UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Machine learning-based motion capture: Exploring the application, limits and opportunities in live performance

by Jamal Knight

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Andrew Johnston and Adam Berry

University of Technology Sydney
Faculty of Engineering and IT

May 2024

Certificate of original authorship

I, Jamal Knight, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

Production Note:
Signature removed prior to publication.

Jamal Knight

May, 2024

This research is supported by an Australian Government Research Training Program Scholarship.

Table of contents

Acknowledgements

I would like to express my sincere gratitude to the individuals and organisations who played pivotal roles in the progress of this thesis. Their unwavering support, expertise, and generosity were instrumental in making this research possible.

I am immensely grateful to Cloé Fournier, whose invaluable contribution went beyond measure. Her dedication and participation in dance experiments provided invaluable insights and contributions to the project's development.

My heartfelt thanks go to my dedicated supervisors, Andrew Johnston and Adam Berry, whose guidance and mentorship throughout the years have been indispensable. Without their expertise and unwavering support, this research would not have been attainable.

I extend my appreciation to Carlos Barreto, whose assistance in navigating coding challenges and expertise in machine learning models were of immense help throughout the project.

I want to recognise the tireless efforts of Thomas Ricciardiello in managing all the data and rehearsals before the performance. His dedication to ensuring projectors worked seamlessly added significant value to this research.

I am thankful to David Clarkson of Box of Birds and Matt Hughes for generously permitting me to observe and test my ideas during their rehearsals.

I would like to express my gratitude to Sam McGilp and Harrison Hall for their trust in incorporating my work into their performance. Their willingness to embrace new ideas and approaches enriches the creative process.

I am thankful to Seng Ung and Andrew Lai for their support at the Animal Logic Academy, which provided a fertile ground for the development of this research.

To Tri, I extend my heartfelt appreciation for your understanding, support and inspiration throughout this journey.

The progress and development of this thesis have been greatly enriched by the unwavering support, encouragement, and patience of these individuals and organisations. I am genuinely fortunate to have had such an incredible network of people who played a significant role in its advancement. Thank you all for being a part of it.

Abstract

This practice-based research investigates the application of machine learning techniques to generate motion capture data for creating abstract animations accompanying live performers in the area of performing arts. Through a literature review, model experimentation, qualitative interviews, collaborative practice, and production of a creative work, the study explores whether machine learning methods can enhance existing motion capture practices and unlock new creative possibilities for performing arts practitioners.

- The literature review examines the evolution of motion capture from early beginnings to current methods, identifying gaps and opportunities for machine learning integration.
- Interviews with experienced performance practitioners reveal key needs like affordability and portability, where modern machine learning models may offer advantages.
- Experimental testing of monocular and multi-camera models on performers evaluates capabilities and limitations. Monocular models exhibit satisfactory detection performance in uncontrolled settings, yet encounter challenges in accurately detecting keypoints with atypical movements, diverse body shapes, and various body positions. Multi-camera models demonstrate improved temporal smoothness compared to monocular models, although requiring a calibration procedure and the need for more cameras and equipment for setup.
- A collaboration with experienced dance choreographers who were interested in incorporating machine learning-based motion capture into their performance was conducted. The collaboration highlighted the utility of machine learning-based models as efficient collaborative tools with expedited results. The imprecise outputs resulting from irregular detections, however, present avenues for artistic exploration.
- A performance piece produced by the author then demonstrates generating animations from performer motion capture using a multi-camera method. The intent is to have the animation play alongside a staged live performance in front of an audience. This practice-based analysis also reveals practicalities of delivering machine learning-based animations to support dance performances. These insights encompass the influence of machine learning technologies on the collaborative workflow, the performer's experience, and the audience's engagement within the context of dance performances. The feedback from the

audience share insights to the blending of digital animation and physical live performance, while the dancer's perspective compares the machine learning workflow favourably to past experiences.

Overall, the research highlights machine learning-based motion capture's democratising potential through providing cost-effective and accessible motion capture to expand artistic capabilities. It provides evidence that machine learning-based pose detection can be integrated into creative performance workflows and used to produce compelling artistic artwork. While refinements remain, this investigation reveals meaningful practical and artistic benefits, setting the stage for machine learning-based motion capture to transform creative expression at the intersection of technology and the performing arts.

List of figures

List of video links

This research incorporates motion analysis, necessitating the inclusion of video examples supplemented by illustrative images. To facilitate a comprehensive understanding of the described movements, a specific video link, denoted as vidStream, is provided for streaming accessibility. Additionally, all videos pertaining to this research are made available for download via the provided link vidDownloadAll. This multimedia approach serves to enhance the clarity of the research findings and enables a more nuanced examination of the intricate aspects of motion under investigation.

List of tables

1 Introduction

1.1 Topic

This practice-based research delves into performing arts, focusing on the use of motion capture data obtained through machine learning methods. Machine learning has emerged as a powerful tool in various domains, revolutionising industries with its ability to learn from data and make informed predictions and decisions. While there is a wealth of research in the area of machine learning for motion capture in particular, little work explores its potential in the context of creating abstract animation for performing arts, which will be the focus of this thesis.

The work will explore the perspectives of performing arts practitioners, the experience of audience members, the practical realities of model deployment and use, and the collaborative process involved in creating a new performance with machine learning-based pose detection at its foundation. The research aims to explore how machine learning-based motion detection unlocks new avenues for creative expression and the complexities and constraints of pursuing those avenues in the performing arts domain.

The research also seeks to understand the current methods utilised by choreographers and performers in obtaining and using motion capture data. By gaining insights into the traditional practices and challenges faced by professionals in the performing arts industry, the study aims to assess how machine learning methods for motion capture can complement or enhance existing techniques. To validate the practicality and feasibility of the identified machine learning models, the research endeavours to test these models with professionals involved in a performing arts production. By collaborating with practitioners in real-world settings, the study aims to gauge the effectiveness and applicability in authentic performance contexts.

As a tangible outcome of this research, a live performance involving a performer was developed to showcase the use of the identified machine learning model. This performance serves as a demonstrative showcase of the technology’s potential and its integration into performing arts practices. It also offers an opportunity to investigate the broader practical impacts of machine learning-based methods for motion capture on choreography, collaboration, rehearsal processes, and technical configurations.

1.2 Definition

Performing arts:

McCarthy et al. (2001) describe the broader definition of performing arts as theatre, music, opera, and dance across a spectrum from traditional high arts to popular contemporary forms. Performing arts encompass live performances in all venues, as well as recordings and mass media distribution through radio, video, television, the internet, and other channels. The performing arts system includes both the production and consumption sides - those involved in creating as well as experiencing the art. Overall, the performing arts cover a diverse range of stage and media arts spanning from classical to contemporary, which engage audiences through live events and audio/visual mediums.

The performing arts, as defined within the context of this thesis, represents a distinctive form of artistic expression characterised by the use of movement and actions performed by dancers or performers. This art is primarily executed in a live setting, often combining elements of spontaneity or scripted choreography, and is typically witnessed by a live audience. The performing arts explored in this thesis employs motion capture technology, an element that translates key elements of the performer's movements into data. This data drives digital animations that are projected within the performance space with the performer, introducing an added visual element and interaction to the overall artistic experience.

Digital Performance:

According to Dixon (2007), Digital Performance can be defined as:

"The conjunction of computer technologies with the live performance arts, as well as gallery installations and computer platform-based net.art, CD-ROMs, and digital games where performance constitutes a central aspect of either its content (for example, through a focus on a moving, speaking or otherwise 'performing' human figure) or form (for example, interactive installations that prompt visitors to 'perform' actions rather than simply watch a screen and 'point and click')."

Digital Performance encompasses a wide range of artistic practices that integrate digital technologies with live performance. It represents a field where traditional performance arts intersect with new media, creating novel forms of expression and audience engagement.

This fusion of the digital and the performative has led to experimentation in theatre, dance, performance art, and installations.

The concept of Digital Performance extends beyond merely using technology as a tool in performance. As Dixon notes, it potentially represents "a new paradigm in theatre and performance," where digital elements become integral to the creative process and the performance itself. These digital components can act as intelligent, sensitive, and subjective entities within the performance space, essentially becoming characters on the stage alongside human performers.

Artificial intelligence:

Artificial intelligence (AI) can be defined as the capability of computer systems to simulate human cognitive functions, particularly learning and problem-solving. AI uses mathematical and logical processes to replicate human-like reasoning in order to learn from new information and make decisions. It encompasses broader capabilities including problem-solving, reasoning, adaptation, and generalised learning. (Piloto, 2022)

Machine learning:

Machine learning (ML) is the process where computers are trained to identify and extract patterns from collected data and apply these patterns to perform new, previously unencountered tasks. It involves systems that can continuously learn and improve from experience by analysing data patterns.

Machine learning functions as a subset of artificial intelligence - it's one of the key methods that enables AI systems to become "intelligent." While AI represents the broader goal of creating computer systems that can mimic human intelligence, ML provides the specific mechanisms for achieving this goal through pattern recognition and data analysis. ML feeds into AI by studying data patterns that data scientists can use to improve AI systems' capabilities. Together, they form a symbiotic relationship where ML provides the learning capabilities that make AI systems more intelligent and effective at producing solutions (Piloto, 2022).

Throughout this dissertation, our focus is specifically on the application of pre-trained machine learning models designed for motion capture, particularly open-access solutions. Rather than developing new machine learning approaches or working with lower-level

computer vision libraries, this research examines how existing, readily available machine learning-based motion capture systems can be applied in performance contexts.

This specific application of machine learning is significant because it represents a practical approach that aligns with how many creative practitioners would likely engage with this technology. Instead of developing custom machine learning solutions, which would require substantial technical expertise and resources, this research explores the creative possibilities available through existing, accessible implementations.

When we discuss machine learning in this context, we are specifically referring to these pre-trained models that can generate 3D mesh animations from video input. While there are many other applications of machine learning in computer vision and motion capture, our focus remains on these implementations that offer a balance between sophisticated capabilities and practical accessibility for creative applications.

This clarification helps position our research as an investigation of how existing machine learning tools can be effectively applied in performance contexts, rather than a technical exploration of machine learning algorithms themselves.

1.3 Research questions

This research examines the intersection of machine learning-based motion capture systems and performing arts, with particular attention to their practical application and value of freely available off-the-shelf systems available to artists today. The study addresses two primary research questions (RQ):

’In performing arts, what are the characteristics creative practitioners look for in motion capture systems?’ (RQ1).

‘What are the benefits, limits, and implications of current machine learning-based motion capture systems in the performing arts space?’ (RQ2).

These questions aim to explore the connection between the characteristics and performance of current machine learning-based motion capture systems and the needs of those involved in performing arts. Through investigating these questions, this research seeks to understand

both the current state and future potential of machine learning-based motion capture in artistic contexts.

This research investigates the practical implementation and effectiveness of off-the-shelf machine learning-based motion capture systems within performing arts contexts. The study examines how these technologies interface with artistic practice, focusing on their functional value and benefits for creative practitioners. Through systematic analysis, we seek to understand both the current applications and potential impact of these systems on artistic creation and performance processes.

1.4 Significance of research

This research aims to explore how these technologies may be used in practice, exploring needs and application through discourse with performing arts practitioners, analysis of experimental applications of the technology by practitioners, and the design, development and realisation of a new performance piece by the author. The lessons here provide insight into the technical realities of deploying the technology, explore practitioner perspectives on the technology, and provide a window into how current versions of the technology may deliver value for, and potentially shape, the performing arts.

This research acknowledges the rich historical context of collaboration between choreography and computing, which spans over 80 years as reflected in Chapter 2. Pioneers such as Merce Cunningham (Vondrak et al., 2012), William Forsythe (Synchronous Objects Archive Site, n.d.), and Wayne McGregor (Bastien Girschig, 2019) have made significant contributions to this field. Even earlier, the work of Loie Fuller (Sommer, 1980a) in the early 20th century laid groundwork for the integration of technology in performance. Additionally, animators like Norman McLaren (McLaren, 2017), Oscar Fischinger (Egan, 2020), and Mary Ellen Bute (Moen, 2019) have explored the connection between abstract animations and embodied movement-based elements, further enriching this interdisciplinary area.

Other notable projects worth mentioning include Living Archive (Bastien Girschig, 2019a), a collaboration between choreographer Wayne McGregor and Google Arts & Culture. This project utilises machine learning to generate new movement material based on McGregor's entire catalog of dance movements. The model can produce hundreds of iterations of dance phrases, significantly accelerating the creative process and providing choreographers with a vast array of options. This tool effectively replicates McGregor's distinctive dance style, allowing for rapid exploration and iteration of movement variations.

Movingstories (Schiphorst & Pasquier, 2015) is an interdisciplinary research group that focuses on integrating movement knowledge into technology design. The project combines body-based disciplines such as Laban Movement Analysis with digital interaction technologies and social contexts. A unique feature of Movingstories is the collaboration between certified movement analysts and AI researchers, bridging the gap between technical expertise and bodily practices. This approach enables a more nuanced understanding of movement in the context of technology design.

WhoLoDancE (Rizzo et al., 2018) is a three-year EU-funded project aimed at applying breakthrough technologies to dance. The project has multiple objectives, including investigating bodily knowledge, preserving cultural heritage, innovating dance teaching methods, enriching choreography, and widening access to dance practice. WhoLoDancE utilizes a range of technologies, including motion capture, volumetric displays, and consumer-grade devices, to create interactive dance learning experiences and tools for choreographers. This comprehensive approach addresses various aspects of dance practice, from pedagogy to creation and preservation.

Casa Paganini – InfoMus(“Casa Paganini InfoMus,” 2019) is an international research center that focuses on scientific and technological research in multimedia systems and multimodal human-computer interfaces. Their work involves developing computational models of non-verbal expressive and social behaviour, with applications in improving quality of life, creating innovative interfaces, and preserving cultural heritage.

“Dance interaction with physical model visuals based on movement qualities” (Alaoui et al., 2013) is a project that developed a novel interface for controlling interactive visuals through full-body dance movements. This system focuses on movement qualities, recognising predefined qualities through gesture analysis and controlling abstract visuals based on physical models. The visuals, in turn, reflect the dancers' expressions, creating a dynamic interplay between movement and visual representation.

Double Skin/Double Mind (DS/DM) (Greco & Scholten, 1997) is a dance method developed by Emio Greco and Pieter C. Scholten. It focuses on basic principles of breathing, jumping, expanding, and reducing, aiming to develop both physical and mental awareness in dancers. The method has been adapted for various educational and research projects, including digital applications and interdisciplinary collaborations. DS/DM demonstrates how a specific dance practice can be transformed into a versatile tool for both artistic and technological exploration.

These projects collectively illustrate the diverse approaches to integrating technology with dance, ranging from machine learning and motion capture to interactive visualizations and pedagogical tools. They represent significant efforts in bridging the gap between state-of-the-art technology and performing arts practice, each contributing unique insights and tools to the field.

Building upon this historical foundation, this research makes distinct contributions by focusing on the application of contemporary, freely available, off-the-shelf machine learning-based motion capture systems in performing arts. Specifically, this work explores the potential of accessible, markerless motion capture using consumer-grade equipment, addressing the need for more affordable and portable solutions as identified in practitioner interviews. Specifically, this work explores the value and potential of markerless motion capture from the perspective of performing arts practitioners via interview, practical deployment, experimentation and performance development.

This research advances the field in several ways:

1. Expanding awareness of current machine learning-based motion capture technologies and their potential value in performing arts spaces: By demonstrating the effectiveness of machine learning-based approaches using consumer-grade equipment, this work opens up new possibilities for a wider range of artists and performers to incorporate motion capture into their practice.

2. Bridging the gap between cutting-edge technology and artistic practice: The thesis provides practical insights into how contemporary machine learning-based systems can be integrated into choreographic processes, addressing real-world challenges faced by practitioners.

3. Exploring new aesthetic possibilities: The research investigates how the unique characteristics of contemporary machine learning-based motion capture, including its imperfections and glitches, may be harnessed for creative expression.

4. Providing a framework for future exploration: By systematically evaluating different machine learning approaches and their applicability in performance contexts, this work lays the groundwork for future research and development in this rapidly evolving field. By identifying limitations of current systems in performing arts contexts suggests potential development pathways for machine learning researchers to refine and specialise machine

learning-based systems specifically for performing arts. Equally, highlighting the use of glitch, may suggest interesting new performance pathways that explore the limits (and beyond) of these systems.

1.5 Structure of thesis

Chapter 2 ‘Literature review’ provides a foundational backdrop to the research by exploring the history of motion capture technology. Traditional motion capture methods are discussed, followed by an investigation into the integration of motion capture in performing arts. The chapter ends by examining machine learning methods in the context of performing arts.

In Chapter 3 ‘Research methodology’, the thesis discusses the framework of practice-based research and reflective practice as a fundamental element of the study. The chapter elaborates on thematic analysis methods employed for the analysis of practitioner and audience interviews, providing insights into the data collection and analysis processes. Additionally, it addresses the ethical considerations integral to the research, ensuring a robust and responsible approach to data gathering and analysis.

Chapter 4 ‘Interviews with practitioners who use motion capture’ explores the need for and application of motion capture in performing arts contexts. These discussions establish a practical and contextual link between the requirements of practitioners and the potential applications of machine learning methods in motion capture for performing arts.

Chapter 5 ‘Monocular pose detection for performance’ focusses on a single-camera machine learning-based motion capture system. This chapter delves into the deployment and testing of the model, shedding light on its capabilities and limitations. It also outlines collaborative efforts with choreographers and directors who use the model’s outputs for their performance, showcasing how it can be practically applied within the performing arts space.
Chapter 6 ‘Multi-camera pose detection for performance’ addresses multi-camera systems, emphasising the camera calibration process necessary for motion capture. Two multi-camera models, EasyMocap and MPP2SOS, are introduced, and their selection rationale is outlined. The chapter provides insights into the practical use of these models and reviews their performance in a practical deployment setting and implications for wider use in the performing arts space.

Chapter 7 ‘Artefact production’ explores the planning, preparation, and execution of a new performance centred on machine learning-based pose detection named *Interlinked*. The performance was a collaboration between the author and a performer over the span of six months. This chapter delves into the production's creative and technical aspects and offers insights from post-performance interviews with the audience. Additionally, it presents an interview with the performer, shedding light on the process and creative dynamics involved in bringing the performance to life. The chapter provides a view of the research’s practical application within a live performance context.

Chapter 8 ‘Future work’ explores the avenues for further development in monocular and multi-camera machine learning models for motion capture. It investigates the feasibility of extending these methods to detect multiple performers, enhancing their temporal smoothness, and achieving real-time capabilities within the model frameworks.

Chapter 9 ‘Conclusion’ encapsulates the research project’s contributions, implications, and findings.

2 Literature Review

2.1.1 Introduction

Modern motion capture can be described as “the process of measuring motion in the physical world and then encoding them into the digital medium” (Okun & Zwerman, 2021). However, motion capture practices have origins far prior to the digital age, but still follow the same principle: Transferring motion data from one medium to another.

Motion capture has a history well before the digital age, where motion was attempted to be transferred to drawings. These early attempts of motion capture are the precursor to modern-day motion capture. Digital motion capture comes in three main forms, marker-based, sensor-based and markerless or sensorless systems. These systems are used extensively in the entertainment industries, but are also used in a number of other industries including performing arts. Figure 2.1 presents a range of digital motion capture systems associated with performing arts, spanning from pre-digital motion capture to contemporary digital methods.

