group j, expressed in the ground reference frame.

This cost function is more complex compared to the state and control tracking functions. Given that the motion data includes quadrupedal walking gaits, the problem becomes challenging. During the walking gait, the cheetah can have up to three points of contact, leading to redundancy in the distribution of the animal's weight and GRFs.

Another factor worth considering is that the tracking problem solves for control state, thus if contact points are not explicitly specified, solutions may arise where some contact points are negated and treated as floating limbs. To address this, contact tracking is incorporated to penalise solutions that do not match the input gait.

Using the ground plane defined in Chapter 4, estimating the contact points based on kinematics becomes straightforward. In the selected trials, the paw is clearly visible, making the evaluation of the method trivial.

To estimate the ground reaction forces based on the ground plane, contact spheres were added to the models' paws and feet. These allow for contact to be specified by the model and the kinematics directly. Figure 6.2 shows the contact sphere definition for the back foot. Ground contact is defined when the contact spheres on the model make contact with the ground plane. Since the kinematics are reasonably accurate and the ground plane has been approximated and aligned, this approach is feasible. After contact detection, the GRFs used for tracking are specified as square-wave functions for the duration of contact. These forces are then input into the cost function. The forces in this case are projected onto the plane perpendicular to the vector $(0, 0, 1)$, allowing transverse forces to be ignored and tracking only the vertical component of the GRFs.

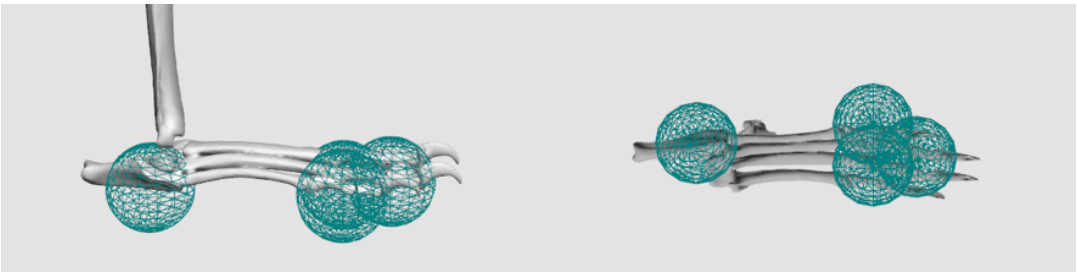

FIGURE 6.2: Contact sphere definition

This method does however require that the contact detection method be evaluated. If the contact tracking tracks incorrect foot or paw contacts, the optimisation process might produce motions which do not match the tracking but are rather forced to constrain the contact timings specified.

Table 6.1 presents the results of contact estimation using two different methods. The first method calculates contact directly from all contact spheres, where contact is defined as the moment the spheres make contact. This method, however, resulted in a significant number of false positives, particularly at the beginning of contact. The issue was most noticeable when the frontmost contact sphere detected contact, even before actual contact occurred. On the other hand, contact estimation at foot takeoff

was much more accurate, highlighting that the initial contact detection method was problematic.

To address this, the second method only considers contact when the foremost contact sphere surpasses a defined height threshold. This approach led to significantly more accurate contact estimations, as shown in Table 6.1.

TABLE 6.1: Results of Contact Estimation

Motion	Frames of Contact	Frames Detected	Success Rate
Binary contact detection	267	296	89%
Threshold contact detection	267	279	95%

Tracking Outputs

The exact outputs of the tracking process are: states, actuator activation and external forces. Additionally, The kinematics from the first stage of optimisation are refined to better match the external forces and actuator activation. The results for a half gait, showcasing some deviation in the kinematics, are shown in Figure 6.3. It is apparent that the joint angles are tracked very closely, with the maximum deviation being less than 5 degrees and the average change in angle for all joint angles remaining under 2 degrees.

As mentioned, the first stage optimisation is not meant to be the final solution, but rather to be used to provide an initial guess to the second stage optimisation. To formulate the initial guess, only the states and actuator controls are used. The contact tracking states are meant purely to ensure that the output states and controls produce a feasible initial guess.

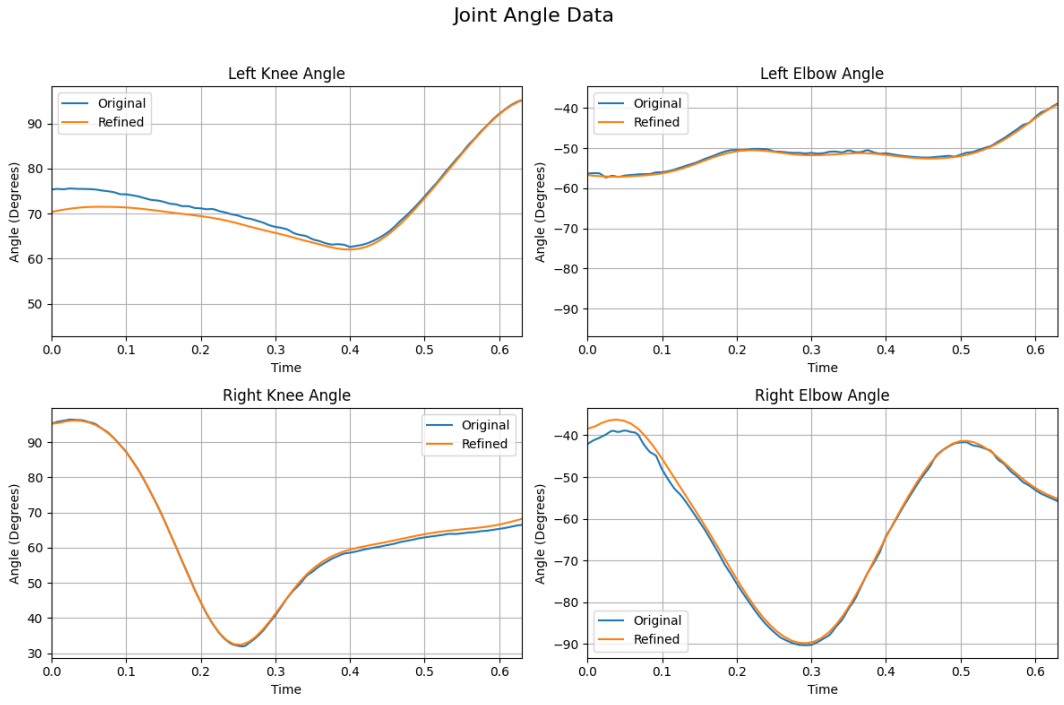

FIGURE 6.3: Refined joint angles

6.2.2 Stage 2 Optimisation

Following the initial optimisation, the second stage focuses on further refining the motions and generating the GRFs required for these motions. This stage shares similarities with the initial optimisation process; however, a key distinction is that not all states are tracked. As a result, the solution has the flexibility to adjust the input kinematics within its constraints to produce viable gaits.