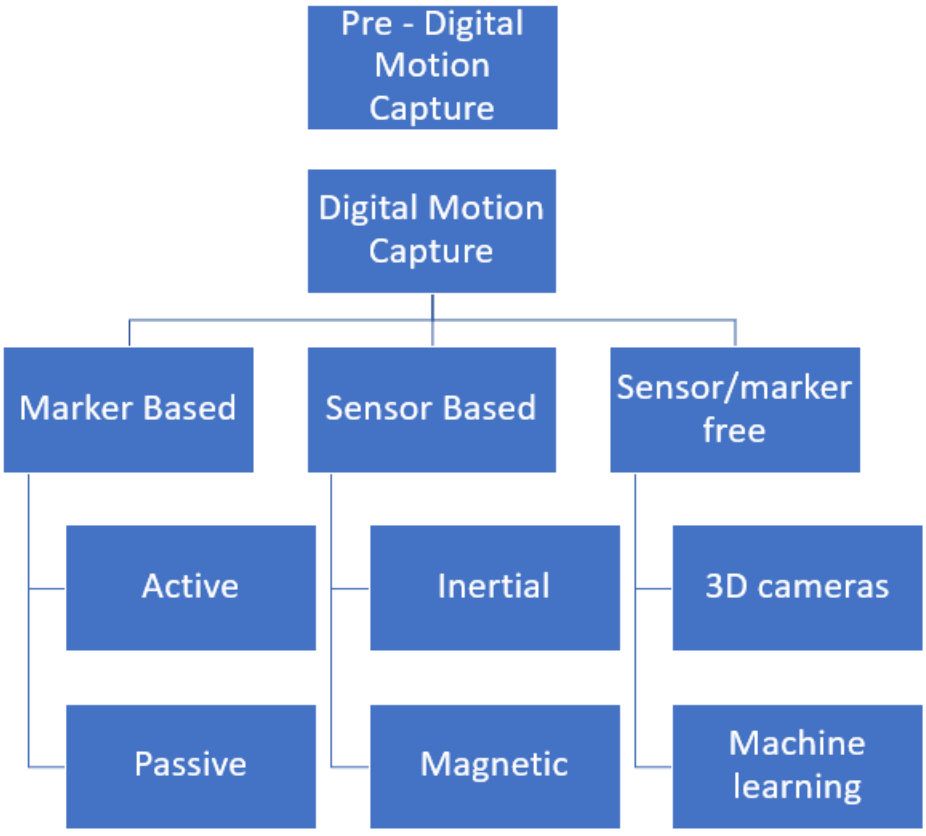

Fig. 2.1: Types of motion capture systems.

The table below is a comparative analysis of motion capture systems from pre-digital to contemporary methods as discussed in this section.

System	Time Period	Technology	Output	Key Advancements	Video Link
Zoopraxiscope	1880s	Multiple cameras in sequence	Play-back on circular disc, painted glass animation	First device to show sequential motion capture, foundation for animation	Link
Chronophotographic Fixed-Plate Camera	1890s	Timed shutter camera	Multiple exposure on single plate	First single camera motion tracking system	Link
Rotoscoping	1915	Tracing from live action film	Cell animation	Realistic human movement in animation, integration of live action with animation	Link
Medical Gait Analysis	1970s	Film cameras	Spatial joint positions	First 3D motion data from multiple cameras	Not available
Early Digital Reference (Brilliance commercial)	1985	Hand-marked joints	Computer animation	Bridge between manual and digital methods, reference-based animation	Link
Passive Marker Based	1990s - Present	Multiple high-speed cameras, reflective markers	Real-time skeletal animation	High accuracy, real-time feedback, multiple actor capture	Link

Active Marker Based	2000s - Present	LED-emitting markers	Skeletal animation	Less sensitive to light interference, easier identification of specific markers	Link
iMocap (ILM)	2006 - Present	Regular film cameras, visible markers and check patterns	CG character animation	Integration with film production, reduced specialised equipment, on set capability	Link

Table 2.1. Motion capture systems from pre-digital to current.

2.1.2 Early days of motion capture

Zoopraxiscope

One of the earliest examples of where motion is recorded was the work of Eadweard Muybridge and his creation of the zoopraxiscope (Muybridge, 1882). He used multiple cameras which photographed the motion of horses in quick succession. It revealed a sequence of frames and gave the illusion of movement. Figure 2.2 below displays a Zoopraxiscope.

Fig. 2.2: Zoopraxiscope (Ulaby, 2010)

The photographs were treated and re-photographed on a circular disc, before an artist would transfer the images by painting them on glass. The painted disc would then be played back on the zoopraxiscope revealing the animation. The zoopraxiscope is considered one of the earliest devices of motion picture, and also motion capture. He published *Animals in Motion* (Muybridge, 1899) and *The Human Figures in Motion* (Muybridge, 1901b) which are still used by animators for reference today.

Chronophotographic fixed-plate camera

Inspired by the work of Muybridge, Etienne-Jules Marey invented a chronophotographic fixed-plate camera (Sipe, 2020). This camera had a timed shutter which exposed images on a plate. This design improved on Muybridge's work by using only one camera to capture the exposures instead of using many. Marey's work included the motion of animals and that of humans. To capture the motion of humans, he created a black motion capture suit that had reflective strips on the limbs. When the captured images are compiled, they look very similar to skeletal motion capture data. The motion capture suit he invented for his work is arguably

the very first motion capture suit created. Figure 2.3 illustrates multiple frames captured from Marey's piece titled "Walking Man", where reflective markers were added to the body of a person walking.

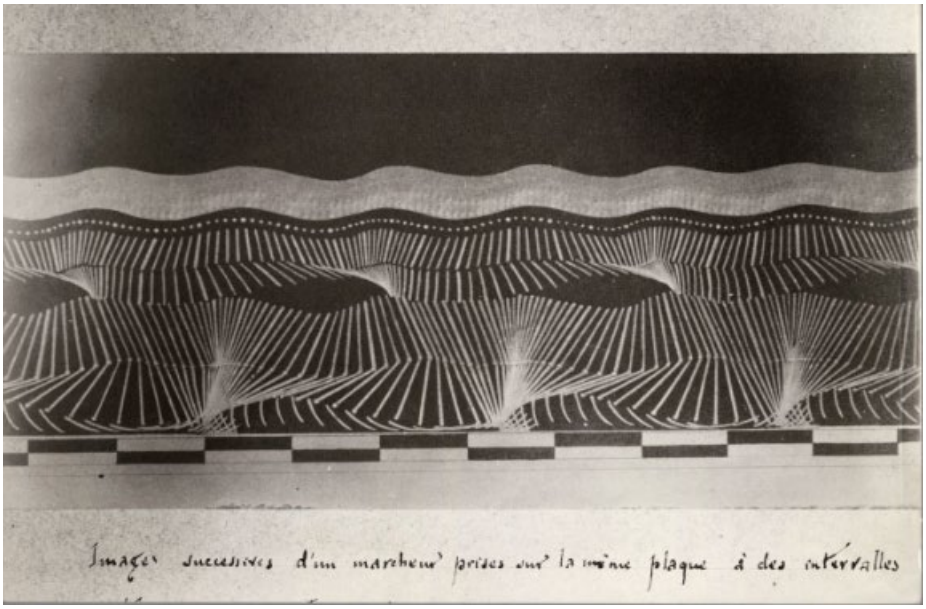

Fig. 2.3: "Walking Man" (Sipe, 2020)

The development of the Zoopraxiscope and subsequent advancement of the technology to produce the Chronophotograph allowed the portability of the motion capture device by containing the camera into one unit. The specialised motion capture suit that Marey developed clearly illustrated the skeletal movement of a human as still images, once overlayed atop of each other, showed the animation over time. Technology has advanced since then, but these early developments in motion capture are still used as motion capture techniques today.

Rotoscoping

Rotoscoping is the process of producing animation by tracing the cells of a live action film frame-by-frame. This method was devised by Max Fleisher, who in 1915 created a film made from rotoscoping with his brother. Fleisher patented rotoscoping in 1917 and produced the animated series *Out of the Inkwell* [1] which combined animation with live action. Fleischer's' brother would be filmed in a clown costume and Fleischer would rotoscope his actions. The complex motion of the animation would appear realistic and smooth due to the rotoscoping

[1] Out of the Inkwell, https://www.imdb.com/title/tt0007151/

process. Fleischer would go on to animate other characters from comics such as Superman, Popeye and Betty Boop.

While Fleischer and his brother were working on their films based on rotoscoping, Disney spent four years on the production of *Snow White and the Seven Dwarfs* in 1937. In some scenes, particularly those which featured Snow White, rotoscoping was used to animate the complex motion of her movement. Disney animators would use live action reference of an actress acting the part, and then drew over their motion to create the animation (Kalmakurki, 2018).

The rotoscoping process is time-consuming as it requires the artist to paint over each image to make an animation. The film ran at twenty-four frames per second, making close to one hundred and twenty thousand individual frames. The concept of the chronophotograph and rotoscoping are similar in that they are capturing the live action motion and producing an animation from that motion.

2.1.3 The beginning of digital motion capture

The film, television and animation industries are most well-known for digital motion capture starting from the 1980's. However, it was in the 1970's when surgeon David Sutherland and engineer John Hagy obtained three dimensional (3D) data from film cameras for gait analysis (Baker, 2007). Three cameras were used to film the subjects which triangulated limb position and animation from joint position. The frames were digitised and five joints of the subject were calculated. The test took about twenty minutes and the calculations from the film could calculate the spatial position of five joints in about two hours. This allowed gait analysis of the patient in 3D space.

In the 1980's, computer technology was in its' infancy and the few computers that were commercially available to produce animation were very expensive. Only a few production companies had access to powerful enough computers to produce animation. In 1985 the commercial *Brilliance*[2] was aired during the Super Bowl for the National Canned Food Information Council by Robert Abel and Associates (Goldman, 1985). The commercial consisted of a computer generated (CG) environment and featured an animated feminine robot. The animation of the robot was perceived to exhibit human-like qualities. The robot was animated using 3D software, but prior to the animation, a live action actor was filmed

[2] Brilliance, https://computeranimationhistory-cgi.jimdofree.com/brilliance-1984/

making the same movements. The actor had markers drawn on her joints and were used as reference for the animation. This technique allowed the animator to capture the motion of the actor and translate it over to the animation. Figure 2.4 below illustrates the actor with markers drawn on her joints, posing as reference for animators (left) and the rendered CG animation used in the commercial (right).

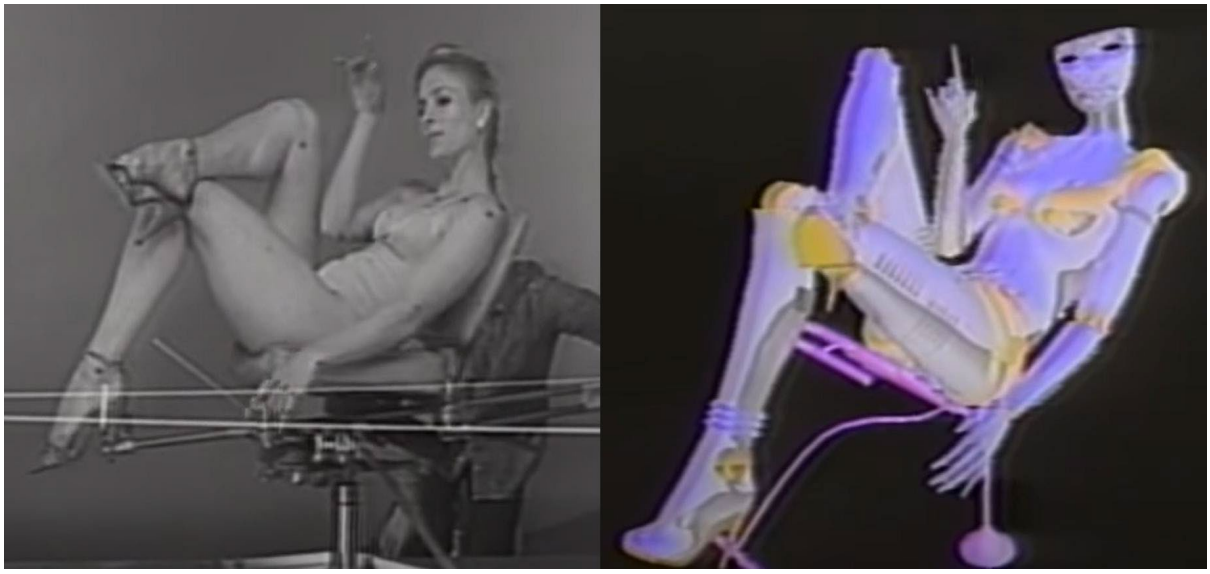

Fig. 2.4: *Brilliance* (Goldman, 1985)

These early methods of mapping joint positions from the perspective of a camera and translated into 3D space paved the way for modern motion capture technology. Human motion is complex and nuanced, but once joint and limb positions are tracked, some degree of that motion can be captured. The animation of the robot in *Brilliance* by today's standards is somewhat crude and subtle intricate motion was not needed to be captured as robot movement has some rigidity.

2.1.4 Passive marker-based

Optical motion capture is the process where the motion of an actor is captured by multiple high-speed cameras positioned on the outskirts of the capture stage. The cameras are equipped with light-emitting diodes (LED's) and would capture the LED reflection of the markers (Okun & Zwerman, 2021). The actors would typically wear tight fitting suits and reflective (passive) markers would be strategically placed over the actors' body. The markers would typically range from 2mm to 14mm in size and calculate accurate movement and rotation of moving joints (Joslin, n.d.). The size of the markers would vary according to the cameras resolution and the subject being captured. The cameras need to have a direct line

of sight to the reflective markers to record accurate data. The capture stage would need to be large enough for the actor(s) to perform. In a typical motion capture session, a monitor would output the motion of the actor in 3D space, either as a skeleton representation or an avatar in real-time (Dent, 2014). This monitoring allows the director or choreographer to visualise the motion capture data. Once the session is over, some data clean-up may be required before handing the skeleton to the person requesting the motion capture services, or the client. The client is then able to apply the motion captured data to a 3D character of their choice. Some editing or retargeting may be required if the dimensions of the 3D character are not similar to the actor. Optical motion capture is widely used in films, television and animation today (Menache, 2011).

The reason why optical motion capture, particularly Vicon systems, are preferred in the entertainment industries is because of its high level of accuracy. An optical motion capture system can run at speeds up to 2000 samples a second and output an animated skeleton at a variety of frame rates (Kitagawa & Windsor, 2008). An animation producer might require multiple actors to be captured at the same time (e.g. for a fight scene). With enough cameras, an optical system can sufficiently capture the motion of groups of people at the same time. Markers may fall off with contact with other actors or if an actor would roll on the ground (“Marker Placement Guide,” 2007). Occlusion of markers may occur if there are more than one actor in the capture stage. These errors are rectifiable after capture in the clean-up process, although these error rates will improve by adding more cameras or markers to the setup. The actors are able to move freely in the capture space because they are not connected to any wires or cables. The precision of an optical Vicon motion capture system is high, typically in the range of a few millimetres per marker (Merriaux et al., 2017). This allows the client to have an accurate digital representation of their actor once clean-up has been performed on the animation due. Clients may see the resulting cleaned up animation the same day if a few shots were processed.
If a VFX (visual effects), animation or games production company requires realistic human-like motion capture, they would typically rent a capture stage with specialised technicians (Okun & Zwerman, 2021). Motion capture stages are usually rented by the hour or the day depending on the amount of shots that need to be captured (Menache, 2011). The client would send a shot list to the motion capture studio and arrange for an actor to be there on the day. The cameras would be calibrated before the shoot started and re-calibrated two or three times during the day. Motion capture stages are temperature controlled and even the slightest degree change can cause the housing of the camera to expand and cause the motion capture data to be inaccurate. After the shoot, the client can choose the appropriate take from the capture session and trim off any unnecessary motion that is not required. The

motion is processed and a Filmbox[3] (.fbx) file with the animated skeleton is delivered to the client. This is not always the case, as Electronic Arts (EA), one of the largest sports games' producers, have their own in-house motion capture stage. EA is well known for creating the FIFA (Federation Internationale de Football Association) series of soccer games and have produced many FIFA games since 1993 (Dent, 2014).

A Vicon equipped motion capture stage is costly to set up. The capture stage at EA consists of 132 Vicon cameras to capture performance. With just two cameras including software licenses costing US$12,500, setting up a studio similar to this one would be too costly for smaller companies, so they are forced to rent (Dent, 2014). EA has a high turnover of sports games incorporating human motion so it makes sense that they have their own motion capture stage. Figure 2.5 depicts an image of the motion capture stage situated in Vancouver, Canada, operated by EA. Table 2.2 outlines the benefits and limitations of the Vicon optical motion capture system.

Fig. 2.5: EA Capture Lab. Vancouver, Canada. Dent, S (2014).

[3]Filmbox, https://www.autodesk.com/products/fbx/overview

Vicon Benefits	Vicon limitations
High accuracy The Vicon system is known for its precise tracking.	**Costly equipment** The initial setup and maintenance costs can be relatively high, making it less accessible for smaller budgets.
Real-time feedback of animation It can provide real-time feedback to monitor the output of the animation as the actor performs.	**Not portable** The stationary nature of the system makes it less portable compared to other systems.
Scalable The system can scale to accommodate various setups, allowing flexibility in capturing different types of motion.	**Needs a dedicated space** Setting up a Vicon system needs a dedicated, controlled environment, which may not be feasible for users with limited space. Suits and markers are needed: Depending on the marker setup, subjects may find wearing suits with affixed markers limiting or uncomfortable, which could impact their performance.
Versatile marker sets It supports various marker configurations enabling customisation for specific applications.	**Suits and markers are needed** Specialised suits fitted with markers are required to be worn by the performers
Broad application Vicon systems are used in a wide range of fields due to their versatility.	**Line-of-sight dependency** Vicon relies on a direct line of sight between markers and cameras, which can lead to occlusion issues when markers are temporarily hidden from view.
	Needs calibration The system requires precise calibration which can be time consuming and complex, particularly in large capture volumes.
	Only for indoor use Vicon systems are fixed and usually in a dedicated space or studio that is temperature controlled.

	Complex post-processing Once the motion capture data is recorded, a computer will process it in order to output usable data to apply to a character.

Table 2.2: Vicon benefits and limitations.

2.1.5 Active marker-based

The other form of optical motion capture uses active markers. Active markers are also positioned on the actors' body, but emit light-emitting diode (LED) light. Each marker has some electronics and is powered by a small battery. Active markers are less sensitive to light interference than passive markers. It is for this reason that active markers are more commonly used on set or a sound stage, and possibly outdoors. Active markers can be controlled by modulating the light emission, or even turning them off to measure background light. This modulation can be used to identify a character. Active markers are brighter than passive markers because they do not need to reflect light. Active markers however work best when they are directly facing the camera, noise may be introduced if they are indirectly facing the camera. Figure 2.6 depicts active markers secured to a motion capture suit.

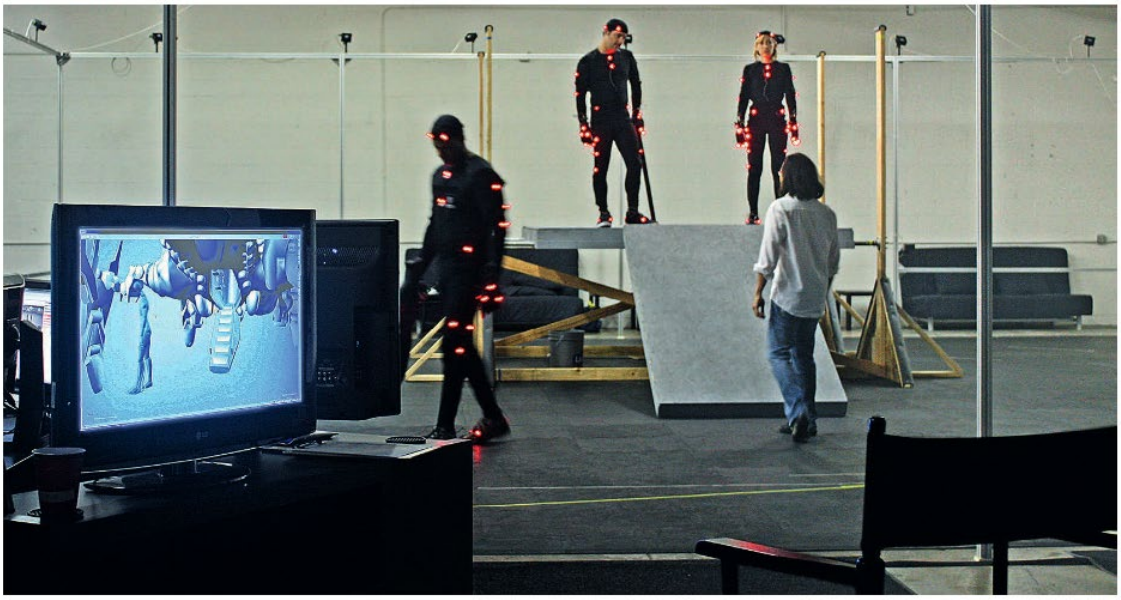

Fig. 2.6: Active markers (Okun & Zwerman, 2021)

2.1.6 iMocap

Industrial Light and Magic (ILM), a world class VFX company with offices around the globe, developed patented technology called iMocap for the film *Pirates of the Caribbean: Dead Man's Chest* in 2006. The characters in the film were completely CG, but because the story took place on a pirate ship the usual procedure of motion capture was not feasible because the characters were interacting within their environment and props. The actors' movement could be observed because they were wearing grey suits with visible markers, as well as black and white checked bands around their arms, legs, waist and head which would provide stable tracking information from each character (Failes, 2019). Figure 2.7 illustrates the iMocap suits worn by actors, featuring visible markers on the left, along with checked bands designed to enhance the precision of tracking computer-generated characters on the right.

Fig. 2.7: Pirates iMocap. (Failes, 2019).

The tracking markers and checked bands on the actors provided valuable data for movement, but the view from one camera was not adequate because of the occlusions on the set of both props as well as the actors themselves. Two or more cameras were used as 'witness cameras' to further refine the animation where the primary camera was occluded, or to just provide extra information on the pose of the actor. Hence no specialised high speed cameras are needed, only RGB cameras, to minimize the amount of equipment brought out onto the set (Failes, 2019). The iMocap system has since evolved and used in films such as *Iron Man*[4]*, Teenage Mutant Ninja Turtles: Out of the Shadows*[5]*, The Avengers*[6]*,* and *Rogue One: A Star Wars Story*[7].

[4] Iron Man, https://www.imdb.com/title/tt0371746/
[5] Teenage Mutant Ninja Turtles: Out of the Shadows, https://www.imdb.com/title/tt3949660/
[6] The Avengers, https://www.imdb.com/title/tt0848228/
[7] Rogue One: A Star Wars Story, https://www.imdb.com/title/tt3748528/

There can be occasions where motion capture may not be suited to a project. Motion capture transfers humanoid movement to a character and if the character in the project does not have a humanoid shape, the transfer will not work. There have been instances where motion capture actors have used stilts or prosthetics to match the dimensions of the character they are portraying. Alan Tudyk in *Star Wars: Rogue One* wore stilts in the motion capture stage to perform in ways that matched his droid character 'K-2SO' (Phillips, 2016). These enhancements can help in bringing the targeted characters to life, but if the characters' anatomy is too far from a human's, it will likely fail and need to be hand animated. Motion capture imparts a distinct and stylistic quality to animations, which may not necessarily align with the desired aesthetic or characteristics of the intended animated character. In games for example, the motion of character may be stylised and move in a particular way. Applying smooth humanoid animation to such a character may not fit into the narrative. Another reason why a studio might not use the services of motion capture is because of expense. Motion capture stages are expensive to rent and out of reach for smaller independent studios.

2.1.7 Conclusion

Motion capture, has a rich history dating back to the early attempt of transferring motion to drawings. From the zoopraxiscope to the chronophotographic fixed-plate camera, early pioneers like Eadward Muybridge and Etienne-Jules Marey laid the groundwork for modern motion capture techniques. Rotoscoping, introduced my Max Fleisher, further refined the process by tracing live-action film frame-by-frame to create realistic animations. The dawn of digital motion capture came in the 1970's when researchers like David Sutherland and John Hagy used film cameras for gait analysis, marking the beginning of 3D motion capture. The 1980's witnessed the use of expensive computer technology to produce animations, and the commercial *Brilliance* demonstrated the integration of live-action reference with computer generated animation.

Optical motion capture, both passive and active marker-based systems emerged as a preferred method in the entertainment industry due to its high level of accuracy. Optical motion capture involves capturing an actor's motion using high-speed cameras equipped with reflective markers placed on a suit worn by the performer. The resulting data is then used to animate digital characters with a real-time preview of the animation during the capture process. Vicon systems, with their high accuracy and frame rates, have become particularly popular in film, television, and animation productions.

Motion capture has come a long way since its early beginnings and has evolved into a powerful tool. Despite the accuracy and realism that optical motion capture offers, it does come with some drawbacks. One significant limitation is its cost, as setting up an optical motion capture stage can be prohibitively expensive for smaller production companies, indie projects or individuals. Additionally, optical motion capture stages are not usually or easily portable, which means that if a team wants to utilise this technology, they would need to travel to the motion capture studio. This lack of portability may pose logistical challenges and restrict access for some projects. As a result, while optical motion capture remains a powerful tool in various industries, researchers and developers continue to explore alternative solutions to make motion capture more accessible and cost-effective.

2.2 Motion capture in performing arts

2.2.1 Introduction

Motion capture has evolved greatly since its first use by Fleischer in 1915 (Kalmakurki, 2018) who rotoscoped each frame by hand. A technology originally built for the entertainment industries, motion capture has since been adopted by dancers and choreographers to capture the human motion in digital form. There are many examples where motion capture data can be used for dance purposes, such as projecting animation for performances, as a means of archiving dance movement, analysing dance movement with other styles or training purposes to name a few.

In examining the evolution of technological tools for dance and choreography, Alaoui et al. (2014) provided a comprehensive analysis of early digital systems designed to support choreographic processes. They identified four main categories of tools: reflective systems for visualizing movement and structures, generative tools for creating new movement material, interactive systems allowing real-time engagement with digital media, and annotative tools for documenting choreographic processes. Their analysis revealed that these early technological approaches tended to be highly idiosyncratic, with each system focusing on specific aspects of the choreographic process rather than providing comprehensive solutions. This fragmented approach to technological tools in choreography highlights why subsequent developments in modern motion capture techniques, which offer more integrated and flexible solutions, represent such a significant advance in the field. While these early systems laid important groundwork in digital dance tools, they were limited by the

technology of their time and often required specialised knowledge or equipment to operate effectively.

In performance, motion capture is often used to capture the performers' motion which is projected nearby (Chang, 2019; Felciano, 1999; Johnston, 2015a; Jung et al., 2012). The projections would be a live representation of their movement, previously rendered offline, or a combination of both. The projections consist of computer animations based on the performers' motion ranging from a literal representation of the performer to the abstract. The animation may not consist of the human form at all, but drive the motion of other objects. In Figure 2.8, we observe an instance demonstrating the integration of projected animations into the realm of performing arts.

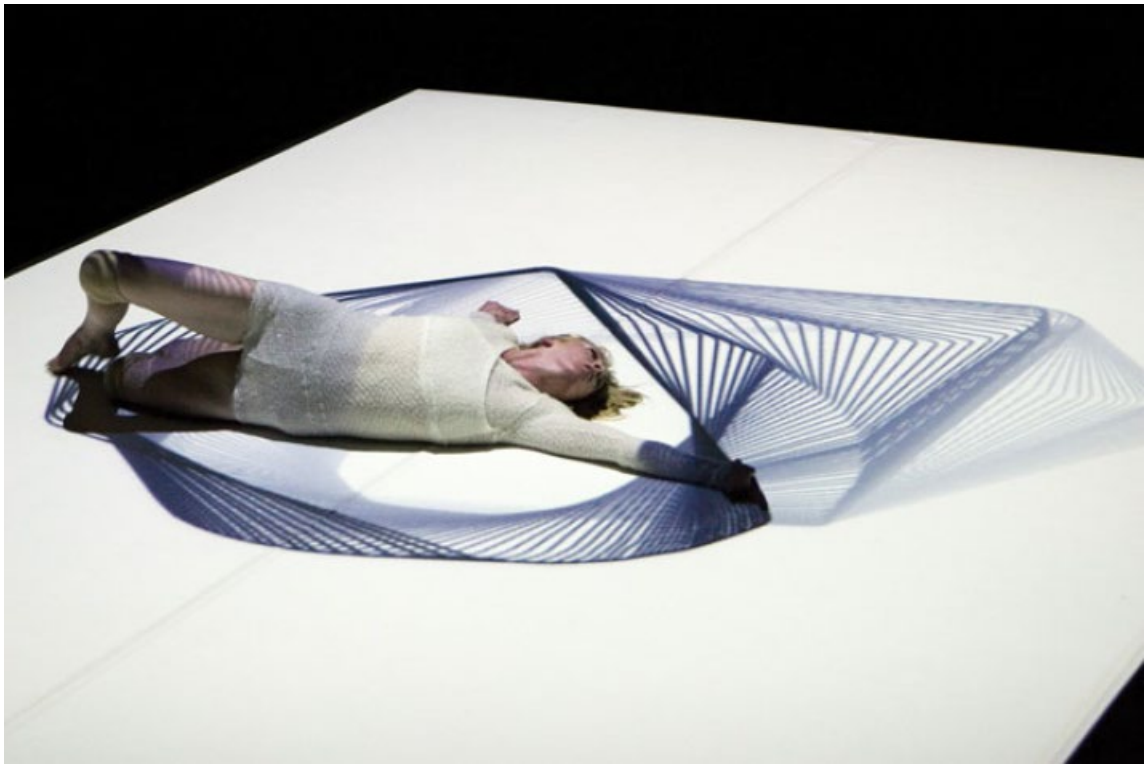

Fig. 2.8: Projected animations (Mullis, 2013).

2.2.2 Early enhancements in performance

The integration of technology with dance performance can be traced back to the late 19th century, when choreographers first began experimenting with visual effects to enhance audience experience. An example was Loie Fuller's Serpentine Dance in 1892, which demonstrated an early understanding of how technology could transform dance performance. Fuller combined coloured lighting with flowing silk costumes, creating what

Sommer (1980b) described as "colour washing in iridescent streams across the moving fabric." This groundbreaking work established the foundation for how technology could extend the expressive capabilities of dance performance.

Building on Fuller's innovations, the early 20th century saw attempts to merge technology with live performance. In 1913, Valentine de Saint-Point pushed these boundaries further by incorporating projected mathematical equations onto multiple fabric screens during dance performances (Goldberg, 1988). This integration of abstract visual elements with dance movement represented a step toward the kind of digital augmentation we see in contemporary dance technology.

The concept of interactive performance with projected media emerged shortly after, notably in Winsor McCay's 1914 touring production of 'Gertie the Dinosaur.' While not strictly a dance performance, McCay's synchronization of live performance with projected animation (Bell, 2000) established a link and interplay between performers and visual media that would later become fundamental to motion capture applications in dance.

The 1920s marked a period of experimentation, particularly in cabaret and music hall settings, where artists refined techniques for combining live movement with projected media. French magician Horace Goldin's work synchronizing live dance with slowed-down film (Bowers et al., 1998) was particularly significant, as it demonstrated early attempts to manipulate time and motion in performance - concepts that would also be used in motion capture technology.

After a brief pause during the 1940s and early 1950s, technological integration in performance took a significant shift with the establishment of Laterna Magika in 1958. Their combination of film, multiple screens, and special effects (Kennedy, 2003) marked the beginning of more complex technological systems in performance, laying groundwork for future digital performance environments.

The late 20th century saw an acceleration in the adoption of visual media in performance, driven by the increasing accessibility of video technology (Dixon, 2007). This

democratisation of technology meant choreographers could more easily experiment with visual elements in their work (Tomlinson, 2014). As motion capture technology has become available, it has enabled new forms of interaction between real and virtual dancers on stage, influencing how choreographers can approach performance creation. The introduction of computers and their creative capabilities has opened possibilities for artistic expression in dance (Mitchell et al., 2018), leading to the motion capture and AI systems being developed today.

2.2.3 Motion capture in performance

An early application of motion capture technology in a performance setting is Merce Cunningham's piece *Biped*[8] (1999), which was created in collaboration with visual artists Shelley Eshkar and Paul Kaiser. The result is a very large animation of dancers projected on a scrim, a finely woven lightweight transparent fabric, which is visible as the real-life dancers performed. The animations were derived from the motion capture data of the dancers, where sensors tracked and recorded their movements in detail (Felciano, 1999). Accompanying the dancers' projection onto the scrim was animated geometric shapes in varying width and colour, making their own rhythmic pattern. The constantly changing visuals seemed to change the perceived size of the live action dancers as the perspective of the projected shapes changed. The live action dancers wore tightly fitted reflective costumes would catch the light and seemed to duet with their projections. The reflective costumes, off-stage lighting and projected shapes illuminated from dark backdrop.

In the same year Kaiser and Eshkar produced *Ghostcatching*[9] in collaboration with choreographer Bill. T. Jones. The installation combined dance and digital imagery created from motion capture of dancers (Chang, 2019). The ghost-like computer generated figures perform alongside the dancers, mimicking their motions. An optical motion capture system was used to capture the movements of Jones, one of the most accurate forms of motion capture available. Kaiser commented "it's true that motion-capture is a process of subtraction, of taking away. The infrared cameras have eyes only for the reflective markers worn by the performing bodies, and not for the bodies themselves" (Rindler et al., 1999). Although the motion capture representation did not speak for the motion of muscle and skin of the dancer, it was not required for the animation that was produced. Kaiser and Eshkars'

[8] Biped, https://www.mercecunningham.org/the-work/choreography/biped/
[9] Ghostcatching, http://openendedgroup.com/artworks/gc.html

projected figures were abstract enough to get an idea of the performer, but the details within the performer was not required for their output.

In the Spring of 2000, David Saltz directed Shakespeare's *The Tempest*[10], a performance whereby live and digital characters are combined on stage. Ariel, the digital character in the play, is a ghostly figure who is not in the same reality as the live action actors. Ariel also has the capability to appear or disappear suddenly and transform into different shapes. Saltz used real-time motion capture technology to animate the character of Ariel, and also give them the characteristics of another worldly figure in the play. The real-time motion capture was displayed on a 32x18 foot rear projected screen. The audience could see the motion capture actor during the performance, with sensors attached to her head, elbows, wrists, hands, knees and ankles. The motion capture data was processed by a computer nearby and the three-dimensional animation of Ariel was brought to life. In addition to motion capture, voice recognition software captured the voice of the actor and animated the mouth of the projection to match. The outcome is a combination of motion capture and voice capture controlling Ariel simultaneously. In Figure 2.9, Jennifer Snow (portraying Ariel) is captured during the rehearsal for the staging of The Tempest.

Fig. 2.9: *The Tempest* (Saltz, 2001).

[10] The Tempest, https://www.newswise.com/articles/innovative-production-of-shakespeares-the-tempest

Saltz goes on to mention the purpose of using technology in his production as: “We propose a new way to use technology that enhances the text, broadens the expressive range of actors and redefines what it means for a performance to be live” (University of Georgia, 2000).

While live motion capture adds an element of interactivity to the performance, some performances use pre-recorded motion capture animation. In the case of Landing/Place (2004), choreographer Bebe Miller incorporates a hybrid approach. Motion captured animation that is pre-rendered and displayed as the dancer is performing has been processed in advance, and does not have to go through the processing stage live. The advantages of pre-recording motion capture animation is that the raw motion capture data can be processed and cleaned to avoid any erroneous keyframes in the data. Furthermore, the aesthetic quality of the animation can be improved by rendering the animation with realistic looking lights, shadows, and materials. This high-quality display of the animation is not possible in real-time, and depending on the length of the animation, could take a powerful computer hours to render (Jobson, 2015; Kocabas et al., 2020). Miller used projections to depict the elements, fire, water, air and earth, interacting with the dancers (Segal, 2005). The dancers movement controlled a dense cluster of dots, and later revealed that each dot is a bird, shaping the movement of the performer (Miller et al., 2006).

Paul Kaiser, Shelley Eshkar and Michael Girard created the visuals for the real-time motion capture performance of *Motion-e*[11]. The performance was developed over a three-year period involving choreographers, visual artists and engineers. The stage was specifically fitted with a motion capture system and the performers wore motion capture suits with reflective markers. In order for the motion capture system to work effectively, issues such as lighting interference, floor reflectivity and capture space dimensions had to be considered. A library of gestures were developed and triggered based on the correlated movement of the dancers. Artificial intelligence (A.I) was used for the production, where positions of the markers attached to the dancers were analysed to discern recurring patterns. Once a particular pattern generated by the dancers is observed, animated shapes connect the dancers, which is projected on the transparent screen in front of them. The projected animation in partnership with the performers communicated a feeling of connectedness between the projection and the dancers (David, 2005). Significant computing power was

[11] Motion-e, https://www.electronicdesign.com/markets/lighting/article/21765559/realtime-motione-capture-makes-dance-a-digital-art

needed to process the visuals in real-time, two Macintosh G5 computers with an additional machine for backup was needed. The custom fitted motion capture system for the stage incorporated 16 infra-red cameras positioned around the stage.

In an attempt to avoid the need for dancers to wear restrictive motion capture suits fitted with markers, (Latulipe et al., 2011) integrated a novel approach whereby the dancers would perform holding wireless gyroscopic mice. The mice tracked where the dancers were spatially, in addition, the dancers could trigger secondary animations by pressing the buttons on the mice. The fact that motion capture suits were not required allowed the producers to create costumes which fitted the performance and did not limit the movements of the dancers. Furthermore, transitioning from the rehearsal space to the performance space was not an issue, due to the method of motion capture. The choreography and the projected visualizations were developed simultaneously. As the choreography was being developed the visual artist was inspired to create based on the motion of the dancer and vice-versa. This created a feedback loop where imagination could be inspired to create new ideas. The lighting of the performance needed to be balanced with the projections. An influx of light from the stage spilling onto the projected animation would dilute the effect. A balanced lighting profile was required to provide sufficient lighting to the performers while retaining the contrast of the animation on the projected wall.

Haag (2008) examines the application of inertial motion capture in live performance, with particular focus on dance contexts. The research explains how inertial motion capture differs from optical systems - while optical systems use cameras to track points in space, inertial systems employ active sensors to measure orientation through wireless transmission to a computer. Although inertial systems may experience translational errors due to the lack of a global reference point, they offer advantages in portability, minimal setup time, and freedom from line-of-sight limitations that affect optical systems.

The study explores several practical applications through projects including "A Brush with the Real World," "Chasing Shadows," "Private Eyes," and "Motionics," demonstrating various capabilities of real-time motion capture in performance settings. These projects highlighted both the potential and limitations of the technology, such as the ability to create real-time visual effects and animation, while also revealing challenges with foot sliding, vertical tracking accuracy, and hardware constraints from wearing sensors.

A significant finding was that while inertial motion capture could function in real-time across various performance scenarios, the animation produced wasn't always consistently stable or robust. The research also identified limitations regarding the obtrusive nature of the electronic equipment worn by performers, which could only be partially concealed through careful costuming. However, Haag notes that many of these technical limitations were being addressed through ongoing technological developments, suggesting increasing potential for the technology's application in live performance contexts.

The research demonstrated that inertial motion capture systems could effectively enable complex interactions between performers and projected visuals, with specific body movements controlling various aspects of the virtual environment. This capability opens up new possibilities for choreographic expression and audience engagement, though considerations around system reliability and hardware obtrusiveness need to be addressed for regular performance use.

Dalmazzo & Ramírez (2019) investigated the use of machine learning for classifying violin bowing gestures, acknowledging the crucial role of gestures in musical expression and sound production. Their study employed a machine learning model combined with motion data captured using a Myo device, which recorded inertial motion information from the violinist's right forearm.

The Myo[12] (Fig. 2.10) is a wearable armband device designed to detect and interpret hand and arm movements. It contains sensors that can recognize various gestures, enabling users to interact with and control different digital devices such as computers, gaming systems, and robotic equipment. The device translates physical movements into digital signals, allowing for remote control and coordination between multiple connected devices through gestural input.