To prevent the solution converging on gaits that deviate significantly from the initial data, a translation tracking term is incorporated into the cost function. This ensures that the forward motion of the solution aligns with the input gait without constraining it as much as state tracking. Additionally, the initial guess for the optimisation is derived from the tracking solution, further guiding the process toward solutions matching the initially captured motions.

Another constraint introduced in this stage is a periodicity bound. Given the challenges of guiding the cheetah to produce consistent gaits, this constraint allows the optimisation process to generate full gait cycles by inputting only half gaits. By leveraging the assumption of periodicity in the gait cycle, the final solution can extrapolate complete gait patterns, enhancing the practicality and applicability of the results.

Cost Function

The second stage optimisation process uses a slightly modified cost function. The notable difference is that there is no contact tracking. Additionally, the initial state tracking cost function is replaced with a translation tracking function. The cost function is given below.

$$J = J_t + J_c, \tag{6.5}$$

where:

- J_t represents the translation tracking cost,
- J_c denotes the effort cost

Translation Tracking

The main difference between the tracking cost function and the state cost function is that rather than tracking the joint angle and its derivative, rather the body position in 3D is tracked. This reduces the number of tracked states and allows for more variation in the refined motion due to only the position being tracked and not the velocity. The cost function is given below.

$$J_t = \int_{t_0}^{t_f} \sum_{i=1}^{N} w_i \left\| \mathbf{p}_i(t) - \mathbf{p}_{\text{ref},i}(t) \right\|^2 dt, \tag{6.6}$$

where:

- N: The number of tracked frames.

- $\mathbf{p}_i(t)$: The position vector of frame i
- $\mathbf{p}_{\text{ref},i}(t)$: The reference position vector for frame i
- w_i: tracking weight
- $[t_0, t_f]$: The time interval over which the cost is evaluated.

6.3 Optimisation Results

Once the optimisation process is completed, the result is a dynamically consistent motion along with the actuation and external forces needed to achieve the motion. Additionally, due to the periodicity bounds, the resulting gait is nearly exactly symmetrical with any deviation removed. This results in motions moving in perfectly straight lines, even if the input motion had some slight variation in off-axis translation. The resulting kinematics, joint moments and ground reaction forces are presented in this section. Given that there is currently no ground truth data, these results are presented as a solution to the problem presented by missing data needed to do further biomechanical analyses rather than being presented as ground truth.

6.3.1 Periodicity Outputs

As previously mentioned, the outputs of the final optimisation for half gaits are complete gait cycles. Notably, the input gaits do not need to be perfectly segmented to produce periodic gaits. Additionally, perfectly symmetrical gaits are not a prerequisite. The result of the optimisation process, however, yields perfectly symmetrical gait cycles as well as nearly perfect symmetry in the gait patterns. The results are illustrated in Figure 6.4.

In this trial, the input data duration is 0.63 seconds, and there is noticeable variance between the left and right joint angles. For the knee joint, the peak variation between the left and right sides is eight degrees. By comparison, the final periodic gait is 1.2 seconds long, with peak variations in joint angles reduced to below two degrees for all measurements.

This demonstrates that optimisation can effectively address the data scarcity problem associated with collecting motion data from wild animals. It is important to note that the periodicity bound is invariant to the definition of start and end points. As long as the motion is approximately cyclic, the exact start and end points do not significantly impact the results.

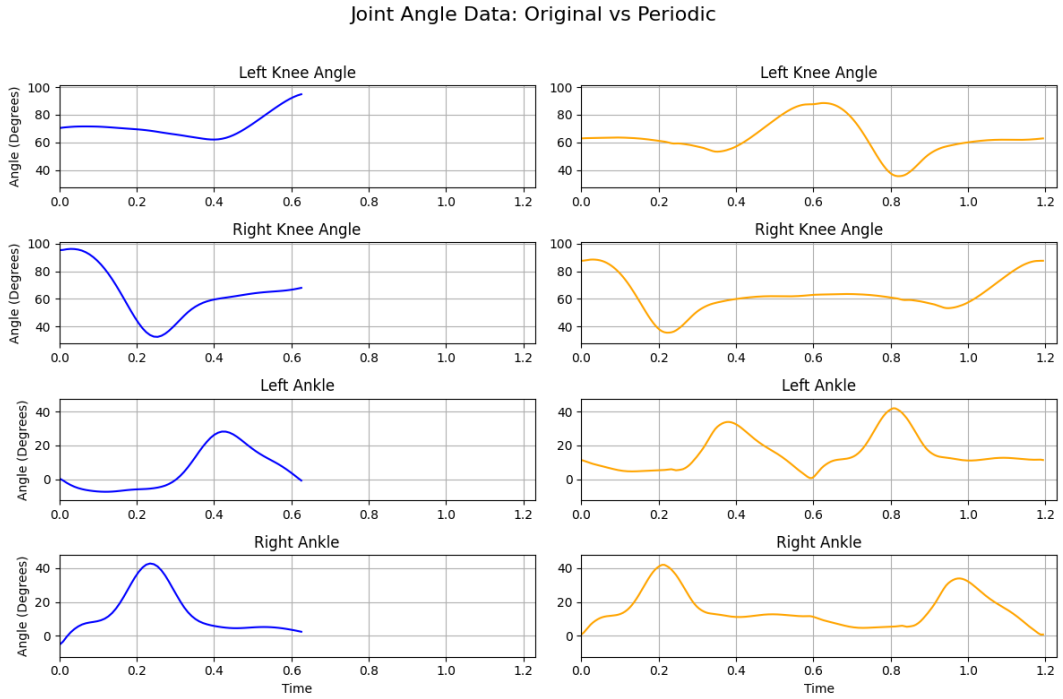

FIGURE 6.4: Periodicity output for gait cycles

6.3.2 Ground Reaction Force Estimation

One of the key motivations behind the optimisation process is the inherent challenge of collecting GRF data from animals in the wild. Therefore, the ability to estimate GRFs solely from kinematic data is of significant importance.

The GRFs are not a direct output of the optimisation process. Instead, the process yields forces acting on each contact sphere used in the contact model. The total GRFs are then obtained by summing these forces, making the estimates somewhat dependent on the placement of the contact spheres rather than assuming a single point of contact, as is done in simpler models. However, as a result of the more complex contact model, the exact output of the simulation model after processing the contact forces is firstly ground reaction forces for each foot/paw, and secondly the centres of pressure for these ground reaction forces.

OpenSim Moco currently supports only a single contact model: OpenSim's "SmoothSphereHalfSpaceForce" model. This model employs a smoothed contact force formulation using hyperbolic tangent (tanh) functions to improve performance in gradient-based optimisations. Additionally, it incorporates Hertz stiffness, a Hunt-Crossley dissipation model, and a Stribeck friction curve [83, 84].