They analysed several bow techniques (Détaché, Martelé, Spiccato, Ricochet, Sautillé, Staccato, and Bariolage) performed by a professional violinist, synchronizing motion data with audio recordings. By extracting features from both motion and audio data, their model achieved over 94% accuracy in identifying different bowing techniques. The success of their approach suggests potential applications in violin pedagogy, where students could benefit from real-time feedback on their bowing technique.

[12] Myo, https://wearables.com/products/myo

In the violin performance studies, researchers have employed various sensor configurations, from simple accelerometers attached to violin bows to more complex IMU setups capturing detailed motion data. These sensor-based systems offer advantages in terms of portability and cost compared to traditional optical motion capture systems, though they may face challenges related to battery management, possible magnetic interference, and potential drift in sensor accuracy over time. The effectiveness of sensor-based systems has been demonstrated across various performance contexts, from musical instrument performance to interactive installations and pedagogical applications.

Fig. 2.10. The Myo device.

(Qianwen, 2024) investigated the application of wearable motion sensor devices for dance movement recognition, proposing a system that combines motion capture technology with wearable devices. The study emphasizes the advantages of wearable devices over computer simulation in capturing the detailed dynamics of dance movements. The research outlines a comprehensive process for collecting and preprocessing dance action data, constructing a human motion capture data-driven system capable of obtaining necessary movement data through wearable devices.

The system architecture comprises three main components: an upper computer, an embedded processor, and a lower computer, designed to capture voltage and acceleration signals during dance movements. The hardware device transmits collected dance action data wirelessly to facilitate subsequent processing and analysis. The system achieved significant accuracy in recognizing dance movements, with recognition rates between 44% and 75% depending on the complexity of the movements and the methods used.

A notable aspect of the study was its use of transfer learning to improve recognition accuracy, which increased from 34.9% to 75.5% after implementation. The system

demonstrated particular strength in handling complex dance movements, similar movements, and self-occlusion scenarios, showing improved recognition rates compared to traditional methods. The research highlights the potential of wearable motion capture technology in dance education and training, offering a cost-effective and efficient approach to movement analysis and feedback.

The study concludes that while motion capture technology was initially limited to animation applications, its integration with wearable devices offers new possibilities for dance movement recognition and analysis. This approach provides a comprehensive and accurate method for capturing dance postures while maintaining relatively low costs and personnel requirements compared to traditional motion capture systems.

An example of where the depth sensing technology of the Kinect was in *as.phyx.i.a* [13] (2015). Maria Takeuchi and Frederico Phillips used the Kinect to explore human motion in dance. Two first generation Kinects were used to capture the movements of Shiho Tanaka, the data was processed and rendered in a near photo-realistic environment (Jobson, 2015). The monochrome rendered animation of *as.phyx.i.a* is lit and produced in a way that feels more realistic than other examples of motion capture animation in performances. In order for animation to be displayed in real-time, a considerable amount of processing power is required. Should the computational demands placed on the animation creation and real-time lighting become excessive, the system may face challenges in maintaining its real-time functionality. To create the photorealistic aesthetic in *as.phyx.i.a*, Takeuchi and Phillips presented the animation as an experimental film, where the processing and rendering was made before the presentation. This processing behind-the-scenes gave the producers time to finesse the animation so that it can be presented in a more realistic way than if it was presented in real-time. The animation was lit by a shaft of light from above with thousands of points representing the shape of Tanaka. Each point is connected to another with a straight line, where the audience feels a sense of restriction as Tanaka moves as though confined in the connected point cloud (Chang, 2019). In Figure 2.11, the experimental film as.phyx.i.a is portrayed, showcasing the utilisation of two Kinect devices for capturing the dancer's motion.

[13] As.phyx.i.a, https://dbini.com/as-phyx-i-a/

Fig. 2.11: *as.phyx.i.a* (Chang, 2019)

The Microsoft Kinect[14] dramatically reduced the cost of motion capture and opened-up wider use. By delivering a markerless technology, the Kinect frees the dancer of restrictive suits as required by optical motion capture systems. It is ironic that Takeuchi and Phillips portray an animation that feels so restrictive, by using a method of motion capture that frees the dancer of a restrictive motion capture suit. While the Microsoft Kinect has clear advantages and is more accessible to motion capture practitioners, it is important to mention that there are limitations. The Microsoft Kinect system relies heavily on the silhouette of the subject and has issues recognizing if the front of the subject is facing the camera or the back (Singh et al., 2022). Subtle movement like toe-tapping is not registered accurately by the Kinect and causes errors (Galna et al., 2014). Lachat et al. (2015) reported that the Kinect does not fare well in outdoor settings with interference occurring if the subject is strongly backlit or if the sun's rays hit the sensor.

[14] Microsoft Kinect, https://learn.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows

2.2.4 Motion capture technology typically suited to performances

2.2.4.1 Depth cameras

In November 2010, the Microsoft Kinect was released. The Kinect was created to track the movements of gameplayers interacting with the X-Box[15] gaming console. The players interacting with the console are not required to wear any specialized suits or markers. The Microsoft Kinect device houses a structured light camera, where a pattern of light is emitted and read by the camera. The depth information of the player in their environment is calculated by projecting a known pattern of light from the camera into the environment and any distortions of that light are perceived as depth by the camera (Colyer et al., 2018). The Microsoft Kinect is able to track the skeletal structure of an entire body. Choreographers and dancers benefitted from the technology because it was commercially available and significantly cheaper than optical systems. A wide range of industry adopted the depth sensing capabilities of the Microsoft Kinect (Q. Wang et al., 2015). In 2013 a second generation of the Kinect was released with the X-Box console and the following year a standalone version was available. The Kinect2 featured a time-of-flight camera, which emits a pulse of light. The depth information of the environment is calculated by the time it takes the pulse of light to return.

The depth cameras of the Kinect capture a range of up to 5 square meters (Lachat et al., 2015, Protopapadakis et al., 2017). The front and back sides of the subject are indistinguishable by the Kinect's depth camera (Protopapadakis et al., 2017, Galna et al., 2014). This limitation does not allow the subject to face away from the camera, which causes an issue when recording the motion of a dancer.

The Kinect represents an important milestone in motion capture technology, employing a sophisticated system that combines multiple sensors with machine learning capabilities. The system integrates hardware components including a color camera, depth sensor, and inputs from both gyroscope and accelerometer, operating across multiple coordinate systems in both 2D and 3D space.

The skeletal tracking system is capable of tracking 32 distinct joints, creating a hierarchical skeleton that flows from the body's center outward to the extremities. The system measures joint positions in millimeters within the depth sensor's frame of reference, with each joint

[15] X-Box, https://www.xbox.com/en-AU/

forming its own right-handed coordinate system. All joint coordinates are absolute within the depth camera's 3D coordinate system.

The machine learning implementation processes multiple sensor inputs simultaneously, estimating joint positions and orientations in real-time. It can create accurate skeletal tracking even with partial occlusion and adapt to different body types and movements, converting raw sensor data into usable skeletal tracking data (Azure Kinect Body Tracking Joints, 2019).

The Kinect detects pose estimation using decision forests. The system takes a single depth image as input and classifies each pixel into one of 31 different body parts (e.g. left upper arm, right hand, etc.) using randomised decision trees. These trees make their decisions using simple depth comparison features that look at the difference in depth between pixels. The system is trained on a large synthetic dataset of depth images with corresponding body part labels, generated by rendering 3D character models in different poses captured from motion capture data. After classification, the system uses mean shift clustering to find the modes of each body part's probability distribution, which gives the final 3D joint position estimates. The approach runs at 200 frames per second on the Xbox 360 GPU and can accurately predict joint positions across variations in body shape, size, clothing and pose. The paper demonstrates that breaking down pose estimation into a per-pixel classification problem, combined with a large synthetic training set, enables robust real-time skeleton tracking without requiring temporal information or kinematic constraints (Shotton et al., 2011).

The Kinect camera system has been used as part of a physical therapy game called Supernova to track patients' movements during exercises. The system uses Microsoft's GestureBuilder software (part of the Kinect SDK) to detect specific movements that patients make, like swatting, reaching, squatting, and pushing. When patients perform these movements in front of the Kinect camera, their actions are translated into gameplay within a space-themed environment displayed on a screen or wall. The Kinect setup is designed to be simple - requiring just a single camera connected to either an Xbox or laptop - making it accessible for both clinical settings and home use. The researchers chose the Kinect because it allows for markerless motion capture (meaning patients don't need to wear any special suits or sensors) and can track movements accurately enough to provide useful feedback during physical therapy exercises (Hobby et al., 2017).

There are three methods of how depth cameras record three-dimensional space:

Time-of-flight

- Time-of-flight cameras illuminate the scene with light and calculate depth by measuring the length of time the light returns to the sensor.

Structured light

- Structured light sensors project a pattern of light out into the scene and the depth is calculated based on how the pattern is deformed, falling onto objects or people in the scene.

Light detection ranging

- Light detection ranging (LiDAR) calculates the depth in a scene by pulsating infra-red light. LiDAR is generally considered more accurate than time-of-flight and structured light depth cameras because LiDAR systems use multiple pulses of light. (Joao Alves, 2021)

Time-of-flight cameras are susceptible to errors if there are surfaces that are highly reflective or very dark. They are suitable in outdoor settings and can handle direct sunlight, unlike structured light cameras. Direct sunlight can cause interference with structured light cameras. Depth cameras are favoured by dancers due to the absence of motion capture suits needed to capture depth in the scene (Meador et al., 2004). Depth cameras are considerably less expensive than optical systems and are easily transported due to their small size. Other limitations of depth cameras such as the Kinect are the lack of fine detail and limited resolution. The Kinect may not capture fine details such as facial expressions and its limited resolution can affect the accuracy of depth perception. While real-time, the Kinect can exhibit slight latency between the user's movements and the display of those movements on the screen.

The RealSense camera system employs a three-component optical arrangement consisting of a conventional RGB camera, dual infrared cameras, and an infrared projector working in concert to generate accurate depth perception. The foundational operation begins with the

infrared projector casting an invisible pattern onto the scene, which serves as a reference point for depth calculation (Tadic et al., 2019).

The system's core functionality relies on stereovision principles, utilising two infrared cameras positioned at distinct vantage points, similar to human binocular vision. These cameras capture simultaneous images of the scene, including the projected infrared pattern, from slightly different angles. The system's processor then analyses the disparities between these two viewpoints, calculating depth values for each pixel by correlating corresponding points between the left and right camera images.

The processing system compiles these depth calculations to generate a depth frame, which provides three-dimensional spatial information about the scene. When multiple depth frames are combined in sequence, the system produces a continuous depth video stream, enabling real-time spatial tracking and analysis. This processing is handled by the RealSense D4 processor, which processes 36 million depth points per second, making it suitable for applications requiring high-speed data processing.

The integration of the conventional RGB camera alongside the depth-sensing system enables the capture of both colour imagery and depth information simultaneously. This dual functionality proves particularly valuable in various lighting conditions, as the infrared system can operate effectively even in suboptimal lighting scenarios. The resulting combination of colour and depth data enables applications such as movement tracking and gesture recognition (Tadic et al., 2019).

An example of RealSense cameras in performing arts can be found in the Co:Lateral project, where depth cameras were used to create a digital dance performance. In this project, Moura et al. (2019) used RealSense technology to capture dancers' movements and transform them into dynamic visual representations. The cameras tracked the performers' movements in real-time, generating what the researchers called "virtual doubles" - digital versions of the dancers that could interact with their physical counterparts. This technology allowed the performers to interact with their own digital reflections through a transparent display positioned between them and the audience. The system was particularly effective at capturing detailed movements, creating volumetric representations of the dancers' bodies, and generating visual effects that responded to their gestures. For instance, when dancers moved, the system could create trails of light following their movements or transform their gestures into abstract patterns. One of the most striking elements of the performance

involved the dancers interacting with virtual vertical bars that initially appeared rigid like prison bars but would respond fluidly to the dancers' touch, eventually transforming into volumetric representations of the moving body (Moura et al., 2019). This use of depth camera technology not only enhanced the visual aspects of the performance but also created new possibilities for artistic expression through the interaction between physical movement and digital response.

While RealSense cameras commonly utilise both RGB and infrared cameras working together, the infrared component alone can be effectively used for performance tracking. A notable example of infrared camera technology in performance art can be found in the production of Encoded (Johnston, 2015b). In this performance piece, the creators employed a specialised camera fitted with an infrared filter, which only captured infrared light while blocking visible light. This setup, combined with carefully positioned infrared lighting, allowed the system to track performers' movements without visible light interference. The tracking data was then used to drive a fluid simulation system, where performers could effectively 'stir' virtual fluids with their movements, creating dynamic visual effects that responded naturally to their actions. The sophistication of this system allowed for real-time adjustments to various parameters such as viscosity and color, enabling the creation of different visual states that could be smoothly transitioned between during the performance. This example demonstrates how infrared tracking technology can be effectively utilised in isolation to create responsive, interactive performance environments, showing that while RealSense cameras offer multiple tracking methods, the infrared component alone can be a powerful tool for performance applications.

LiDAR (Light Detection and Ranging) technology represents an advanced sensing system that operates on principles similar to radar but utilises light waves instead of radio waves. The fundamental operation of LiDAR involves the emission and detection of laser light pulses to measure distances and create three-dimensional representations of the environment (Raj et al., 2020).

The operational process of LiDAR consists of several key components working in sequence. The system initiates by emitting rapid pulses of laser light, typically in the invisible spectrum. These pulses reflect off various objects in the environment, including structures, vegetation, vehicles, and terrain. Upon reflection, the system's detectors capture the returning light pulses. The time-of-flight measurement between emission and detection allows for precise distance calculations. Through rapid repetition of this process combined with scanning mechanisms, the system generates comprehensive three-dimensional spatial data.

This scanning process produces what is known as a point cloud, which provides a detailed three-dimensional digital representation of the scanned environment. The technology's ability to capture thousands of measurements per second with high precision has led to its widespread adoption across multiple fields.

The applications of LiDAR technology have expanded significantly across various sectors. The technology serves crucial functions in autonomous vehicle navigation, geographical surveying, robotics, architectural planning, and agricultural management.

Recent technological advances have focused on reducing the size, weight, and cost of LiDAR systems while maintaining or improving their accuracy and performance. This evolution has transformed LiDAR from a specialized, expensive instrument into a more accessible technology suitable for broader applications (Raj et al., 2020)

LiDAR technology has found novel applications in the field of performance arts and dance analysis. As demonstrated by Naik and Supriya (Smys et al., 2021), LiDAR's ability to capture precise three-dimensional data makes it particularly valuable for analyzing and classifying dance movements and postures. The technology can capture surface data of dancers' movements, creating detailed point cloud representations that preserve the spatial relationships and depth information crucial for understanding complex dance forms. In the context of Indian classical dance, for example, LiDAR scanning can capture the intricate hand mudras, leg postures, and body positions that characterize different dance styles. This three-dimensional data collection offers significant advantages over traditional two-dimensional video or image analysis, as it provides complete spatial information about the performer's movements and positions. This application of LiDAR technology not only aids in the classification and analysis of dance forms but also has potential applications in dance education, allowing for precise documentation and study of dance techniques in a digital format.

2.2.4.2 Sensor-based motion capture

Sensor-based motion capture systems include inertial sensors or inertial measurement units (IMU) that are used to determine the motion of the performer. The sensors consist of a combination of accelerometers, gyroscopes and magnetometers which calculates the force, angular rate and position of the body. While a suit is needed to attach the sensors to it, the sensors are not required to be visible like optical systems. The sensors are not reflective and do not need to be visible to the camera. Performers are able to wear these sensors

underneath costumes and still record relevant data. The sensors are battery-powered and could require calibration, which limits the duration of continuous use. Suits with a magnetometer can be susceptible to magnetic interference and may not be suitable for all locations. Sensor-based motion capture does have similar limitations as the Kinect system like line-of-sight dependency, limited range, sensitivity to lighting conditions, and limited capture of fine detail, but not as severe as Kinect systems. Sensor-based systems can cause inaccuracy of the detection over time due to the subtle drift of sensors. Sensor-based motion capture brands are Rokoko, Xsens and Perception Neuron.

Researchers have explored other sensor-based approaches for motion capture and analysis. (Maranan et al., 2014) developed EffortDetect, a system that used a single wearable accelerometer to recognize Laban Movement Analysis (LMA) effort qualities in real-time. While achieving only 70.71% average accuracy compared to optical systems, their work demonstrated that even simple sensors could capture meaningful movement qualities. The authors argued that accelerometer-based approaches had advantages in being highly portable, usable in varying environmental conditions, and more practical for public installations or performances. However, like the Kinect, such systems were limited in their ability to capture fine movement details and faced challenges with accuracy and consistency.

While inertial and magnetic motion capture systems are indeed significant technologies in the field, and their benefits and limitations are acknowledged in this section, this research deliberately focuses on optical/markerless systems using standard cameras. This focus was chosen primarily because consumer-grade cameras are more readily accessible to the target users of this research, being performing artists and choreographers working with limited resources.

Furthermore, the field of machine learning-based motion capture using standard cameras is experiencing rapid advancement and innovation. This emerging technology presents new opportunities and challenges that warrant focused investigation. While this thesis does discuss inertial and magnetic systems to provide context and completeness, an in-depth analysis of these systems falls outside the scope of our primary research objectives, which focus on making motion capture technology more accessible through widely available consumer hardware.

This focused approach allows for a more thorough examination of the specific challenges and opportunities presented by machine learning-based optical systems, while acknowledging the broader ecosystem of motion capture technologies available to practitioners.

Table 2.3 below lists the various types of motion capture systems using in performing arts and their characteristics.

Characteristics	**Optical Motion Capture**
Cost	High
Portability	Limited (fixed studio setup)
Accuracy	High
Setup requirements	Complex Controlled environment Calibration needed
Wearable required	Reflective markers and suits
Limitations	Line of sight required Markers can detach Space constraints High maintenance costs
Benefits	High accuracy Reliable for complex movement Animation/Games industry standard
Real time capability	Yes
Best suited for	Professional productions with high budget

Characteristics	**Depth Cameras (Kinect/RealSense)**
Cost	Low – moderate (commercially accessible)
Portability	Highly portable
Accuracy	Moderate
Setup requirements	Simple setup, minimal requirements
Wearable required	None (markerless)
Limitations	Limited range (up to 5m)

	Issues with sunlight Cannot distinguish front/back Limited resolution
Benefits	No markers needed Cost effective Easy to use Accessible
Real time capability	Yes (with some latency)
Best suited for	Interactive installations Small performances

Characteristics	**Sensor-Based Systems (IMU)**
Cost	Moderate
Portability	Highly portable
Accuracy	Moderate with drift over time
Setup requirements	Simple setup, requires calibration
Wearable required	Sensors/suits with IMU's
Limitations	Battery dependency Magnetic interference Sensor drift Calibration needs
Benefits	Works under any lighting No occlusion issues Can be work under costumes
Real time capability	Yes
Best suited for	Mobile performances

Characteristics	LIDAR
Cost	Historically expensive but becoming more accessible
Portability	Moderately portable
Accuracy	High
Setup requirements	Moderate setup complexity
Wearable required	None (markerless)
Limitations	Processing power requirements Complex for real-time use
Benefits	Precise 3D data Works in various lighting Captures detailed spatial data
Real time capability	Depends on processing power
Best suited for	Spatial analysis

Table 2.3 Motion capture systems and their characteristics.

The Movement and Computing (MOCO)[16] conference represents a significant component of the computer vision and performance research landscape. MOCO functions as an international conference bringing together academics and practitioners who study computational approaches to movement and generation of movement information.

The conference operates at the intersection of art and science, addressing specific technical challenges in representing embodied movement knowledge within computational models. While movement itself centres on bodily experience, its computational implementation requires abstracting lived embodied cognition into representational models.