The contact parameters were determined through experimentation and based on previous literature [41, 47]. Specifically, the plane strain modulus was set to 10 MPa, static and dynamic friction coefficients to 0.8, viscous friction to 0.5, and dissipation to 0.1 $s \cdot m^{-1}$. The smoothing behavior of the contact model which is controlled via two parameters: Hertz smoothing was set to 300, while Hunt-Crossley smoothing was set to 50.

A particular challenge in this study is the lack of measured GRF data for the cheetahs in motion. Without ground truth measurements, the evaluation of the results is limited to assessing their realism as compared to previously reported measurements in cheetahs and in other similar quadrupeds rather than verifying their exact accuracy. The GRF estimates obtained from the optimisation process are directly related to joint moments and for that reason they form an integral part of estimating the dynamics of the animal from motion capture alone.

As mentioned in Section 5.6.1, a challenge faced when doing these evaluations is consistency of movements over time. GRFs are no different. Previous studies have shown a strong correlation between GRFs and movement speed [85]. To address this, qualitative analyses of the estimated GRFs were conducted across five trials of similar speed to assess their consistency and plausibility. The results are presented in Figure 6.5.

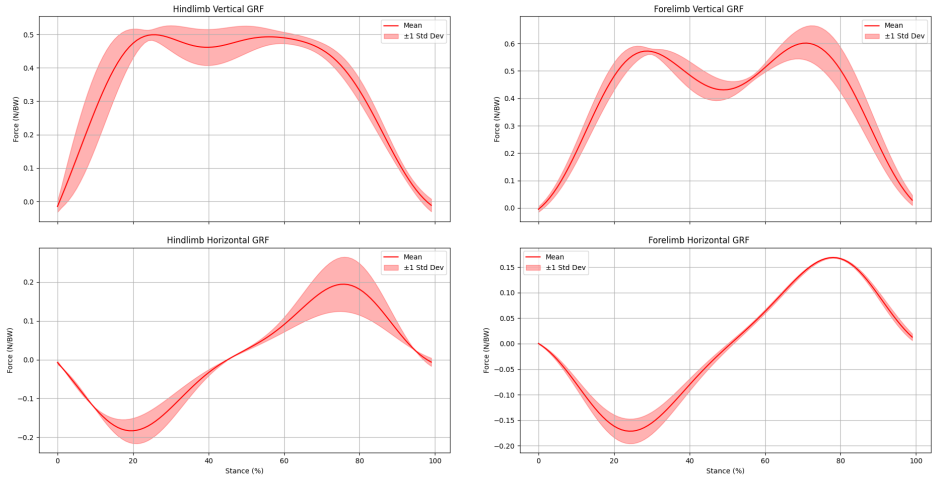

FIGURE 6.5: Ground reaction forces for hind- and forelimbs, normalised for body weight

From the resulting GRFs, it is evident that the proposed method converges to a consistent force pattern over time. These findings are promising, especially considering the limited research on GRF prediction for quadrupeds and the fact that the method proposed in this project is novel.

To strengthen the analysis in the absence of measured GRF data, the estimated vertical and horizontal GRF profiles can be compared to those reported for similar quadrupeds. While GRF data for quadrupeds is less abundant than for humans, relevant studies do exist. To facilitate comparison, six studies that explicitly plotted ground reaction force profiles for dogs and cats were reviewed [86–91]. Further analysis of these studies showed that the estimated profiles closely matched reported measurements, particularly in force relative to body weight. The horizontal component aligned almost exactly, while the vertical component exhibited some peak variation. This is expected given the various differences across datasets. Overall, these comparisons validate the results and support the realism of the GRF estimates.

Moreover, the specific gait analysed maximum of three points of contact at any given time. This poses significant challenges for analysis using standard bipedal methods, which are designed to handle only two points of contact [92]. Despite this complexity, the method demonstrates a robust capability to produce accurate GRFs even during phases with three points of contact. Lastly, the periodicity constraint ensures symmetry between the forces generated by the left and right legs. Given that straight-line motion is assumed in the selected trials, this assumption remains valid.

6.3.3 Inverse Dynamics

The final step in estimating the dynamics of the animal is calculating the joint moments. Without GRFs, these moments cannot be directly determined. While, in theory, the forces could be derived using Newton's second law, their accuracy depends entirely on the precise measurement of each body's acceleration. This dependency often results in joint moments with high variability. Moreover, the absence of a contact model further compromises the accuracy of the calculations. Therefore, an alternative approach is necessary.

In the proposed method, the optimisation process inherently solves for joint actuation and, consequently, joint moments. However, these moments are often noisy due to the nature of the optimisation. To mitigate this issue, the resultant kinematics and GRFs are first filtered to reduce noise. The joint moments are then computed using inverse dynamics, modelling the system as a series of rigid body links and applying Newton's second law to solve for the moments. The inverse dynamics problem can be directly solved for in OpenSim.

While the joint moments derived through this approach are not definitive or to be considered "ground truth," they provide a robust estimate of what the joint moments and forces might be in the absence of additional external data. The joint moments over the stance phase, normalised to body weight, are shown in Figure 6.6.

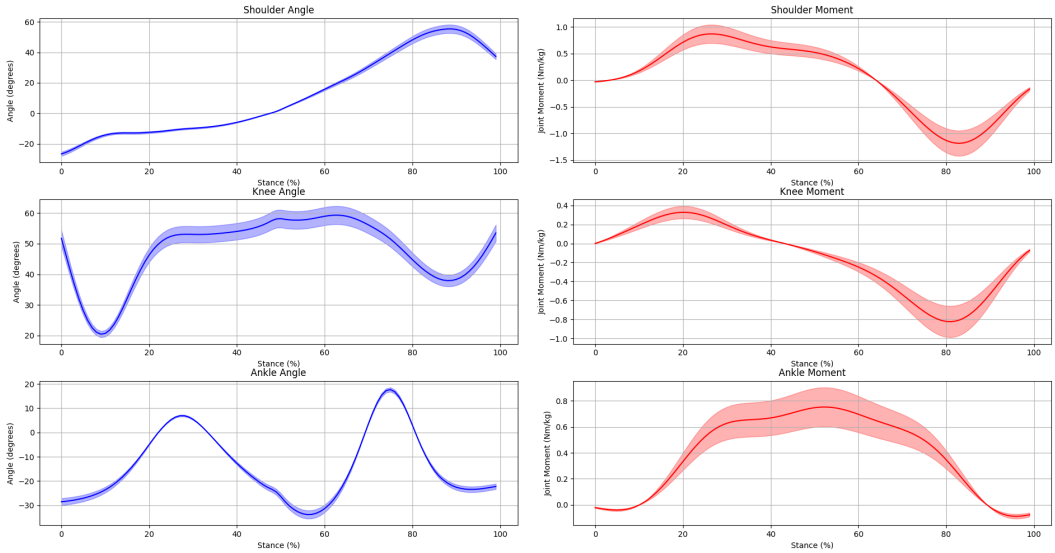

FIGURE 6.6: Joint moments and kinematics over the stance phase for the hind limb.