The field of movement computation research has expanded across multiple disciplinary domains. Interaction design has incorporated movement principles into interface development and user experience. Human-Computer Interaction (HCI) has integrated movement analysis into new forms of computer interfaces and interaction paradigms. In education, movement computation has informed new teaching methodologies and learning

[16] Movement and computing, movementcomputing.org

assessment tools. Machine Learning applications have begun to analyse and generate movement patterns, contributing to both artistic and functional applications.

The MOCO community maintains a moderated mailing list that serves multiple functions within the field. The list disseminates conference announcements and calls for papers to keep the community informed of academic opportunities. It facilitates event notifications and enables the sharing of datasets and software among researchers. The platform also hosts research discussions between students, artists, and researchers working in movement computation.

Within computer vision applications for performance, MOCO's work documents the relationship between movement experience, cognition, and computational representation. Their research addresses the technical requirements of movement analysis while acknowledging the need to account for contextual and embodied aspects of movement in computational frameworks (Movementcomputing.org, n.d.)

2.2.5 Conclusion

Choreographers have often sought to enhance the audience experience of a performance by looking for creative uses of technologies available at the time. From the simplicity of long flowing costumes interacting with coloured lighting (Sommer, 1980b) to photorealistically rendered 3D animation (Jobson, 2015), the choreographer is drawing an emotion for the audience to experience. Technology in performance is prevalent and even expected from choreographers who have a reputation to produce technologically infused dance productions based on the motion of the dancer (Downie & Kaiser, 2018).

As motion capture technology has advanced to facilitate the needs of film, television, and games, choreographers have incorporated this technology for their own needs. Optical motion capture has the accuracy to animate believable digital avatars but implies constraints that limit the range of possible performances. Optical motion capture systems like Vicon have several drawbacks that limit their accessibility and flexibility. The initial setup and ongoing maintenance costs for the cameras, software, and computing hardware can be prohibitively high, making optical motion capture less feasible for smaller budgets. The system is not portable due to its stationary camera setup, so users are confined to a dedicated studio space. Setting up the system requires a controlled occlusion-free

environment, which may not be available to users with limited space. The need for subjects to wear specialised suits and have markers affixed to their bodies can also impact comfort and restrict natural movement. Additionally, optical systems rely on maintaining a direct line of sight between markers and cameras, which can cause data loss when markers are temporarily blocked from view. Markers can detach if the subject is moving in a way that causes excess strain or friction against the suit material. Precise camera calibration is required and can be time-sensitive, especially for large capture volumes. Overall, while optical motion capture produces high-accuracy data, the logistical constraints make it less adaptable than some other motion capture methods.

Restrictive motion capture suits may prompt dancers and choreographers to gravitate towards markerless systems like the Microsoft Kinect depth camera or sensor-based systems for their price point and portability. However, these systems also have their limitations: Kinect systems are constrained by the inability to capture details and their restricted resolution. Facial expressions may elude its capabilities and its limited resolution can compromise depth perception. Despite real-time functionality, the Kinect can have slight latency, causing a delay between the user's motion and the representation on the screen. The Kinect system also grapples with a finite range, rendering the front and back of the subject indistinguishable. Direct sunlight can cause interference and the Kinect needs supplementary equipment, such as a computer, to compute depth data.

Sensor-based motion capture methods present several noteworthy limitations. Firstly, these systems often rely on battery-powered sensors which could require calibration, thus imposing constraints on continuous usage duration. Suits equipped with magnetometers may be susceptible to magnetic interference, limiting their suitability in diverse environments. Sensor-based motion capture shares certain constraints with Kinect systems, including line-of-sight dependency, a restricted capture range, sensitivity to lighting conditions, and challenges in capturing intricate details. It is worth noting that these limitations are typically less severe in sensor-based systems than in Kinect systems. A potential issue with sensor-based systems is the gradual drift of sensors over time which can result in accuracy degradation in motion detection.

2.3 Machine learning methods of generating motion capture for performing arts

2.3.1 Introduction

In contrast to conventional motion capture systems, an emerging approach utilises machine learning for human pose detection. This involves training a model to identify a set of keypoints on the human body using input from one or more RGB cameras (Kocabas et al., 2020, Shuai, 2021). Such models are trained on extensive datasets of millions of sample images with annotated keypoints (Ionescu et al., 2014). The massive volume of training data enables the model to generalise and reliably predict keypoints across a wide range of novel test images and poses, provided they demonstrate reasonable consistency with the distribution of examples used during training.

Machine learning models learn by analysing examples. They learn patterns and relationships in these examples. When they are presented with unseen data, they use what they have learned to make predictions or decisions. Modern machine learning is powerful because it can learn from millions of different examples. However, this can also be a problem. If we teach a model to recognise cats and dogs using cat and dog pictures, it won't know what to do when it is presented with something completely different, like a picture of a mouse. It has not learned what makes a mouse different from cats and dogs, so it's performance becomes unpredictable in such cases (Sahni et al., 2022).

This relates to how a model is trained using machine learning techniques for motion capture. If a model is trained solely on detecting keypoints on a human walking, it will likely encounter errors when presented with data of a person dancing. This is because the model is not familiar with the different movements and poses associated with dancing, as it has only been exposed to walking motions during its training process. For accurate motion capture of dancing, the model needs to be trained on a diverse dataset that includes various dance movements and choreography.

Before exploring specific applications of machine learning in dance and performance, it is worth noting Levin's (2006) foundational work in computer vision for artists and designers. Levin provided early documentation of how computer vision algorithms could be implemented by novice programmers to create interactive artworks. He outlined several basic but effective techniques that could be implemented by artists with limited programming experience, including frame differencing for detecting motion, background subtraction for

detecting presence, and brightness thresholding for object detection. Levin emphasised that while sophisticated computer vision algorithms require advanced knowledge of image processing and statistics, many widely used techniques could be implemented by novice programmers in a short timeframe. His work helped demystify computer vision for the arts community and provided practical examples of how these technologies could be integrated into interactive performances and installations. Levin also stressed the importance of considering both the software implementation and physical environment when designing computer vision systems, noting that clever physical staging could often achieve better results than more complex algorithms.

The process of observing and analysing movement is crucial when developing systems for motion capture and generation. (Alaoui et al., 2015) investigated techniques for observing movement in embodied design, identifying three key approaches: attunement (preparing to perceive sensory information), attention (consciously focusing awareness on specific aspects of movement), and kinesthetic empathy (understanding movement through one's own physical response). Their research highlighted challenges that remain relevant to modern machine learning approaches, particularly around articulating and translating observed movement qualities into computational systems. The authors noted that while observers can develop sophisticated awareness of movement qualities, converting these observations into formal specifications that can be implemented technologically remains difficult. This challenge persists in contemporary machine learning systems for motion capture, where translating nuanced human movement observations into computational features and parameters continues to be an active area of research.

2.3.2 Machine learning origins

Machine learning has been around since the 1950s, where one of the first uses was to recognise letters of the alphabet (Fradkov, 2020). Machine learning has since become relevant to many industries, including robotics (Bonsignorio et al., 2020), autonomous cars (Muthalagu et al., 2021) and medicine (Goyal et al., 2021), leveraging a computer's capacity to 'learn' based on training data. Machine learning-based computer vision techniques allow analysing the pixels in images or videos to understand and interpret the content. For instance, in pose detection, machine learning models can identify the locations of human joints and limbs within an image. Recent advancements in this field present a potential alternative approach to motion capture for the performing arts, potentially addressing the

cost and equipment limitations associated with marker-based or sensor-based motion capture systems. Once the joints of a human form can be identified, this information can be stored as data in the form of keypoint positions in three-dimensional (3D) space. A computer can learn how these joints move in relation to each other. When a machine learning model detects limbs and keypoints in an image or video, they can be tracked over time and reconstruct a mesh of the subject in 3D space.

The typical (contemporary) process for training a pose detection machine learning model starts with collecting lots of images showing people in different poses like sitting, standing, walking, running or dancing. These are used as the training images. The images are then pre-processed, often to ensure consistency or to adhere to the requirements of the target model (e.g. resizing, converting to grayscale, etc.). Each image is then annotated to highlight limb/skeleton/keypoint positioning on the human body (typically, this is manual, though automated and/or semi-automated approaches may also be deployed). This annotated image data is then used to train a neural network pose detection model (noting that the image is often cropped to focus only on the human, using an additional object detection algorithm). The model learns associations between features of the image and the annotated human body keypoints through this training process. After enough training, the model can predict the pose (a simplified skeleton from the collection of body keypoints) from new (previously unseen) images.

Motion capture systems need to not only track movement mechanically, but also capture the qualitative aspects of how movement is performed. (Fdili Alaoui et al., 2017) developed an innovative approach for computationally modelling Laban Movement Analysis (LMA) Effort qualities by combining multiple sensing modalities: positional data from motion capture, dynamic data from inertial sensors, and physiological data from electromyography (EMG). Through interviews with Certified Movement Analysts, they identified key features that correlate with different Effort qualities - using the norm of jerk from accelerometer data to detect Time Effort (Quick vs Sustained), EMG muscle activation patterns to detect Weight Effort (Strong vs Light), and measures of bodily extension/contraction from motion capture to detect Space Effort (Direct vs Indirect). Their evaluation showed that combining these multimodal data sources allowed for better characterization of movement qualities compared to using any single modality. This work demonstrates how machine learning approaches can move beyond just tracking positions to capture more nuanced expressive qualities of movement that are important for dance and performance applications.

2.3.3 3D pose detection

The technology of detecting keypoints in human pose has advanced significantly in the last few years. This technology is not aimed to a specific industry, but as a task that any industry can adopt and benefit from. For this reason, the literature available for machine learning pose estimation is general and not specifically for performing arts. In the realm of machine learning-based pose detection, models utilise a human body framework for pose detection. Within the literature, the most frequently employed 3D human body model is the skeleton, akin to a stick figure, which consists of interconnected bones and joints, effectively representing the human body's structure (Sarafianos et al., 2016).

To enhance the accuracy of these models, various constraints are implemented. For instance, limb constraints ensure that limb lengths, limb-length proportions, and joint angles adhere to specific predefined rules. Additionally, researchers have integrated rules for managing occlusions, enabling the generation of more realistic poses where certain body parts, such as legs or arms, can be hidden from view by others. Appearance constraints, rooted in the symmetry of left and right body part appearances (Gupta et al., 2006), are another commonly applied technique. Lastly, smoothness constraints come into play regarding joint angles, smoothing out abrupt changes between consecutive video frames. In Figure 2.12, a skeletal representation of the human body is presented, comprising various joints.

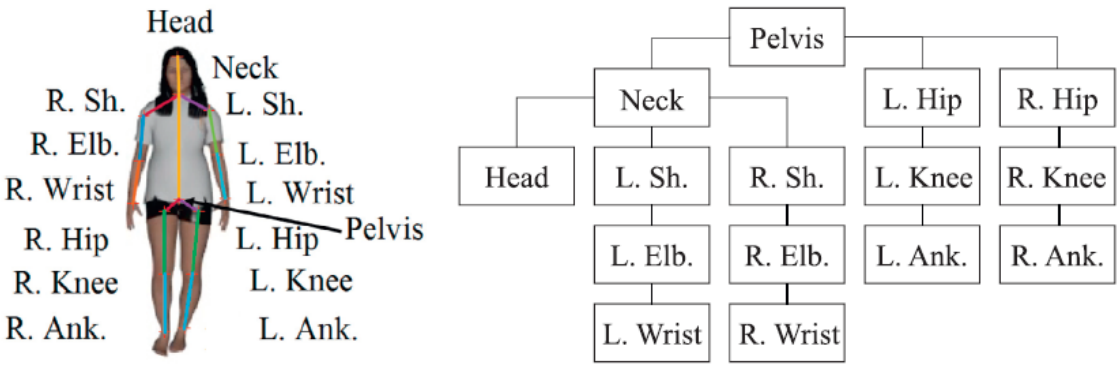

Fig. 2.12: Skeletal joints or keypoints (Sarafianos et al., 2016).

A machine learning model has the capability to comprehend human poses by analysing numerous images or videos and can be deployed to identify humans in newly introduced images of videos. Several datasets exist in which keypoint annotations are available, aiding in the training of models in 2D or 3D representations. When the individuals in the images are isolated, with the use of segmentation (He et al., 2017), the model can determine the locations of the joints based on the training data.

The H3.6M dataset (Ionescu et al., 2014) for example, consists of 3.6 million images of human pose captured by an optical motion capture system. Each image has a corresponding location of each joint in 3D space. A machine learning 3D pose detection model may learn this corresponding data and when presented with unseen images, attempt to predict 3D joint positions on new data.

2.3.4 A summary of 3D human pose detection models suitable for performing arts

2.3.4.1 Introduction

The area of machine learning for analysing and understanding human motion and poses is rapidly evolving and advancing. Newly improved models are released frequently. The models listed in this review are relevant as of February 2023. Below, is a table summarising the various machine learning models used in a performance context which will be discussed in more detail in this chapter. Table 2.4 provides a summary of the utilisation of machine learning in a performance context.

Name of work	**Machine learning usage**	**Outcome**
Learning to dance with a human (Mccormick et al., 2013).	Used to learn the movement of a dancer and when recognised, respond to that movement from learned choreography.	Projected stereoscopically in a performance with a dancer.
Embodied design of dance visualisations (Brenton et al., 2014)	Used to create interactive visuals that responded to the improvised movements of a non-professional dancer.	Created visualisations based on dancer's movement.

Generative choreography using deep learning long short-term memory (Crnkovic-friis & Crnkovic-friis, 2016)	Used to create relevant choreography.	The production of a creative tool to enhance the process of creating choreography in a particular style of dance.
Dance dance convolution (Donahue et al., 2017)	Used to automate the creation of step charts in a video game.	Video game enhancements.
Living archive (Bastien Girschig, 2019b)	Used to generate new choreographic material.	The model can replicate dance movement in the style of the artist that it was trained on, including iterations.
AI dancer scene (McDonald, 2019)	Used to create dances in the style of another dancer.	Dance sequences were generated from videos captured of the audience prior to a performance.
Beyond imitation: Generative and variational choreography via machine learning (Pettee et al., 2019)	Used to generate novel choreography based on input.	Authentic looking dance sequences generated.
A renowned dancer performed with an AI model – Can AI stimulate the dancer's creativity? (Tokui, 2020)	Used to train and interact with the stepping patterns of a dancer.	A performance where the dancer interacts with feedback from a model trained based on their steps.
Dance self-learning application and its dance pose evaluations (Choi et al., 2021)	Used as an aid to learn dancing.	An application that teaches students to dance.
Artificial intimacy (Miyoshi, 2021)	Used to interact with a dancer, by imitating dance movement sequences.	Authentic looking dance sequences generated.
Dance-on (Payne et al., 2021)	Used for pose detection to create an interactive experience.	An educational application that creates animations that

		interact with the body's movement.
Forgery (Rose, 2021)	Used to generate instructions for performers to create novel choreography.	A performance where dancers movements are determined by a machine learning model.
Puppeteering an AI (Bisig & Wegner, 2021)	Used to control a 'puppet' based on training data from a dancer.	An animated skeleton.

Table 2.4: Machine learning usage in performing arts.

The section presenting various dance systems that use machine learning serves to illustrate the diverse range of performance contexts where this technology has been deployed. Rather than attempting to provide an exhaustive list, which would be beyond the scope of this thesis given the rapidly growing number of implementations, these examples were selected to demonstrate the breadth of applications in the field. The examples highlight how different practitioners and researchers have approached similar technological challenges in varying contexts, establishing that there is a robust community of practice in this space. This overview connects to the broader themes of the thesis by showing how our work builds upon and contributes to an active, evolving field where technology and performance intersect in numerous ways. While many other examples exist, these selected cases provide sufficient context to situate our research within the larger landscape of performance technology.

2.3.4.2 Monocular 3D pose estimation models

One method to estimate the 3D human pose and also the shape of the human from a single image is described by (Bogo et al., 2016) in their approach called SMPLify. A full 3D mesh is estimated from 2D body joint locations. This model predicts 3D joints from 2D keypoints by utilising a well-trained 3D human body model. This model is trained using thousands of 3D scans of human bodies, which helps it understand how human body shapes vary across the population and how they change with different poses. This model can be applied with very limited data because it already contains a wealth of information about human body shapes. SMPLify automatically addresses the shape of the subject as well by capturing pose and shape data from a 2D image. The SMPL model is based on thousands of 3D body scans and represents the surface of the human body as a mesh consisting of triangular geometry.

The 2D joints of the image are matched to the 3D mesh where the shape and pose is matched. Furthermore, SMPLify addresses the issue of interpenetration of the 3D mesh.

Previous work has estimated a 3D skeleton shape from 2D joints, which does not reveal if there is any intersection of the mesh from limbs. Intersection occurs when geometry passes through one another. In Figure 2.13, an illustration is provided featuring intersecting geometry where the characters hand intersects the body, traversing neighbouring geometry throughout the animation process.

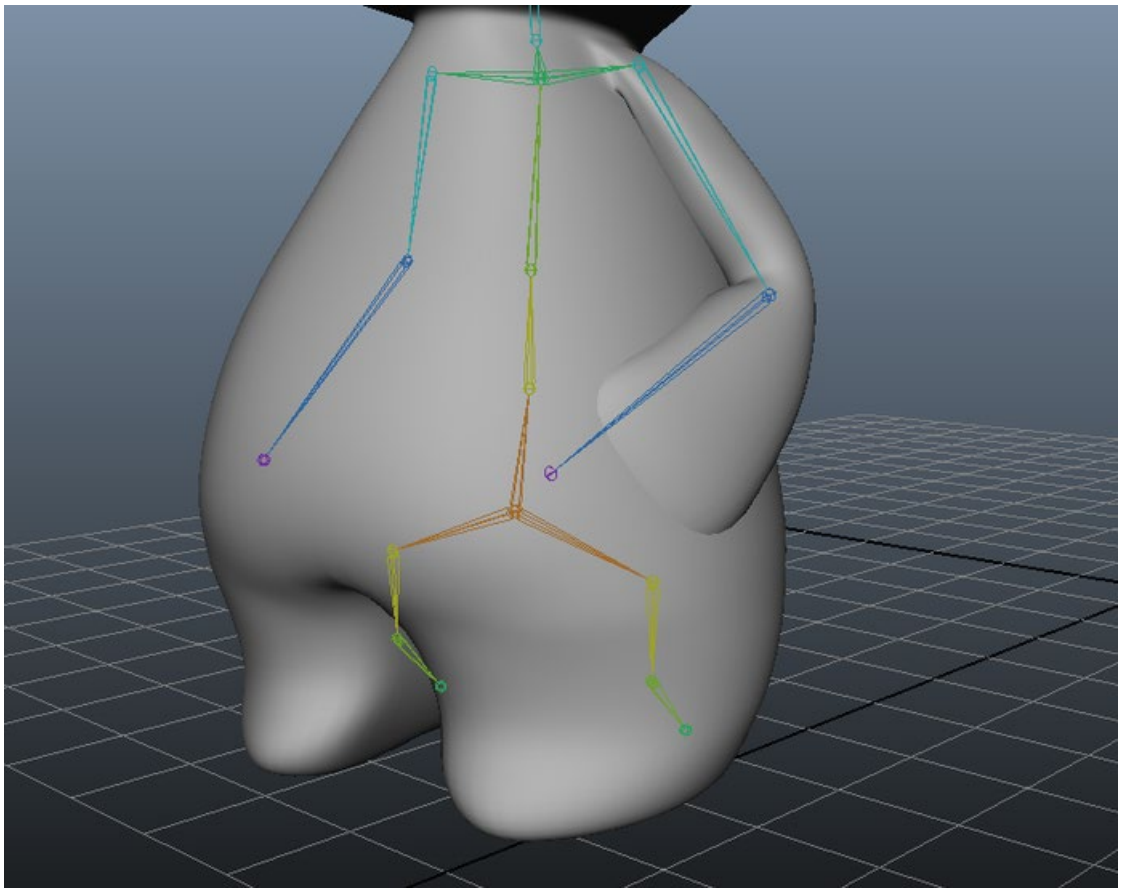

Fig. 2.13: Intersecting geometry (*Intersecting Geometry*, 2014)

Determining interpenetration on a mesh is done by defining capsules that resemble the body shape which allows calculating interpenetration efficiently. This method is evaluated on the

Leeds Sports Pose Dataset [17] (LSP) (Johnson & Everingham, 2010), the HumanEva-I[18] (Sigal et al., 2010) dataset as well as the Human3.6M[19] (Ionescu et al., 2014) datasets proving to be more accurate than previous methods. The LSP dataset contains 10,000 images of athletes participating in sports such as parkour, gymnastics, and athletic activities. The HumanEva-I dataset contains video sequences paired with corresponding 3D motion capture data. It includes seven calibrated video sequences with four subjects performing six common actions such as walking, jogging, gesturing, and others. The Human3.6M dataset contains a large set of 3.6 million 3D human poses paired with corresponding images. The images feature eleven professional actors, captured in seventeen different scenarios. The scenarios range from discussions, smoking, taking photos, talking on the phone, and other everyday activities. To evaluate a model, the data is divided into three subsets, a training set, a validation set and a test set. The validation set serves an interim evaluation purpose which tests the performance of the model. Having this subset of the data is important because evaluating on the test set would bias the model (Jordan, 2017).