The results presented can only be evaluated qualitatively. Since no direct measurements of joint moments exist for cheetahs, these results must be assessed based on the methodology. Given that the kinematics and ground reaction forces are generally accepted and that the inverse dynamics results are a direct consequence of these inputs, the process therefore demonstrates that joint moments can indeed be estimated using the proposed method. For that reason, the joint moments presented in this section serve rather as a means of showcasing the methods' ability to calculate joint moments given mocap and GRF data.

6.4 Discussion

The ultimate goal of this chapter was to evaluate whether, given input kinematics, GRFs and joint moments could be reasonably calculated. This had previously not been implemented at all in cheetahs, and so to address this problem, a novel solution was formulated. By employing the two-stage optimisation process proposed in this chapter, it was showcased that it was possible to generate GRF profiles and subsequent joint moments using only motion capture data.

It was assumed that the motion capture data itself was not perfect and thus to address this, the optimisation refines the input kinematics and ensures that they are physically realisable by relying solely on torque actuators at each joint produce the movement. From these refined gaits, the GRF profiles are estimated.

The analysis of these GRF profiles, as compared to measurements recorded in previous literature, demonstrates that the method effectively estimates forces within reasonable accuracy. Furthermore, it highlights that GRF moments can be derived solely from kinematics, even in cases involving multiple foot contacts—something that is highly complex using conventional methods. This capability allows for further analysis of cheetah movement even in the absence of explicitly measured GRF data.

However, certain simplifications were made to obtain the resulting GRF profiles. The most significant assumption was that all analysed movements were straight line motions. This simplification constrained the problem by assuming periodic motion. This is justified as it enabled the analysis of incomplete gaits. This was necessary due to the cheetah's uncontrolled behavior, which made it difficult to record complete gait cycles.

Estimating GRFs directly from kinematics enables a straightforward inverse dynamics analysis of the model. This conclusively addresses the initial research question, demonstrating that GRFs and joint moments can indeed be estimated in the absence of additional data beyond joint angles. While the proposed framework is not presented as a definitive solution, it offers a first approximation for estimating the dynamics of cheetahs in the wild.

Chapter 7

Conclusion

7.1 Discussion

This project aimed to develop a methodology for 3D motion capture and dynamic analysis of wild cheetahs. The goal was to approach the problem from a biomechanics perspective, aligning the proposed method with current biomechanics standards rather than strictly from an engineering or computer vision perspective.

A review of the existing literature on similar problems revealed that current markerless motion capture techniques were insufficient for further application. To address this, a photogrammetry-based approach was proposed, leveraging the spotted pattern of cheetahs to transform the markerless motion capture problem into a pseudo-markered approach.

After conducting simulations and real-world testing, the system was used to record cheetahs in the wild. Walking and running sequences were captured, and 3D reconstructions of the cheetahs' outer surfaces were generated. The resulting point clouds were generated over the videos and evaluated both qualitatively and quantitatively. These results were generally accepted as accurate representations of cheetahs in 3D space. However, a notable limitation was the reduced density of point clouds during running trials compared to walking trials. This issue is likely attributable to the quality of the low-cost cameras used for data collection. Overall, the point clouds were able to reconstruct 90%+ of the cheetahs outer surface as compared to the silhouette in both walking and running trials. The results were deemed suitable for further analysis.

The next step was to complete the motion capture process. Strategic spots on the cheetah's body were selected to serve as markers and were projected directly onto the point clouds, minimising errors associated with triangulation. Inverse kinematics was performed using a skeletal model in OpenSim, adhering to biomechanics standards. Sensitivity analyses indicated that the resulting joint angles were accurate. While the limited number of markers introduced some error, additional markers could potentially enhance accuracy. However, due to the absence of ground truth measurements, further evaluation was not conducted. The sensitivity analyses as well as a qualitative presentation of the results indicated that the method was successful in reliably estimating the joint positions in the recorded motions.

Finally, the calculated kinematics were used to estimate ground reaction forces and joint moments. A two-stage trajectory optimisation approach successfully estimated GRFs within acceptable error ranges across various motions. Using these GRFs and joint kinematics, joint moments were estimated through a classic inverse dynamics framework. While the methods cannot be validated exactly without ground truth data, the optimisation approach and its application for estimating dynamics in quadrupeds are novel and from the limited analyses are shown to be realistic.

This study demonstrates the feasibility of 3D motion capture and dynamic estimation of cheetahs in uncontrolled environments, overcoming key challenges for non-invasive biomechanical analyses. While limitations exist, the methodology lays a foundation for future innovations in wildlife biomechanics, with potential applications ranging from robotics to conservation biology. This work provides an important first step toward estimating the dynamics of cheetahs in the wild.

7.2 Future Work

The methodology proposed in this project is by no means entirely conclusive. Certain simplifications were made to align with the research objectives, leaving room for further refinement and improvement.

One primary concern during the point cloud synthesis phase is the limitation posed by the cameras. Due to sensor quality and effective resolution, the cameras need to be positioned close to the cheetahs during recording to achieve high-quality point clouds. The use of low-cost cameras, lacking hardware synchronisation, further degrades the quality of the results. Higher-quality cameras could address these issues by enabling longer strides to be captured from greater distances and improving overall motion capture fidelity.

This study focuses on using the cheetah's spots as natural markers. To explore the broader applicability of this approach, it would be valuable to test it on animals without such markings. Alternative methods could include applying non-intrusive markers, such as painted patterns, or projecting a texture onto the animal to facilitate photogrammetry and marker tracking.

Regarding the inverse kinematics process, the current model assumes negligible skin deformation. However, skin deformation is significant, particularly during running motions. A potential improvement would be to establish a more precise mapping between the point clouds and the underlying skeleton. This could eliminate the need for markers entirely; for instance, an iterative closest point algorithm could directly use the point clouds to determine joint positions.

The skeletal and kinematic model also involves simplifications in terms of degrees of freedom. The joint ranges in the model are more constrained than those observed in reality. By increasing the number of markers, it is feasible to incorporate additional degrees of freedom into the model, thereby improving its accuracy. Furthermore, the contact model currently uses discrete contact spheres to approximate foot-paw interactions with the ground, simplifying what is, in reality, a continuous surface.

Exploring more realistic contact modelling could further enhance the estimates of GRFs and joint moments.

At present, the model is torque actuated. However, OpenSim and Moco are inherently designed to support musculoskeletal models. Using OpenSim as the foundation ensures that the complexity of the model can be progressively expanded. Adding musculature to the model would provide deeper insights into the cheetah's dynamics, allowing for a more biologically accurate representation of motion.

Finally, a significant limitation of the project is the lack of ground truth measurements for joint angles and ground reaction forces. The proposed optimisation approach is broadly applicable to other animals, making it worthwhile to validate the method against properly captured motion data and recorded GRFs. This would allow for a more thorough assessment of the method's accuracy and practical utility.