VNect (Mehta et al., 2017) is the first real-time method that captures the 3D pose of a human. The method is temporally stable (smooth and with minimal jitter) and uses a single RGB camera. It does not require the image to be cropped around the subject and fits a kinematic skeleton and not a 3D mesh like previous methods. Less processing is required by fitting a kinematic skeleton thus making the processing faster. The approach of this method was to be able to run it in an interactive gaming session, similar to the application of the Kinect gaming 3D camera. As the speed of this model is 30 frames-per-second (fps) it is on par and occasionally exceeds the results of the Kinect in tests. Additionally, this method can cope well with outdoor scenes, an area where the Kinect has detection problems due to sunlight interference. In comparison to the Kinect, this model is less bulky, has less power consumption, higher resolution and range and a single RBG camera is more widely available and not as expensive as the Kinect.

(Martinez et al., 2017) proposed a method of detecting three dimensional keypoint positions from 2D joint locations. The 2D detections are obtained by leveraging the work of (Newell et al., 2016). 3D point data was detected by lifting 2D points in a single image. Although there

[17] Leeds Sports Pose Dataset, https://dbcollection.readthedocs.io/en/latest/datasets/leeds_sports_pose_extended.html
[18] HumanEva-I Dataset, http://humaneva.is.tue.mpg.de/
[19] Human3.6M Dataset, http://vision.imar.ro/human3.6m/description.php

are numerous methods for lifting, the process is essentially obtaining 3D point data from 2D points in an image (Nie et al., 2021). The task of detecting points in 3D space from a 2D image can be difficult because of factors such a lighting, the background, clothing shape and image imperfections. The model is trained on paired datasets of images with 2D joint detections, and the corresponding 3D motion capture data of those poses. This learned "3D body concept" is then used to teach the 2D branch of the network how to produce 3D keypoints from 2D inputs, so when given a new 2D image, the model combines its 2D evidence with imagined 3D structure based on its body model. In this way, the network learns to "lift" 2D joint inputs into plausible 3D pose estimates, guided by its learned knowledge (Nie et al., 2021). This method considers the camera coordinate frame when detecting 3D keypoints and improves on previous 2D-to-3D keypoint prediction models.

Table 2.5 displays a chart of monocular 3D pose estimation models, presenting their output and the training dataset utilised.

Model	**Output**	**Training Dataset(s)**
SMPLify (Bogo et al., 2016)	3D mesh	Leeds Sports Pose, Human3.6M
VNect (Mehta et al., 2017)	3D keypoints/skeleton joints	MPI-INF- 3DHP, Human3.6M
A simple yet effective baseline for 3d human pose estimation (Martinez et al., 2017)	3D keypoints/skeleton joints	MPII[20], Human3.6M
Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision (Mehta, Rhodin, et al., 2018)	3D keypoints/skeleton joints	MPI-INF-3DHP
Learning to Estimate 3D Human Pose and Shape from a Single Colour Image (Pavlakos et al., 2018)	3D mesh	UP-3D, CMU

[20] MPII dataset, http://human-pose.mpi-inf.mpg.de/

End-to-end Recovery of Human Shape and Pose (Kanazawa et al., 2018)	3D mesh	Human3.6M, MPI- INF-3DHP
Single-shot multi-person 3D pose estimation from monocular RGB (Mehta, Sotnychenko, et al., 2018)	3D keypoints/skeleton joints	MPI-INF-3DHP, MuCo-3DHP
XNect (Mehta et al., 2020)	3D keypoints/skeleton joints	MS-COCO, MuCo-3DHP
VIBE (Kocabas et al., 2020)	3D mesh	AMASS

Table 2.5: Monocular 3D pose estimation models, output and training sets.

A new benchmark in the detection of 3D human pose in indoor and outdoor scenes was reached with the approach of (Mehta, Rhodin, et al., 2018), using a low cost RGB camera. They recognised that in-the-wild scenes are particularly difficult to detect and created their own 3D human pose dataset. This dataset named MPI-INF-3DHP[21], consists of subjects captured in indoor and outdoor scenes and has subjects wearing everyday clothing, having a large range of motions, and interacting with objects. Their methods of detecting 3D human pose consisted of three steps. Firstly, a bounding box around the subject is calculated and the 2D keypoints of the subject are positioned on the image. The second stage crops the image around the subject based on the bounding box data, and the third stage calculates the 3D human pose joint positions. The third stage works on a cropped image, so the estimation is relative to the crop. To transform it back to the full image, the known parameters of the crop undo the effect of the cropping, giving the correct orientation. Then the predicted 3D joints are aligned with the 2D joints from the previous step. Solving the alignment gives the 3D positions in x,y and z. This method obtains state-of-the-art results on the Human3.6M and HumanEva datasets.

Continuing the work from SMPLify (Bogo et al., 2016), Pavlakos et al. (2018) also considered estimating the 3D human keypoint locations and shape from an image. The SMPL (Loper et al., 2015) body shape model was incorporated, the SMPL mesh has 6890 vertices and is a high quality mesh that is generated by a limited number of parameters. The

[21] MPI-INF-3DHP Dataset, https://vcai.mpi-inf.mpg.de/3dhp-dataset/

model was evaluated on the UP-3D[22] (Lassner et al., 2017), SURREAL[23] (Varol, 2018) and Human3.6M (Ionescu et al., 2014) datasets. This approach improved on previous work of 3D keypoint detection from an image and the performance has a substantially faster running time to previous models. The efficiency of the model comes from simpler inputs, where the prediction of pose and shape comes from just 2D keypoints and silhouettes, making calculations faster as opposed to using raw images as input.

In an attempt to have a mesh representing human pose run in real-time, (Kanazawa et al., 2018) presented a model that outperforms previous models, provided that there is a bounding box calculated around the subject prior. Their model also contains data for part segmentation, where the head, limbs and torso are segmented from each other and also uses the SMPL human body model mentioned previously. This model learns the angle of joints of the human body for a more accurate representation, it also considers the orientation of the head and limbs. If any ambiguities occur when relationships between joints may seem unnatural, they are sent to a discriminator network where the joint angles are determined to be human or not. If they are not, they are sent back to be recalculated. Previous models infer a 3D mesh from 2D keypoints, whereas this model infers the mesh directly from the image. This model outperforms previous methods in terms of run-time and 3D joint error.

To solve the problem of estimating 3D pose of multiple people in a scene from a single camera, (Mehta, Sotnychenko, et al., 2018) proposed a model that inferred 3D positions of joints that have been occluded by other people in the scene. Previous multi-person 3D joint estimation models calculate 3D joint positions by combining single-person occurrences. This model calculates the 3D joint positions of all people in the scene jointly in a single pass, not requiring bounding boxes for separate people. The model excels on multi-person scenes where others fail, it also performs competitively on images that contain a single person.

XNect (Mehta et al., 2020) is a real-time multi-person 3D pose detection model, capable of calculating skeleton joint positions at speeds over 30 fps. Occlusions from subjects from the view of a single camera can be problematic in detecting pose, but XNect provides a fitted skeleton for each subject in the scene. Multi-person tracking requires more computational power to compute the joint position for all people and previous work detecting multiple

[22] UP-3D dataset, https://files.is.tuebingen.mpg.de/classner/up/
[23] SURREAL dataset, https://www.di.ens.fr/willow/research/surreal/data/

people reached 10-15 fps (Dabral et al., 2019; Rogez et al., 2019). An advantage of this model is that the pose detections can be used to drive other end applications requiring joint positions. Furthermore, only a single camera is required which results are of a similar quality as the Kinect depth camera.

VIBE (Video Inference for Human Body Pose and Shape Estimation) (Kocabas et al., 2020) is a markerless 3D pose detection system. It was developed at the Max Planck Institute for Intelligent Systems[24]. VIBE estimates the 3D shape and pose of the person in a video clip. VIBE exploits a dataset of 3D motions called AMASS[25] (Archive of Motion Capture as Surface Shapes) (Mahmood et al., 2019). The AMASS dataset includes 15 different optical marker-based motion capture libraries, 42 hours of data and 346 subjects. Included are datasets that have varied movements, such as HumanEva[26] (Sigal et al., 2010), CMU[27] (Joo et al., 2016) and the Pose Limits dataset (Akhter & Black, 2015), which include movement relevant to dancing and performing arts. VIBE is trained on human movement that is not limited to basic movement like walking or running, but also on irregular human poses. This model is well suited for the movement associated with performing arts because of its wide range of dancing motion in the dataset it is trained on. Figure 2.14 demonstrates how the VIBE pose detection model has improved compared to earlier methods, especially when dealing with challenging real-world video situations.

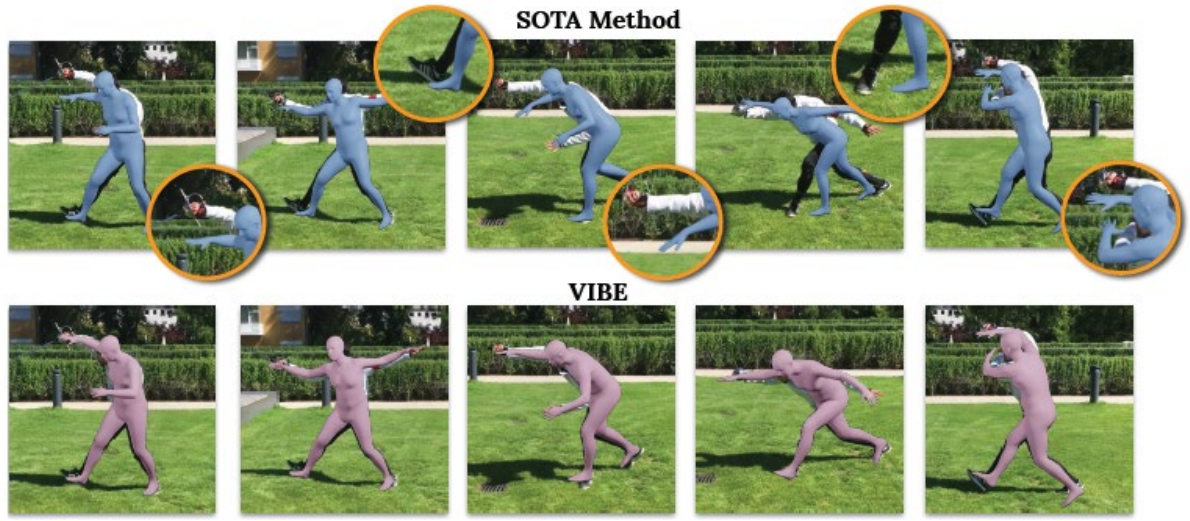

Fig. 2.14: VIBE (Kocabas et al., 2020).

[24] Max Planck Institute for Intelligent Systems, https://is.mpg.de/
[25] AMASS Dataset, https://amass.is.tue.mpg.de/
[26] HumanEva Dataset, http://humaneva.is.tue.mpg.de/
[27] CMU Dataset, http://domedb.perception.cs.cmu.edu/

There are online-based paid platforms where 3D human pose can be detected by submitting video. Ipisoft[28], Move.ai[29], Radical[30], Deepmotion[31], Kinetix Tech[32], and Wonder Dynamics[33] are paid web-based services which will provide an animated mesh from an input video from a single camera. The machine learning models used in this thesis, VIBE (Kocabas et al., 2020), EasyMocap (Shuai, 2021) and MPP2SOS (Barreto, n.d.), are open source and available for free on the Github code repository for anyone to use.

2.3.4.3 Multi-view 3D pose estimation models

Multi-view 3D human pose detection models have the benefit of viewing the subject at different angles, thus reducing the number of subject occlusions in the scene. It also makes estimation of where a point is in 3D space easier, once a point has been identified across all images. If one view is occluded, then the pose data can be obtained from another view if there is a direct line of sight to the subject. (Pavlakos et al., 2017) developed an approach which processed each view separately, producing heatmaps for each joint on a 2D image for each camera view. The heatmaps provide probability of where the joints may be. Once the heatmaps are calculated for all available views, they are combined to determine where the joints of the human pose are in 3D space. From this stage, 3D annotations can be created, creating a skeleton with joints in 3D space. This method attained state-of-the-art results on common benchmarks for multi-view pose detection models.

EpipolarPose (Kocabas et al., 2019) is a multi-camera 3D pose estimation model that predicts the 3D joints of human pose by using at least two RGB cameras. This model sets state-of-the-art results on the Human3.6M and MPI-INF-3DHP datasets. EpipolarPose is trained on the MPII Human Pose dataset[34] (MPII) (Andriluka et al., 2014). When compared to previous datasets, MPII contains a wide variety of human motion and from varying viewpoints. For evaluation, the Human3.6M dataset was used, which still remains one of the largest datasets for 3D human pose estimation.

In 2020 and 2021, the most accurate multi-camera 3D pose estimation listed on the prominent paperswithcode.com performance leader board (when ranked on the H3.6M

[28] IPi Soft- Markerless Motion Capture, https://www.ipisoft.com/
[29] Move Ai, https://www.move.ai/
[30] Radical – AI-Powered 3D Animation, https://radicalmotion.com/
[31] Deepmotion - AI Motion Capture & Body Tracking , https://www.deepmotion.com/
[32] Kinetix Tech, https://www.kinetix.tech/
[33] Wonder Dynamics , https://wonderdynamics.com/
[34] MPII Human Pose dataset, http://human-pose.mpi-inf.mpg.de/

dataset benchmark) was (Iskakov et al., 2019). This model took about a week to train on the H3.6M dataset. The accuracy of the model is increased by the number of cameras used for the pose estimation, up until about 14 cameras where the change in error rate is negligible.

EasyMocap (Dong et al., 2021) is 3D human pose estimation model that fits a SMPL mesh to calculated joints in 3D space. Multiple calibrated RGB cameras are used to capture the subject. The footage of the subject is synced and the distortion of the lens is removed for a more accurate detection. The model requires the subject to be in the frame of all views for accurate detection, if the subject leaves the detection area errors will increase and glitching might occur. The camera calibrations stage allows the model to know where each camera is in 3D space in relation to the subject. Once processed, a 3D mesh of the subject is produced in the form of a filmbox (.fbx) file which can be used in other 3D applications. Furthermore, a skeleton of the joints is provided which is parented to the 3D mesh, this allows further editing and smoothing where needed. Parenting allows the mesh to be paired to the skeleton, so that when the skeleton joints are animated, the mesh follows the same motion. The ability to edit the mesh post-detection is attractive to practitioners who wish to further art direct or smooth the motion in some way.

TesseTrack (Reddy et al., 2021) is a 3D multi-person multi-view 3D pose estimation model, ranked at first place on the H3.6M dataset for monocular and multi-views on paperswithcode.com at the time of writing. For multiple view its MPJPE (Mean Per Joint Position Error) average is 18.7mm, whereas its closest rival was AdaFuse (Zhang et al., 2021), with an average MPJPE of 13.55mm. Tessetrack identifies each subject in the scene, identifies each person's body joints in a skeletal structure and tracks the joints over time. This model is well adapted to real-world scenarios of multiple-people interacting in close proximity who are outdoors where fast motion and occlusions are considered. The authors recognise that solving 3D joint positions using multiple stages can be prone to errors that cannot be recovered. Thus, their approach of simultaneously addressing these stages improves the result. An improvement in accuracy in MPJPE is reported on the H3.6M, Campus and Panoptic datasets.

A recent open-source library fuses three existing models to generate a skeleton from multiple views called MPP2SOS. The models are Blazepose (Bazarevsky et al., 2020; *MediaPipe | Google Developers*, n.d.-a), pose2sim (Pagnon et al., 2022b) and opensim (*SimTK: OpenSim: Project Home*, n.d.). Blazepose generates the 2D poses for all cameras, pose2sim triangulates the cameras and converts the data to be read by opensim. Opensim acts as a solver for keypoints in 3D space and allows easy import of the data into the

Blender 3D application. There is no associated academic paper for MPP2SOS, but the outcome of the model produces an articulated 3D mesh based on multiple views and includes a skeletal structure parented to joints for further editing or smoothing. MPP2SOS requires camera intrinsic and extrinsic data in order to triangulate the camera positions and remove lens distortion. Views of two or more angles are required and the video footage needs to be synced. The subject is detected in the image and bounding box data is used to detect joints in 2D in all views. 3D joint positions are then detected from 2D, where joints and bones (skeleton) are connected to the keypoints are output. The skeleton is in .fbx format which can be imported in 3D applications for further processing.

2.3.4.4 Machine learning in artistic performances

If choreographers would like to explore the possibilities of using a machine learning model to capture the motion of a dancer, there are many to choose from including open-source code to run them. Examples of where choreographers use machine learning models directly for motion capture for projecting animation in a performance are sparsely documented, especially in academic literature. There is, however, evidence of where machine learning has been used in choreography and dance performance.

John McCormick and his team set up a machine learning model to learn the movement of a dancer and, when recognised, respond to that movement from learned choreography (Mccormick et al., 2013). The dancer recorded and supplied material for the machine learning agent to learn from, incorporating spontaneous dance movements. After observing the agent's learning process, the dancer improvises new movements inspired by and building upon the refined motions generated by the agent's responses. This refined motion and interactivity between the dancer and agent become the vocabulary for the production between the dancer and agent. The motion of the agent is projected stereoscopically onto a 3D environment as the dancer performs. The method of digitising the dancer’s motion was with the use of an optical-based motion capture system which recorded forty reflective markers on the dancer’s suit for training. It was demonstrated that a machine learning agent that can learn dance choreography from a dancer can be integrated into the collaborative creative process.

Brenton et al (2014) used Interactive Machine Learning (IML) to create interactive visuals that responded to the improvised movements of a non-professional dancer (Brenton et al., 2014). Interactive machine learning is a process where a user (or group of users) gradually

improves a computer model through repeated rounds of providing input and feedback. The user builds and refines the mathematical model that the computer uses to understand a particular concept or process. This happens through an iterative cycle - the user gives the computer some input data or guidance, and then reviews how well the computer model has learned from that input. The way the user provides input to refine the model can take many forms - they may give examples that represent the concept, describe key features the model should look for, or adjust high-level settings of the model. Through this iterative loop of the user inputting data/guidance and then reviewing/correcting the model's current state, the model gradually gets refined and improved to accurately capture the concept the user has in mind. In essence, it's an interactive process of the human user teaching the computer model through repeated cycles of input and correction, until the model's understanding matches what the user intends (Dudley & Kristensson, 2018).