Appendix A

A.1 Joint Angles and Constraints

The joints and there accompanied joint bounds are shown in Table A.1. The pelvis is used as the root joint for the model and thus, the DOF associated with it remain unbound. The "zero" position for all joints are given in Figure 6.1, with clockwise taken as positive.

TABLE A.1: Joints andjoint bounds

Joint	Bounds(Radians)
pelvis_rotation	$[-\infty, \infty]$
pelvis_tilt	$[-\infty, \infty]$
pelvis_lift	$[-\infty, \infty]$
pelvis_tx	$[-\infty, \infty]$
pelvis_ty	$[-\infty, \infty]$
pelvis_tz	$[-\infty, \infty]$
spine	[-0.78, 0.78]
neck	[-1.05, 1.05]
head	[-0.78, 0.78]
tailbase	[-1.57, 1.57]
hip_flexion_l	[-1.25, 1.50]
hip_flexion_r	[-1.25, 1.50]
knee_l	[-0.43, 1.91,]
knee_r	[-0.43, 1.91,]
ankle_l	[-1.05, 1.57]
ankle_r	[-1.05, 1.57]
toes_l	[-1.05, 1.05]
toes_r	[-1.05, 1.05]
scapula_l	[-0.52, 0.95]
scapula_r	[-0.52, 0.95]
shoulder_l	[-1, 1]
shoulder_r	[-1, 1]
elbow_l	[-2.09,0.43]
elbow_r	[-2.09,0.43]
wrist_l	[-1.05, 1.57]
wrist_r	[-1.05, 1.57]
finger_l	[-1.05, 1.05]
finger_r	[-1.05, 1.05]

Bibliography

[1] Daniel Joska, Liam Clark, Naoya Muramatsu, Ricardo Jericevich, Fred Nicolls, Alexander Mathis, Mackenzie W. Mathis, and Amir Patel. "AcinoSet: A 3D Pose Estimation Dataset and Baseline Models for Cheetahs in the Wild". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2021, 13901–13908. DOI: 10.1109/icra48506.2021.9561338. URL: http://dx.doi.org/10.1109/ICRA48506.2021.9561338.

[2] L. Wade, L. Needham, P. McGuigan, and J. Bilzon. "Applications and limitations of current markerless motion capture methods for clinical gait biomechanics". In: *PeerJ* 10 (2022), e12995. DOI: 10.7717/peerj.12995.

[3] K. Das, T. de Paula Oliveira, and J. Newell. "Comparison of markerless and marker-based motion capture systems using 95% functional limits of agreement in a linear mixed-effects modelling framework". In: *Scientific Reports* 13 (2023), p. 22880. DOI: 10.1038/s41598-023-49360-2.

[4] Tanmay Nath, Alexander Mathis, An Chi Chen, Amir Patel, Matthias Bethge, and Mackenzie W Mathis. "Using DeepLabCut for 3D markerless pose estimation across species and behaviors". In: *Nature Protocols* (2019). URL: https://doi.org/10.1038/s41596-019-0176-0.

[5] K. Iskakov, E. Burkov, V. Lempitsky, and Y. Malkov. "Learnable Triangulation of Human Pose". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019.

[6] P.C. Bala, B.R. Eisenreich, and S.B.M Yoo. "Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio". In: *Nature Communications* 11.1 (2020), p. 4560. DOI: 10.1038/s41467-020-18441-5.

[7] Silvia Zuffi, Angjoo Kanazawa, David Jacobs, and Michael J. Black. "3D Menagerie: Modeling the 3D Shape and Pose of Animals". In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[8] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. "SMPL: A Skinned Multi-Person Linear Model". In: *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34.6 (Oct. 2015), 248:1–248:16.

[9] Silvia Zuffi, Angjoo Kanazawa, and Michael J. Black. "Lions and Tigers and Bears: Capturing Non-Rigid, 3D, Articulated Shape from Images". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2018, pp. 3955–3963.

[10] W. I. Sellers and R. H. Crompton. "A system for 2- and 3-D kinematic and kinetic analysis of locomotion, and its application to analysis of the energetic efficiency of jumping in prosimians". In: *Z. Morphol. Anthropol.* 80 (1994), pp. 99–108.

[11] Stefano Corazza, Lars Mündermann, and Tom Andriacchi. "A framework for the functional identification of joint centers using markerless motion capture, validation for the hip joint". In: *Journal of Biomechanics* 40.15 (2007), pp. 3510–3515. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2007.05.029. URL: https://doi.org/10.1016/j.jbiomech.2007.05.029.

[12] G. K. Taylor, M. Bacic, R. J. Bomphrey, A. C. Carruthers, J. Gillies, S. M. Walker, and A. L. Thomas. "New experimental approaches to the biology of flight control systems". In: *Journal of Experimental Biology* 211 (2008), pp. 258–266. DOI: 10.1242/jeb.012625.

[13] P. Windes, X. Fan, M. Bender, D. K. Tafti, and R. Müller. "A computational investigation of lift generation and power expenditure of Pratt's roundleaf bat (Hipposideros pratti) in forward flight". In: *PLoS ONE* 13.11 (2018), e0207613. DOI: 10.1371/journal.pone.0207613.

[14] M. Sandau, H. Koblauch, T. B. Moeslund, H. Aanæs, T. Alkjær, and E. B. Simonsen. "Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane". In: *Medical Engineering & Physics* 36.9 (2014), pp. 1168–1175. DOI: 10.1016/j.medengphy.2014.07.007.

[15] W. I. Sellers and E. Hirasaki. "Markerless 3D motion capture for animal locomotion studies". In: *Biology Open* 3.7 (2014), pp. 656–668. DOI: 10.1242/bio.20148086.

[16] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. 2nd ed. New York, NY, USA: Cambridge University Press, 2003. ISBN: 0521540518.

[17] K. Bartoš, K. Pukanská, and J. Sabová. "Overview of Available Open-Source Photogrammetric Software, its Use and Analysis". In: *International Journal for Innovation Education and Research* 2.4 (2014), pp. 62–70.

[18] E.-K. Stathopoulou, M. Welponer, and F. Remondino. "Open-source image-based 3d reconstruction pipelines: review, comparison and evaluation". In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2/W17 (2019), pp. 331–338. DOI: 10.5194/isprs-archives-XLII-2-W17-331-2019.

[19] S. Bianco, G. Ciocca, and D. Marelli. "Evaluating the Performance of Structure from Motion Pipelines". In: *J. Imaging* 4.8 (2018), p. 98. DOI: 10.3390/jimaging4080098.

[20] S. Liu, W. P. Bonelli, P. Pietrzyk, and A. Bucksch. "Comparison of open-source three-dimensional reconstruction pipelines for maize-root phenotyping". In: *The Plant Phenome Journal* 6.1 (2023), e20068. DOI: 10.1002/ppj2.20068.