The visualizations are generated through the subject's dance movements and have provided the researchers with insights into interactive interfaces, specifically in the context of spontaneous and unique dance expressions. The work made embodied interaction more accessible using cheaper sensing equipment than previous methods. The parallels that the authors reference are those in which a game player would use a Microsoft Kinect or Nintendo Wii to interact with a game when performing free-form dance. Two prototypes were created prior, the first used a Microsoft Kinect to serve as an input medium to control visual elements. Feedback from users of this system reported a disconnect between their motion and the visualisations on screen, in regards to scale and rotation. Another mentioned the latency between their motion and the visualisations which was disconcerting. The system was designed to recognise larger motions, so if a subject’s dancing style had smaller motions, the response is not as evident. The second prototype utilised IML considering the limitations of hardwired mapping from the Kinect sensor. The IML method allowed the subject to train a dataset without them having knowledge of technical coding. Although an improvement on the first prototype, the performance was underwhelming as some actions were not identified. The final prototype was created and utilised the Oculus Rift, a virtual reality headset, which displayed the visualisation in three dimensions to make the experience more immersive for the subject.

Chen et al. (2017) introduced Adversarial PoseNet, an approach to human pose estimation that focused on making pose predictions more natural and human-like. Traditional deep learning systems often struggled when parts of the body were hidden or when multiple people overlapped in images, sometimes producing unrealistic body poses. Taking

inspiration from how humans can naturally understand body positions even with partial information, Adversarial PoseNet was designed to learn the natural constraints of how human bodies can move and be positioned. The system used a neural network (GAN) that essentially learned to tell the difference between realistic and unrealistic poses. This approach proved particularly effective when dealing with partially hidden body parts, demonstrating significant improvements over previous methods. The success of Adversarial PoseNet showed the importance of teaching AI systems about the basic rules of human body structure when performing pose estimation.

Google developed TensorFlow as a powerful tool to help researchers and developers work with machine learning more effectively. According to Abadi et al. (2016), TensorFlow was built to be flexible - it can run on a single laptop or scale up to work across many computers in large data centres. What makes TensorFlow special is how it organises machine learning calculations in a way that makes them easy to understand and modify. This design allows researchers to try new ideas and approaches without having to rebuild everything from scratch. The tool has become extremely popular in the machine learning community, with thousands of researchers using it and over a million downloads of its software.

TensorFlow itself is not a machine learning model - instead, it's more like a toolbox that helps people build and run their own machine learning models. As Abadi et al. explain, TensorFlow provides the building blocks and tools that researchers and developers need to create different types of machine learning systems. TensorFlow provides the basic pieces and instructions, but researchers use these pieces to build their own unique machine learning solutions (Abadi et al., 2016).

As a source of inspiration for dance choreography, the machine learning model chor-rnn was developed as a collaborative tool which interacts between human and machine (Crnkovic-friis & Crnkovic-friis, 2016). Their goal was to find if a computer could create relevant choreographies. Previous literature mentioned that choreography has three levels of abstraction; style (a dancers expression), syntax (the language of the work) and semantics (the overarching theme that combines the work coherently) (Blacking & Keali'inohomoku, 1980). They also argue that dance lacks a comprehensive notation system like music. Systems like Benesh (Neagle et al., 2004) and Laban (Barbacci, 2002) notation do not capture style and are not widely adopted because they are difficult to learn (Guest & Ryman, 1998). Motion capture systems contain far more information of the performer by capturing the movement of their joints in 3D space. The authors used the Microsoft Kinect V2 sensor to detect 25 joint positions at a speed of 30 fps. Five hours of a dancer's movement was

collected for training data. After six hours of training the system learned the basic motion of the dancer from the training data and after 48 hrs the results show that the chor-rnn system can create new dance choreography in the style of the training data and understands joint relationships to each other. The work determined that the chor-rnn system can be used as a creative tool to enhance the process of creating choreography in a particular style of dance.

Dance Dance Revolution (DDR) (Höysniemi, 2006) is a popular video game where players can dance to popular songs while triggering switches on a dance platform on the floor when dancing. Dance Dance Convolution (Donahue et al., 2017) is a machine learning-based model which improves on the game based on feedback from players. Some negative feedback about DDR is that players do not have access to a wide variety of songs to dance to, step charts are similar in different songs and although step charts can be customised, the process is arduous and requires considerable know-how. The authors automate the creation of step charts so that players can dance to a more varied sequence of steps to any song they choose. Step charts are generated from raw audio and based on difficulty level. Beginner charts are restricted to slower motion and contain quarter and eighth notes. More challenging levels contain faster motion going up to 16^{th} and 32^{nd} notes and may include triplet patterns. Step placement is identified by timestamps and step selection involves selecting which steps occurs at each timestep. The machine learning model predicts step placement that are aligned with significant audio features in the song, considers difficulty level and constructs a step chart for four switches that the player will hit during the song. Each switch has four states which are on, off, hold and release.

Carlson, Schiphorst, et al. (2015) developed iDanceForms, a mobile application that allows choreographers to capture dance movements through camera still frames and create movement sequences. The system uses a camera to photograph dancers' poses, matches these against a database of skeletal data, and allows choreographers to sequence and manipulate these captured movements. Through a study with professional choreographers, they found that the mobile platform enabled new creative possibilities by allowing artists to work directly in the studio space while capturing and manipulating movement material. Notably, while the system was designed for straightforward pose capture, choreographers discovered creative ways to use it - photographing from unusual angles, capturing architectural elements instead of bodies, and experimenting with multiple subjects simultaneously. This unexpected creative appropriation demonstrates how dancers and choreographers can adapt motion capture tools beyond their intended technical purposes to serve artistic exploration.

Carlson et al. (2014) subsequently developed Cochoreo, a sub-module within the idanceForms (Carlson, Tsang, et al., 2015) platform that uses genetic algorithms to generate novel keyframe animations for choreography. The system was designed to bridge the gap between creativity support tools and autonomous creative systems, allowing choreographers to interactively develop movement material. Cochoreo generates body positions using a fitness function based on Laban Movement Analysis parameters, enabling choreographers to adjust generation options based on their preferences. The system was evaluated in a pilot study with fourteen novice choreographers, who found that the generated movements suggested options they would not have developed themselves. Interestingly, the researchers found that when the system produced lower resolution movement data, it prompted more creative interpretation from the choreographers, while higher resolution data led them to focus more on precise physical mapping from screen to body. This finding highlights an important consideration in machine learning approaches to dance - that sometimes less precise or "incomplete" computational outputs can stimulate more creative responses from dancers. Like other systems of its time, Cochoreo demonstrated early potential for machine learning to support creative movement generation, though it was limited to 2D data and required manual intervention to create full movement sequences.

Choreographer Wayne McGregor partnered with Google Arts and Culture to use machine learning to generate new movement material (Bastien Girschig, 2019a). McGregor handed over his entire catalogue of dance movement, which consisted of thousands of hours of material, so that the machine learning model could learn the language of his dancing. In a typical rehearsal scenario McGregor would start a dance phrase and iterate on it to get something that works. With this model, however, hundreds of iterations of that phrase could be generated to speed up the creative process and to have many more options to choose from. The model can replicate dance movement in the style of McGregor and allow for many iterations to choose from.

Kyle McDonald, an artist and coder exploring machine learning models in the context of dance, worked with fellow artist Daito Manabe and choreographer Mikiko to investigate training a machine learning model to create dances in the style of one of Mikiko’s dancers, Maruyama Masako (McDonald, 2019). In addition to this, dance sequences were generated from videos captured of the audience prior to the performance. Masako performed on stage with a digital representation of herself performing next to her, closely matching her movements. In Figure 2.15, a digital portrayal of the dancer from the production *Discrete Figures* is presented.

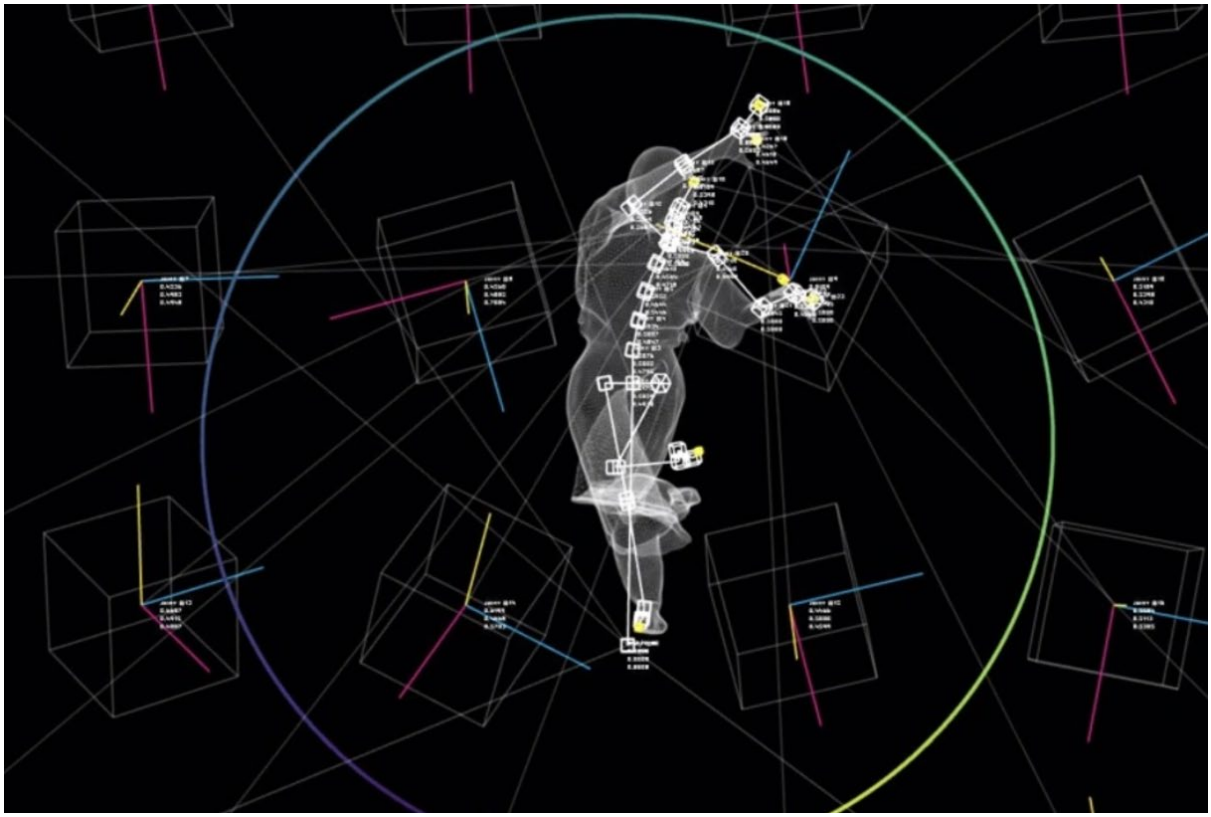

Fig. 2.15: *Discrete Figures* (McDonald, 2019)

Before the performance started, video footage of audience members was recorded and detection was run using OpenPose (Cao et al., 2018). The rendering of each person to an image was done by pix2pixHD (T. C. Wang et al., 2018). The rendered video is checked fifteen minutes before the performance starts. The outcome is very similar to the AI dance project *Everybody Dance Now* (Chan et al., 2019) where motion capture data is transferred onto a textured mesh.

Pettee et al. (2019) developed machine learning tools to assist choreographers to output a potentially infinite number of novel dance sequences. By adding controlled amounts of noise corresponding to an input sequence, allows potentially infinite subtle to significant variations, while maintain some connection to the original input phrase (Pettee et al., 2019). Their model also has the ability to fine-tune the model to create subtle or widely varying variations. This application deploys a machine learning model to enhance choreographic innovation by enabling choreographers to explore their embodied knowledge, resulting in a high array of movement possibilities. They believe this research builds upon the current dance notation, but also raises questions where using such a model might lead choreographers to envy the algorithm and doubt their own created choreographies. Furthermore, the authors brought up

a good point in that by using tools like these may lead to the analysis of what well-constructed choreography means and may deny humans the space to be creative.

Nao Tokui documented his work using machine learning methods in collaboration with flamenco dancer Israel Galvan (Tokui, 2020). Galvan wanted to work on a project where he could “dance with his alter ego”. Custom flamenco shoes were constructed with sensors on the heel, toe and middle of the foot. The sensors would record each step, up to ten steps per second in parts. The shoes transmitted step data wirelessly during the performance. The data was collected and the machine learning model learned the rhythm of Galvan’s steps, which predicted a new sequence of steps from input data. In order to respond to Galvan’s dancing, small machines were constructed with solenoids that hit the floor. In Figure 2.16, we observe the presence of custom flamenco shoes and solenoids worn on stage by the performer Israel Galvan.

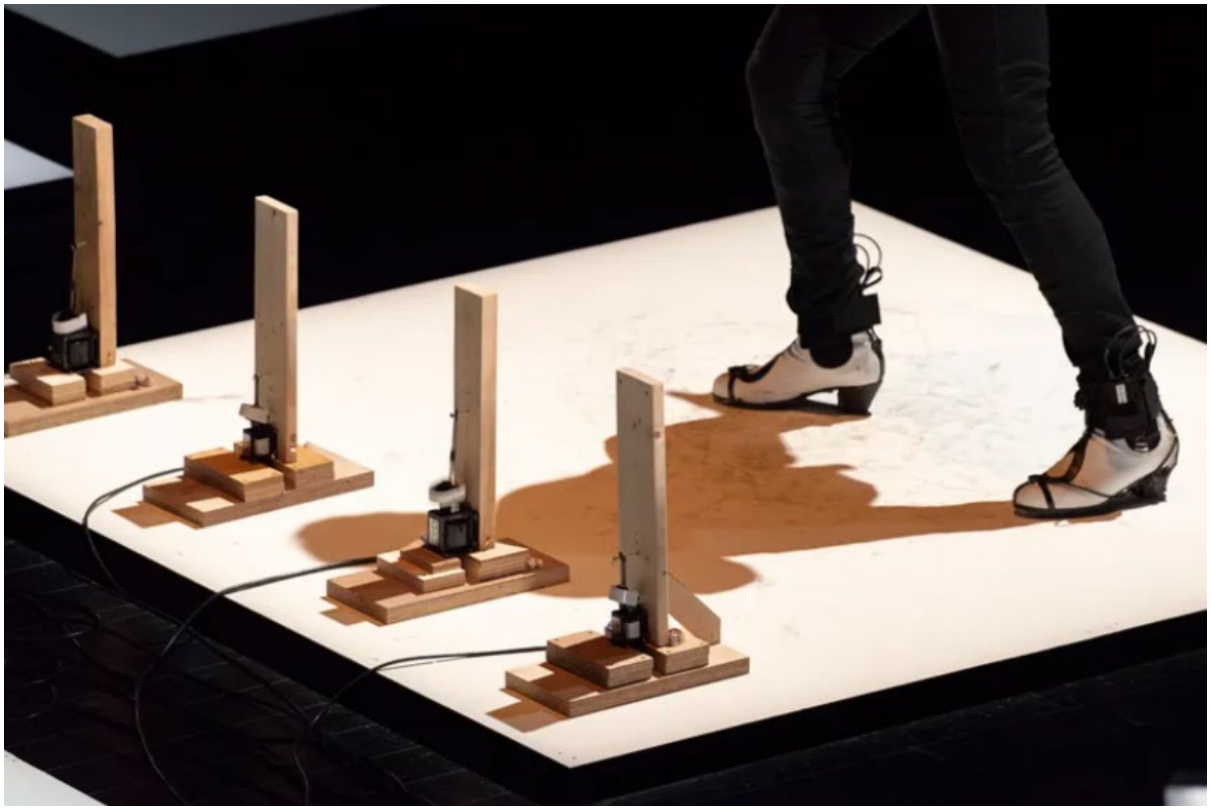

Fig. 2.16: Performer Israel Galvan (Tokui, 2020).

After training the model the feedback Tokui got from Galvan was that the response was “too flamenco-y”. Galvan’s style of dancing was not true flamenco but had a contemporary feel to it. After some adjustment of adding some randomness and fine tuning the model, the performance was ready. After the performance Galvan commented that the response from

the machine learning model was inspiring and that he sometimes forgot that he was dancing alone.

Choi et al. (2021) noted that for people often learn dance moves from their K-Pop (Korean popular music) idols. They developed a smartphone-based application which assists the user to learn a particular dance. Machine learning is used to obtain 2D joint data of the student learning the dance moves from the application. It then evaluates joint position and angular similarities between the student and the keypoints of the application. A rule-based system is then used to compare the difference. The application is geared towards self-learning, where the user prefers to learn choreography on their own instead of going into a dance class or having group lessons and is more convenient than learning from online resources such as YouTube. Previous methods have attempted to solve this problem but have not been broadly adopted because 3D cameras are required, which are relatively expensive. The authors improved on previous work where occluded limbs were not detected and the rotation of the torso was erroneous. They found that this affordable method of learning dance moves provided accurate criticism to people learning to dance, with the convenience of having the application in a smartphone.

Ephraim Wegner and Daniel Bisig collaborated with dancer Emi Miyoshi to create the project Artificial Intimacy (Miyoshi, 2021). Miyoshi interacts with an avatar previously trained on the movements of several other dancers. She has seven sensors fitted into her shoes which are pressure sensitive as well as IMU sensors around her body measuring acceleration and orientation. The sensors transmit data wirelessly through Wi-Fi. Miyoshi is also recorded with RGB and depth cameras in the form of the Microsoft Kinect V2. The avatar that Myoshi performs with is presented by a stick figure. The data that is recorded from her shoes and the Kinect is used to train a motion to music translation model. The model that is trained with motion capture data is called Granular Dance (Bisig, 2021a) and the model that creates audio from motion capture data is called RAMFEM (Raw music from free movements) (Bisig, 2021b). Granular Dance uses a neural network to compress motion capture data of dance poses in smaller representations. The model can then generate new dance sequences by randomising these smaller representations and turning them into dance pose sequences. The audio component takes a sequence of captured poses and outputs phrases of audio which are blended together. Myoshi experimented with the interaction of the avatar and the audio that was produced. She tried mimicking its motions, and also tried to get alternative results by changing the orientation of the IMU's that recorded her movement. The conversation that Myoshi has with the feedback of the audio ranges from an immediate response to waiting a moment and then responding to the audio that she hears. The method

of designing the production was open-ended where initially there was a lot of experimentation between the technology and the dancer. This feedback, along with regular communication open up more questions about what the aesthetic should look and sound like. Myoshi described this phase as the “searching” phase in the production and once the production deadline came near, the knowledge that was learned was consolidated into a coherent piece.

DanceON (Payne et al., 2021) is an interactive application which allows users to create on-screen graphics as they dance. The system educates users creatively and technically by combining an engaging way to include simplified coding and dance. The application consists of a media player with built-in pose detection algorithms which allows the user to connect graphical shapes to body positions with a few lines of code. Users can upload dance videos and create animations based on pose detection and pixel data. The authors recognised an under-representation of women of colour in STEM (Science, Technology, Engineering and Mathematics) fields, and presented the DanceON application to young women on a two-week summer camp. The interactive animation that the users created are presented as cultural themes and relevant issues, such as “Dripping with Power”, “Black Lives Matter”, “Colorful Escape” and “Love Yourself”. The goal was to use this educationally-focussed interactive system to engage users by creating artistic animation based on body movement into relevant themes while further manipulating a graphical interface with minimal coding. The entry-level coding element is designed to allow users to get a glimpse of the interactivity of coding.

The Australian production *Forgery*[35] utilised machine learning with dance in a novel way by giving instructions to dancers to perform live on stage. Every show was unique in that the instructions were different each time. The instructions would tell the dancers where to move, and provide a piece of music to perform to. The model is a text generator which is converted into the voice of the model which then articulates the instructions. On some occasions the audience could hear the voice of the model uttering instructions, other times the text of the instruction is projected, but sometimes the audience does not know what the instructions are. The creator of the model mentions that there are “literally trillions” of possible outcomes of instructions from the model.