[21] Pierre Moulon, Pascal Monasse, Romuald Perrot, and Renaud Marlet. "OpenMVG: Open multiple view geometry". In: *International Workshop on Reproducible Research in Pattern Recognition*. Springer. 2016, pp. 60–74.

[22] Ch. Wu. *VisualSFM Ver. 0.5.22*. USA. 2013. URL: http://homes.cs.washington.edu/~ccwu/vsfm/.

[23] Johannes Lutz Schönberger and Jan-Michael Frahm. "Structure-from-Motion Revisited". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[24] T. Lindeberg. "Scale Invariant Feature Transform". In: *Scholarpedia* 7.5 (2012), p. 10491. DOI: 10.4249/scholarpedia.10491.

[25] L. Righetti, A. Nylén, K. Rosander, and A. J. Ijspeert. "Kinematic and Gait Similarities between Crawling Human Infants and Other Quadruped Mammals". In: *Front. Neurol.* 6 (2015). DOI: 10.3389/fneur.2015.00017.

[26] M. S. Fischer, N. Schilling, M. Schmidt, D. Haarhaus, and H. Witte. "Basic limb kinematics of small therian mammals". In: *Journal of Experimental Biology* 205.9 (2002), pp. 1315–1338. DOI: 10.1242/jeb.205.9.1315.

[27] J. A. Vilensky. "Locomotor behavior and control in human and non-human primates: Comparisons with cats and dogs". In: *Neuroscience & Biobehavioral Reviews* 11.3 (1987), pp. 263–274. DOI: 10.1016/S0149-7634(87)80013-1.

[28] X. Zhang, C. Zhao, Z. Xu, and S. Huang. "Mechanism analysis of cheetah's high-speed locomotion based on digital reconstruction". In: *Biomimetic Intelligence and Robotics* 2.1 (2022), p. 100033. DOI: 10.1016/j.birob.2021.100033.

[29] C. Nielsen, S. M. Stover, K. S. Schulz, M. Hubbard, and D. A. Hawkins. "Two-dimensional link-segment model of the forelimb of dogs at a walk". In: *AJVR* 64.5 (2003), pp. 609–617. DOI: 10.2460/ajvr.2003.64.609.

[30] E. Andrada, L. Reinhardt, K. Lucas, and M. S. Fischer. "Three-dimensional inverse dynamics of the forelimb of Beagles at a walk and trot". In: *AJVR* 78.7 (2017), pp. 804–817. DOI: 10.2460/ajvr.78.7.804.

[31] S. B. Williams, J. R. Usherwood, K. Jespers, A. J. Channon, and A. M. Wilson. "Exploring the mechanical basis for acceleration: pelvic limb locomotor function during accelerations in racing greyhounds (Canis familiaris)". In: *J Exp Biol* 212.Pt 4 (2009), pp. 550–565. DOI: 10.1242/jeb.018093.

[32] R. G. Ellis, J. W. Rankin, and J. R. Hutchinson. "Limb Kinematics, Kinetics and Muscle Dynamics During the Sit-to-Stand Transition in Greyhounds". In: *Front. Bioeng. Biotechnol.* 6 (2018), p. 162. DOI: 10.3389/fbioe.2018.00162.

[33] J. Becker, E. Mermoz, and J.-M. Linares. "Determination of biological joint reaction forces from in-vivo experiments using a hybrid combination of biomechanical and mechanical engineering software". In: *Mechanics & Industry* 21.6 (2020), p. 623. DOI: 10.1051/meca/2020088.

[34] J. Becker, M. Emmanuel, and L. Jean-Marc. "Joint Loading Estimation Method for Horse Forelimb High Jerk Locomotion: Jumping". In: *J Bionic Eng* 16 (2019), pp. 674–685. DOI: 10.1007/s42235-019-0054-z.

[35] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen. "OpenSim: Open-source Software to Create and Analyze Dynamic Simulations of Movement". In: *IEEE Transactions on Biomedical Engineering* (2007).

[36] Thomas K. Uchida and Scott L. Delp. *Biomechanics of Movement: The Science of Sports, Robotics, and Rehabilitation*. Cambridge, MA: The MIT Press, 2021. ISBN: 9780262044202.

[37] C. R. Hargraves and S. W. Paris. "Direct trajectory optimization using nonlinear programming and collocation". In: *Journal of Guidance, Control, and Dynamics* 10.4 (1987), pp. 338–342. DOI: 10.2514/3.20223.

[38] C. L. Dembia, N. A. Bianco, A. Falisse, J. L. Hicks, and S. L. Delp. "OpenSim Moco: Musculoskeletal optimal control". In: *PLoS Comput Biol* 16.12 (2020), e1008493. DOI: 10.1371/journal.pcbi.1008493.

[39] F. C. Anderson and M. G. Pandy. "Dynamic Optimization of Human Walking". In: *Journal of Biomechanical Engineering* 123.5 (2001), pp. 381–390.

[40] P. A. van Bijlert, A. J. van Soest, A. S. Schulp, and K. T. Bates. "Muscle-controlled physics simulations of bird locomotion resolve the grounded running paradox". In: *Science Advances* 10.39 (2024), eado0936. DOI: 10.1126/sciadv.ado0936.

[41] Pasha A. van Bijlert, Thomas Geijtenbeek, Ineke H. Smit, Anne S. Schulp, and Karl T. Bates. "Muscle-Driven Predictive Physics Simulations of Quadrupedal Locomotion in the Horse". In: *Integrative and Comparative Biology* 64.3 (2024), pp. 694–714. DOI: 10.1093/icb/icae095.

[42] L. Bottini, G. D. Raimondo, B. A. Killen, Z. Sawacha, and I. Jonkers. "IMU-based ground reaction force estimation using OpenSim Moco". In: *Gait & Posture* 106 (2023), S48. DOI: 10.1016/j.gaitpost.2023.07.061.

[43] X. Zhang, C. Zhao, Z. Xu, and S. Huang. "Mechanism analysis of cheetah's high-speed locomotion based on digital reconstruction". In: *Biomimetic Intelligence and Robotics* 2.1 (Mar. 2022), p. 100033. DOI: 10.1016/j.biometrics.2022.100033. URL: https://doi.org/10.1016/j.biometrics.2022.100033.

[44] Z. da Silva, S. Shield, P.E. Hudson, A.M. Wilson, F. Nicolls, and A. Patel. "Markerless 3D kinematics and force estimation in cheetahs". In: *Scientific Reports* 14.1 (2024), p. 10579. DOI: 10.1038/s41598-024-60731-1. URL: https://doi.org/10.1038/s41598-024-60731-1.

[45] P.E. Hudson, S.A. Corr, and A.M. Wilson. "High speed galloping in the cheetah (Acinonyx jubatus) and the racing greyhound (Canis familiaris): spatio-temporal and kinetic characteristics". In: *Journal of Experimental Biology* 215.Pt 14 (2012), pp. 2425–2434. DOI: 10.1242/jeb.066720. URL: https://doi.org/10.1242/jeb.066720.

[46] A. Kewley, J. Beesel, and A. Seth. *OpenSim Creator (0.5.19)*. Zenodo. 2025. DOI: 10.5281/zenodo.14733890. URL: https://doi.org/10.5281/zenodo.14733890.

[47] P.A. van Bijlert. *MuSkeMo: Open-source software to construct, analyze, and visualize human and animal musculoskeletal models and movements in Blender*. bioRxiv (preprint). 2024. DOI: 10.1101/2024.12.10.627828. URL: https://doi.org/10.1101/2024.12.10.627828.

[48] H. Goldstein. *Classical Mechanics*. 2nd. Reading, MA: Addison-Wesley, 1980. Chap. The Euler Angles and Euler Angles in Alternate Conventions, pp. 143–148, 606–610.

[49] The MathWorks Inc. *MATLAB version: 9.13.0 (R2024b)*. Natick, Massachusetts, United States, 2024. URL: https://www.mathworks.com.

[50] Blender Community. *Blender - a 3D modelling and rendering package*. http://www.blender.org. Amsterdam, 2018.

[51] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. "MeshLab: an Open-Source Mesh Processing Tool". In: *Proceedings of the Sixth Eurographics Italian Chapter Conference*. 2008, pp. 129–136.

[52] Itseez. *Open Source Computer Vision Library*. https://github.com/itseez/opencv. 2015.

[53] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. "Open3D: A Modern Library for 3D Data Processing". In: *arXiv preprint arXiv:1801.09847* (2018). URL: https://doi.org/10.48550/arXiv.1801.09847.

[54] Alessandro Muntoni and Paolo Cignoni. *PyMeshLab*. Jan. 2021. DOI: 10.5281/zenodo.4438750.

[55] Penny E. Hudson, Sandra A. Corr, and Alan M. Wilson. "High speed galloping in the cheetah (Acinonyx jubatus) and the racing greyhound (Canis familiaris): spatio-temporal and kinetic characteristics". In: *Journal of Experimental Biology* 215.14 (July 2012), pp. 2425–2434. ISSN: 0022-0949. DOI: 10.1242/jeb.066720. eprint: https://journals.biologists.com/jeb/article-pdf/215/14/2425/1280636/2425.pdf. URL: https://doi.org/10.1242/jeb.066720.

[56] Cenek Albl, Zuzana Kukelova, Andrew Fitzgibbon, Jan Heller, Matej Smid, and Tomas Pajdla. "On the Two-View Geometry of Unsynchronized Cameras". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 2321–2329. DOI: 10.1109/CVPR.2017.250. URL: https://arxiv.org/abs/1704.06843.

[57] M. Svedman, L. Goncalves, N. Karlsson, M. Munich, and P. Pirjanian. "Structure from stereo vision using unsynchronized cameras for simultaneous localization and mapping". In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2005, pp. 3069–3074. DOI: 10.1109/IROS.2005.1545431.

[58] Andrea Bönsch and Andrew Feng. "Volumetric Video Capture using Unsynchronized, Low-cost Cameras". In: *14th International Conference on Computer Graphics Theory and Applications*. 2019, pp. 255–261. DOI: 10.5220/0007373202550261.

[59] Gwon Hwan An, Siyeong Lee, Min-Woo Seo, Kugjin Yun, Won-Sik Cheong, and Suk-Ju Kang. "Charuco Board-Based Omnidirectional Camera Calibration Method". In: *Electronics* 7.12 (2018). ISSN: 2079-9292. DOI: 10.3390/electronics7120421. URL: https://www.mdpi.com/2079-9292/7/12/421.

[60] Francois Rameau, Jinsun Park, Oleksandr Bailo, and In So Kweon. "MC-Calib: A generic and robust calibration toolbox for multi-camera systems". In: *Computer Vision and Image Understanding* (2022), p. 103353. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2021.103353. URL: https://www.sciencedirect.com/science/article/pii/S1077314221001818.

[61] Tony Lindeberg. "Scale Invariant Feature Transform". In: *Scholarpedia* 7 (2012). DOI: 10.4249/scholarpedia.10491.

[62] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. "SAM 2: Segment Anything in Images and Videos". In: *arXiv preprint arXiv:2408.00714* (2024). Available at: https://doi.org/10.48550/arXiv.2408.00714. URL: https://doi.org/10.48550/arXiv.2408.00714.

[63] Johannes Lutz Schönberger and Jan-Michael Frahm. "Structure-from-Motion Revisited". In: *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[64] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. "LoOP: local outlier probabilities". In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. CIKM '09. Association for Computing Machinery, 2009, 1649–1652. ISBN: 9781605585123. DOI: 10.1145/1645953.1646195. URL: https://doi.org/10.1145/1645953.1646195.

[65] Herbert Edelsbrunner and Ernst P. Mucke. "Three-Dimensional Alpha Shapes". In: *ACM Transactions on Graphics* 13.1 (1994), pp. 43–72. DOI: 10.1145/147130.147153.

[66] Michaela Servi, Andrea Profili, Rocco Furferi, and Yary Volpe. "Comparative Evaluation of Intel RealSense D415, D435i, D455, and Microsoft Azure Kinect DK Sensors for 3D Vision Applications". In: *IEEE Access* 12 (Aug. 2024). Received 17 July 2024, Accepted 6 August 2024, Date of Publication 9 August 2024, Date of Current Version 20 August 2024, pp. 12345–12356. DOI: 10.1109/ACCESS.2024.3441238.

[67] Lucas C. Nascimento, Waleson A. Melo, Emanuelle S. Gil, Alternei S. Brito, Felipe G. Oliveira, and José LS. Pio. "Kinect V2 vs Intel RealSense D435: A Comparative Study on 3D Mapping". In: *Proceedings of the 18th Computational Vision Workshop (WVC)*. São Bernardo do Campo/SP, 2023. Porto Alegre: Brazilian Computer Society, 2023, pp. 108–113. DOI: https://doi.org/10.5753/wvc.2023.27541.

[68] L. Mündermann, S. Corazza, and T.P. Andriacchi. "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications". In: *Journal of NeuroEngineering and Rehabilitation* 3.1 (2006), p. 6. DOI: 10.1186/1743-0003-3-6.

[69] Bruce D. Lucas and Takeo Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision". In: *Proceedings of the International Joint Conference on Artificial Intelligence*. 1981, pp. 674–679.

[70] P. Cerveri, A. Pedotti, and G. Ferrigno. "Kinematical models to reduce the effect of skin artifacts on marker-based human motion estimation". In: *Journal of Biomechanics* 38.11 (2005). Epub 2004 Nov 10, pp. 2228–2236. DOI: 10.1016/j.jbiomech.2004.09.032.

[71] A.J. van den Bogert, P.R. van Weeren, and H.C. Schamhardt. "Correction for skin displacement errors in movement analysis of the horse". In: *Journal of Biomechanics* 23.1 (1990), pp. 97–101. DOI: 10.1016/0021-9290(90)90374-c.

[72] Sam Coatham, William Sellers, and Thomas Püschel. *Convex hull estimation of mammalian body segment parameters*. 2021. DOI: 10.5061/dryad.cfxpnvx4k. URL: https://doi.org/10.5061/dryad.cfxpnvx4k.

[73] A. Boczek-Funcke, J. P. Kuhtz-Buschbeck, and M. Illerl. "Kinematic analysis of the cat shoulder girdle during treadmill locomotion: an X-ray study". In: *European Journal of Neuroscience* 8 (1996), pp. 261–272.

[74] F. Caliebe, J. Haußler, P. Hoffmann, M. Illert, J. Schirrmacher, and E. Wiedermann. "Cat distal forelimb joints and locomotion: An X-ray study". In: *European Journal of Neuroscience* 3 (1991), pp. 18–31.

[75] George E. Jr. Goslow, Robert M. Reinking, and Douglas G. Stuart. "The cat step cycle: Hind limb joint angles and muscle lengths during unrestrained locomotion". In: *Journal of Comparative Physiology* 85 (1973), pp. 177–192.

[76] M. Guillot, P. Gravel, M. L. Gauthier, H. Leblond, M. Tremblay, S. Rossignol, J. Martel-Pelletier, J. P. Pelletier, J. A. de Guise, and E. Troncy. "Coxofemoral joint kinematics using video fluoroscopic images of treadmill-walking cats: development of a technique to assess osteoarthritis-associated disability". In: *Journal of Feline Medicine and Surgery* 17.2 (2015), pp. 134–143. DOI: 10.1177/1098612X14537261.

[77] J. P. Kuhtz-Buschbeck, A. Boczek-Funcke, M. Illert, and C. Weinhardt. "X-ray study of the cat hindlimb during treadmill locomotion". In: *European Journal of Neuroscience* 6 (1994), pp. 1187–1198.

[78] OpenSim Documentation. *OpenSim Documentation - Global Site*. Accessed: 2025-01-22. 2017. URL: https://simtkconfluence.stanford.edu:8443/display/OpenSim/OpenSim+Documentation.

[79] J. J. Dunne, T. K. Uchida, T. F. Besier, S. L. Delp, and A. Seth. "A marker registration method to improve joint angles computed by constrained inverse kinematics". In: *PLoS One* 16.5 (2021), e0252425. DOI: 10.1371/journal.pone.0252425.

[80] T. K. Uchida and A. Seth. "Conclusion or Illusion: Quantifying Uncertainty in Inverse Analyses From Marker-Based Motion Capture due to Errors in Marker Registration and Model Scaling". In: *Front. Bioeng. Biotechnol.* 10 (2022), p. 874725. DOI: 10.3389/fbioe.2022.874725.

[81] J. L. Hicks, T. K. Uchida, A. Seth, A. Rajagopal, and S. L. Delp. "Is My Model Good Enough? Best Practices for Verification and Validation of Musculoskeletal Models and Simulations of Movement". In: *Journal of Biomechanical Engineering* 137.2 (2015), p. 020905. DOI: 10.1115/1.4029304.

[82] M. Ackermann and A. J. Van Den Bogert. "Optimality principles for model-based prediction of human gait". In: *Journal of Biomechanics* 43.6 (2010), pp. 1055–1060. DOI: 10.1016/j.jbiomech.2009.12.012.

[83] Michael A. Sherman, Ajay Seth, and Scott L. Delp. "Simbody: multibody dynamics for biomedical research". In: *Procedia IUTAM* 2 (2011). 2011 Symposium on Human Body Dynamics, pp. 241–261. DOI: 10.1016/j.piutam.2011.04.023. URL: https://doi.org/10.1016/j.piutam.2011.04.023.

[84] Gil Serrancolí, Antoine Falisse, Christopher Dembia, and Jonas Vantilt. "Subject-Exoskeleton Contact Model Calibration Leads to Accurate Interaction Force Predictions". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* PP.99 (2019), pp. 1–1. DOI: 10.1109/TNSRE.2019.2924536.

[85] T. S. Keller, A. M. Weisberger, J. L. Ray, S. S. Hasan, R. G. Shiavi, and D. M. Spengler. "Relationship between vertical ground reaction force and speed during walking, slow jogging, and running". In: *Clinical Biomechanics (Bristol, Avon)* 11.5 (1996), pp. 253–259. DOI: 10.1016/0268-0033(95)00068-2.

[86] Ronald J Corbee, Huub Maas, Arie Doornenbal, and Herman A. W. Hazewinkel. "Forelimb and hindlimb ground reaction forces of walking cats: assessment and comparison with walking dogs." In: *Veterinary journal* 202.1 (2014), pp. 116–127. URL: https://api.semanticscholar.org/CorpusID:30708.

[87] Maxim Moreau, Bertrand Lussier, Laurent Ballaz, and Eric Troncy. "Kinetic measurements of gait for osteoarthritis research in dogs and cats". In: *The Canadian veterinary journal. La revue veterinaire canadienne* 55 (Nov. 2014), pp. 1057–1065.

[88] T. M. Griffin. "Biomechanics of quadrupedal walking: how do four-legged animals achieve inverted pendulum-like movements?" In: *Journal of Experimental Biology* 207.20 (2004), pp. 3545–3558. DOI: 10.1242/jeb.01177.

[89] E. Schnabl-Feichter, A. Tichy, and B. Bockstahler. "Coefficients of variation of ground reaction force measurement in cats". In: *PLOS ONE* 12.3 (2017), e0171946. DOI: 10.1371/journal.pone.01719.

[90] G. Helms, B.A. Behrens, M. Stolorz, and et al. "Multi-body simulation of a canine hind limb: model development, experimental validation and calculation of ground reaction forces". In: *BioMed Eng OnLine* 8 (2009), p. 36. DOI: 10.1186/1475-925X-8-36. URL: https://doi.org/10.1186/1475-925X-8-36.

[91] E. Andrada, G. Hildebrandt, H. Witte, and M. S. Fischer. "Positioning of pivot points in quadrupedal locomotion: limbs global dynamics in four different dog breeds". In: *Front Bioeng Biotechnol* 11 (2023), p. 1193177. DOI: 10.3389/fbioe.2023.1193177.

[92] Erik J. Dijkstra and Elena M. Gutierrez-Farewik. "Computation of ground reaction force using Zero Moment Point". In: *Journal of Biomechanics* 48.14 (2015), pp. 3776–3781. DOI: https://doi.org/10.1016/j.jbiomech.2015.08.027.