[35] Forgery – Australasian Dance Collective Dance Against The Machine, https://scenestr.com.au/arts/forgery-australasian-dance-collective-dance-against-the-machine-20210819

Bisig and Wegner improved on their machine learning model Granular Dance by animating a machine-learning-based virtual puppet (Bisig & Wegner, 2021). Their work, inspired by puppeteering, has the input of a single limb that is animated either manually or by an algorithm. The joints that are not controlled by the algorithm move in a physical way, similar to the joints of a puppet when controlled by a puppeteer. The machine learning model searches for poses that are similar to the training data as well as suitable to the controlled limb. Thus, the puppet is being controlled mostly by the model which makes pose decisions independently. A conversation is created between the puppeteer and the puppet, where a limb is moved and the rest of the body responds, where both are improvising and the puppet retains some of the creative nuances of the original dance from where the model was trained. Although not directly controlled by motion capture, the machine learning model interpreted dance poses from its training, from a dancer whose movement was captured by motion capture.

Alemi et al. (2015) explored a different approach to movement generation using Factored Conditional Restricted Boltzmann Machines (FCRBMs). They developed a system that could create new dance movements while controlling their emotional qualities. They recorded two professional actors performing various movements while expressing different emotions, ranging from happy to sad, and energetic to calm. Using machine learning, their system learned to generate new movements that could express these emotional states and even smoothly transition between them. When they tested their system with viewers, people could successfully recognize the energy levels (high vs. low) in the generated movements, though they found it harder to identify emotional qualities like happiness or sadness. This study showed both the potential and limitations of using machine learning to create emotionally expressive movements - while it worked well for capturing energy and intensity in movement, capturing more subtle emotional qualities remained challenging.

For creative applications like dance performance, it's important that computer systems can learn and understand new movements as they happen. (Liu et al., 2019) developed a system that could automatically group similar movements together without needing to be pre-programmed with specific movement categories. Their system works in three steps: first, it simplifies the movement by keeping only the most important frames; second, it identifies key joint angles; and finally, it uses machine learning techniques to group similar movements together. When tested with 104 different movements captured using a Kinect sensor, the system could successfully separate different types of movements - for example, grouping hand gestures together and keeping them separate from leg movements. While the system worked well with simple movements, it struggled with complex movements that used multiple

body parts at once. It also couldn't recognize that the same movement done with the left or right arm should be considered similar. Despite these limitations, their work shows how motion capture systems can move beyond just recognizing pre-defined movements to learning new movements through interaction with performers.

Trajkova et al. (2024) conducted focus groups with dance students to understand how collaborative movement improvisation works between dancers, with the goal of improving AI (artificial intelligence) dance partners. Through their research, they developed a model showing how dancers make movement choices based on immediate influences (from themselves, their partner, and environment), movement strategies, and collaborative guidelines. Their findings are being used to improve LuminAI, an interactive AI system that can improvise dance movements with humans. The system projects an AI avatar onto a screen that can detect and respond to human dancers' movements in real-time. This research demonstrates how understanding human dance improvisation can help create more natural and collaborative AI dance partners.

A comprehensive systematic review by(Nogueira et al., 2024)examined how machine learning techniques impact dance performance across four key areas: choreographic creation support, dataset collection and training, human pose detection techniques, and new visual representations of dance movement. The review analysed twenty-three papers published between 2010-2023 and found that while machine learning offers significant benefits for enhancing creativity and improving pose detection accuracy, there are notable challenges around limited engagement with dance professionals and questions of artistic ownership. The authors identified that many studies developing pose detection systems did not involve professional dancers in the data collection process, potentially leading to datasets that don't fully capture the technical nuances of specific dance styles. They also highlighted ethical concerns around attributing credit for machine-generated choreography and maintaining cultural sensitivity when algorithms are used to generate dance movements. This thorough analysis provides valuable context for understanding both the opportunities and limitations of applying machine learning to dance performance.

Gunerli et al. (2024) developed a novel video processing pipeline specifically designed for analysing dance movements through computer vision. Their system combines pre-trained MediaPipe (*MediaPipe | Google Developers*, n.d.) models with human input to create a four-stage process: interactive image segmentation, video batching, keyframe extraction, and precise timestamping. A key advantage of their approach is that it eliminates the need for motion capture suits or sensors, allowing dancers to move naturally without physical

interference. The pipeline specifically addresses challenges in capturing complex dance movements by using multiple specialized algorithms working together, while also incorporating human feedback at crucial stages. The system is designed as part of LuminAI, an interactive installation where an AI agent can improvise movements with human dancers. Their work demonstrates how computer vision techniques can be effectively combined with human expertise to create more accurate and practical motion capture solutions for dance applications, though it is currently limited to analysing only one dancer at a time.

2.4 Conclusion

This literature review covers the diverse facets of motion capture ranging from its origins to current applications. The early history traced pioneering works like the zoopraxiscope, chronophotography, and rotoscoping which planted the early seeds for transferring motion from a physical medium to digital representation. The advent of computer technology enabled the emergence of digital motion capture systems to flourish in fields like film and animation. Optical marker-based systems such as Vicon, deliver high accuracy and precision but face limitations in terms of cost, portability and restrictiveness due to their suits, line-of-sight dependency and its need for precise calibration. Alternative approaches based on depth cameras like the Microsoft Kinect or sensor-based motion capture alleviate some of these constraints providing more accessible and flexible motion capture, but are also constrained by line-of-sight dependency, limited range, sensitivity to lighting conditions, limited capture of fine details such as face and hands, and also the potential of losing accuracy due to unwanted subtle drift of sensors over time.

Recent developments in machine learning-based motion capture models introduce new prospects for motion capture. Monocular and multi-camera models can infer 3D skeletal motion without markers or sensors, using only input from one or more RGB cameras. Performance benchmarks illustrate the rapid improvement of such models on public datasets. Though constraints around occlusion and subjects performing unexpected movements that the model is not trained on remain, the low cost and convenience of such markerless approaches has prompted novel explorations in dance and performance contexts. Artists are able to harness these models in collaborative performances, where motion capture data and interactive visuals shaped by the model occur.

Very little research has investigated how machine learning-based motion capture could benefit choreographers and dancers directly in the creative process. Most machine learning applications focus on the technical aspects of human pose estimation, while the literature lacks examples of machine learning-based motion capture enabling new forms of animated expression for live performance. This represents a significant opportunity for further research. There is untapped potential for performers to leverage machine learning-based motion capture to push the boundaries of their practice. The research now needs to explore how choreographers can harness these tools to generate innovative styles of animated motion and integrate that digital animation into dance and theatre productions. By focussing machine learning on the creative needs of performers themselves, exciting new directions for performing arts integrating animation and machine intelligence could emerge.

3 Research Methodology

3.1 Introduction

This research investigates how machine learning-based motion capture can be effectively integrated into performing arts practices through a mixed-methods approach combining practice-based research with qualitative investigation. The study is guided by two key research questions:

RQ1: 'In performing arts, what are the characteristics creative practitioners look for in motion capture systems?'

RQ2: 'What are the benefits, limits, and implications of current machine learning-based motion capture systems in the performing arts space?'

To address these questions, the research design incorporates the following methodological strands. First, a qualitative investigation using semi-structured interviews with experienced performers, choreographers and artists provides insights into current uses and needs around motion capture technology, perceived benefits and limitations of existing systems, and potential opportunities for machine learning-based approaches. The interview data is analysed using thematic analysis to identify patterns and insights relevant to RQ1.

Second, through practice-based research involving the development and presentation of a performance piece incorporating machine learning-based motion capture, the study examines practical implementation challenges and opportunities. This involves testing and evaluation of different machine learning models, collaborative development with performers, and analysis of public performance outcomes and audience response. This practical component directly addresses RQ2 by demonstrating and evaluating the technology in a real performance context.

The research methodology is deliberately structured to gather multiple perspectives on the research questions. The interviews capture practitioner insights and requirements, while the practice-based component offers concrete evidence of benefits and limitations through actual implementation. This dual approach enables a comprehensive understanding of both the technical and artistic implications of integrating machine learning-based motion capture into performing arts practice.

Table 2.6 below represents the different forms of data collection and their relationships.

Data Collection Method	Source	Purpose	Research Question Relevance
Practitioner Interviews	Motion capture practitioners, performers, choreographers	Understand practitioners needs, practices	RQ1: Understanding characteristics practitioners need in motion capture systems
Performance Observations	Performer rehearsals and live performance	Document interactions with technology, creative processes, technical challenges	RQ2: Understanding practical benefits and limitations of machine learning-based systems
Performer Interviews	Performer (Cloé)	Gather insights on experience with machine learning-based motion capture compared to traditional methods	RQ2: Understanding implications of machine learning-based systems from a performer perspective
Audience Interviews	Performance attendees	Gather feedback on perception of technology integration and performance impact	RQ2: Understanding effectiveness and implications of machine learning-based systems in live performance
Self-reflection	Researcher observations and documentation	Document processes, challenges, and insights during development and implementation	RQ1 and RQ2: Overall understanding of system capabilities and implications

Table. 2.6 Data collection types and relationships.

This chapter details the methodology used to achieve the objectives of this research. The research can be categorized as follows:

1) Literature review (Chapter 2): Search the literature, including existing artworks, to identify types of technology, trends and methods used for motion capture in performance works. The literature review outlines previous research related to the topic, tracing key developments over time. It highlights open questions and limitations, while identifying gaps in the literature.

2) Thematic analysis:
    a. Collect data through interviews with experienced performers, choreographers and artists to establish their needs and current uses of motion capture in performing arts.
    b. Analyse the data to find where machine learning-based motion capture methods can fit in with their workflow and identify any benefits, improvements or shortcomings.
3) Machine learning-based motion capture pipeline design and testing, including the production of a performance piece.
    a. With the needs of choreographers and dancers needs in mind, identify suitable machine learning-based motion capture models.
    b. Evaluate the process of deploying those models and the performance and behaviour of those models in practice.
    c. Apply animation to the motion-captured animated mesh of the performer, and present it in a performance setting.

3.2 Practice-based research

3.2.1 Introduction

Candy defines ‘creative practice’ as the novel creation of an artefact (or artwork) that uses the essential processes and techniques inherent in the domains of art, music, design, engineering or science (Candy & Edmonds, 2011). The practice-based component of this research is the creation of a performance piece that demonstrates the use of machine learning methods to create animation. It also demonstrates the collaborative experience of the performance which is referring to the experience of the performer while developing the work, audience response to the performance, and feedback from the performer. Modern machine learning-based pose detection models will be used to analyse and capture human

movement and produce an animated skeleton matching the performer's motion. A 3D mesh will then be combined with the skeleton, and abstract animation applied to accompany a dancer in a performance. The entire pipeline, from the acquisition of the motion capture data from the performer including their choreography, through to the performance, is included in this practice-based research. In this context, the significance of practice-based research lies in its ability to bridge the gap between theoretical understanding and practical implementation. By creating animation to accompany performances, this research contributes to the body of knowledge by demonstrating the potential of machine learning in generating abstract animation. Additionally, it examines the collaborative performance, audience experience, interdisciplinary collaboration between technology and the performing arts, and explores innovative ways to integrate emerging technologies into live performances. This research aims to understand the potential for machine learning-based motion capture in creative practices. By merging this technology with established performing arts practices, innovative and transformative works may emerge, firmly situating this research within Candy's definition of 'creative practice'.

Practice-based research is deeply rooted in the practice and expertise of the researcher (Candy & Edmonds, 2011). Candy and Edmonds clarify that although 'practice-based' and 'practice-led' are used interchangeably, they denote different focuses. Practice-led research strives to uncover fresh understandings about practice. At the same time, practice-based primarily involves the creation of artefacts, thus being intrinsically linked to producing an intrinsic outcome or object during the research period. As such, the research presented in this thesis is practice-based, as the outcome is the creation of artefacts, and documentation of the process is included. This documentation includes my own reflections, as well as contributions from the audience, the performer, and industry practitioners.

This aligns with Fraleigh's (Fraleigh, 1999) understanding that dance research requires a recognition of 'things we do' as distinct objects of study, while acknowledging that in dance 'this is not a material thing that can be positioned apart from ourselves.' This perspective is particularly relevant to practice-based research where the researcher is deeply embedded in the creative process.

Frayling's (1993) work on research in art and design provides a foundational framework for understanding different approaches to practice-based research. He identifies three distinct categories: research into art and design, which encompasses historical, aesthetic, and theoretical investigations; research through art and design, which includes materials

research, development work, and action research; and research for art and design, where the end product is an artifact that embodies the research thinking. Frayling's framework emphasises documentation and communication of process – particularly in research through art and design, where he argues that the research diary and contextual report are essential elements that distinguish research from mere practice. This documentation transforms creative practice into research by making explicit the knowledge generated through the creative process. Frayling argues that while the artifact itself may embody knowledge, it is the documented investigation and communication of the research process that validates practice-based research in academic contexts (Frayling, 1993).

This research aligns with Frayling's influential categorisation of art and design research, particularly his concept of 'research through art and design' where practice serves as an investigative methodology. Frayling mentions that such research must involve systematic documentation and communication of the creative process and its outcomes. In this study, the development and presentation of the performance piece exemplifies research through practice, as it involves 'action research - where a research diary tells, in a step-by-step way, of a practical experiment in the studios, and the resulting report aims to contextualise it' (Frayling, 1993). The research also incorporates elements of what Frayling terms 'research into art and design' through its theoretical and historical investigation of motion capture technologies, and 'research for art and design' in the creation of the final performance artifact. These approaches allow for an investigation of machine learning-based motion capture's potential in performing arts contexts.

This research aligns with Leavy's (2020) framework of arts-based research practices, which emphasizes how art and science can work together in research endeavours. Leavy argues that arts-based practices are particularly valuable for research projects that aim to 'describe, explore, discover, or unsettle' - goals that align directly with this study's investigation of machine learning-based motion capture in performing arts. Furthermore, these practices are especially suited to projects that are 'problem-centered' and transdisciplinary, as is the case with this research which bridges technology and performing arts.

Leavy emphasizes that arts-based research practices can be employed during any phase of research, including data generation, analysis, interpretation, and representation. This study incorporates arts-based practices across multiple phases: in data generation through the creation of performance pieces, in analysis through reflection-in-action during rehearsals and development, and in representation through the final performance piece. Additionally, Leavy notes that arts-based approaches are particularly effective at forging micro-macro

connections - in this case, connecting individual performance experiences with broader implications for the field of performing arts technology.

The holistic nature of arts-based research, as described by Leavy, is reflected in this study's integrated approach to research design, in which theory and practice are intertwined throughout the process. This is evidenced in how the practice-based components (performance development, technological implementation) inform and are informed by the theoretical framework and interview findings (Leavy, 2020).

The data collection approach in this research aligns with Creswell's (2007) framework for qualitative inquiry, which emphasizes that qualitative researchers collect data in natural settings with sensitivity to the people under study, and analyse their data inductively to establish patterns or themes. As Creswell notes, the researcher should engage in 'extensive data collection, drawing on multiple sources of information'. This research reflects this principle through its combination of observations, interviews, and performance analysis.

The interview process followed Creswell's guidance that qualitative research interviews should be designed to 'focus on understanding the meaning that the participants hold about the problem or issue'. This was particularly important in understanding the practitioners' perspectives on motion capture technology and its applications in performing arts. The semi-structured nature of the interviews aligns with Creswell's emphasis on allowing participants to shape their responses rather than being constrained by rigid questions.

The thematic analysis approach used in this study incorporates Creswell's recommendation that analysis should move from 'particulars to general' through 'multiple levels of abstraction'. This is reflected in the systematic process of coding and theme development, where specific observations and interview data were gradually synthesized into broader theoretical understandings about the application of machine learning-based motion capture in performing arts.

3.2.2 Reflective practice

When engaging in practice-based research, it is important to recognise how the act of practice itself can be viewed as a rigorous endeavour, contributing to the generation of knowledge. Scrivener highlights the relevance of Schön’s theory of reflective practice (Schön, 1992), originally oriented towards problem-solution based design, in understanding the process involved in a ‘creative-production’ task (Scrivener, 2000). According to

Scrivener, Schön's theory offers valuable insights into comprehending the nature of the creative-production process, the integration of past personal and collective experiences, the assessment of actions, the rigour within creative production, and the practitioner's stance (Scrivener, 2000).

Schön describes a process of reflection that naturally occurs during creative design practice while the practitioner is actively engaged in their professional activities. In a creative domain, practice involves continual moments of self-reflection, where practitioners assess the outcomes of their actions and adjust their processes accordingly. This ongoing process of action, observation, and reflection, occurring in immediate connection with the act of creation, is referred to as 'reflection-in-action', which takes place after the act of creation to assess what has been produced and its degree of success. Schön suggests that 'reflection-in-action' generates intuitive or 'tacit' knowledge within the practitioner (Schön, 1992).

Walkerden(2009) expands on Schön's concept of reflective practice by emphasising the importance of practitioner experience. He notes that 'practitioners orient themselves in their professional practice situation, developing a sense of what is at stake and what direction(s) it makes sense to head in.' This aligns with how reflection-in-action was employed during the development of the performance piece.

Reflection-in-action

The reflection-in-action process is similar to conducting research on a micro time scale. Practitioners involved in reflection-in-action do not seek unified models or frameworks at this stage but rapidly explore multiple aspects of a specific task. They rely on their previous expertise to guide their actions, but the knowledge gained through this process remains problem-specific (Schön, 1992).

The process of reflection-in-action is particularly pronounced when the creative task at hand "talks back" to the practitioner, presenting unexpected outcomes that trigger moments of reflection. This back-talk prompts reflection in problem-solution projects where the problem's exact nature is clearly defined. However, if the outcome of an action aligns precisely with expectations, reflection is not necessary, and the practitioner can proceed to the next task. Reflection and reassessment only occur when the situation responds with the unexpected

results, prompting the practitioner to generate intuitive knowledge about the outcomes of their actions and reconsider or reframe the task from a different perspective (Schön, 1992).

Reframing the situation is particularly crucial in artistic and creative projects, where the nature of the task often unfolds through the process of undertaking it. The back-talk encountered during reflection-in-action becomes vital to the discovery process, allowing artists to grasp the deeper potential meanings of the content they strive to create. An example of where reflection-in-action will play a pivotal role will be in the planning and execution of the performing arts piece in my research, particularly in the design of the performance. When collaborating with the performer, it became apparent that various aspects needed careful consideration, such as the position of the animated projection and the performer within the performance space. This iterative process involved a constant back-and-forth dialogue between myself and the performer. A delicate balance was needed, ensuring that neither element overshadows the other. For instance, if the animation is too dominant, it will risk drawing the audience's attention away from the performer's actions, potentially diluting the emotional impact of the live performance. Conversely, if the performer's actions take centre stage, the animated elements might be overlooked, diminishing the immersive quality of the experience.

To address this challenge, we engaged in continuous reflection-in-action. This process involved experimenting with different positions, scales, lighting configurations, and orientations of the animation within the performance space. Key adjustments were considered and evaluated during rehearsals. We observed how changes in lighting affect the visibility of the animation and how altering the performer's position influences the audience's focus. This dynamic feedback loop allowed us to fine-tune the performance layout iteratively.

Furthermore, reflection-in-action extend to the coordination of timing and synchronization between the performer's actions and the animation. We constantly adjusted cues and transitions to achieve a seamless integration of live performance and digital animation, fostering a sense of unity and narrative coherence. Additionally, unforeseen technical challenges surfaced during rehearsals, necessitating on-the-fly problem-solving and adaptations. These unexpected situations demanded quick thinking and immediate adjustments to maintain the integrity of the performance.

Reflection-on-action

While reflection-in-action unfolds during the active practice itself, reflection-on-action occurs after the practice has concluded. Reflection-in-action persists until the situation's back-talk is addressed and reframed, whereas reflection-on-action can span a more extended time scale. Scrivener distinguishes the purpose of reflection-on-action from that of reflection-in-action, stating that reflection-on-action is not driven solely by the unexpected but by the desire to learn from experience. It is a deliberate discipline rather than a necessity for further action (Scrivener, 2000).

The reflection process is further supported by Walkerden's (2009) framework which identifies three types of experimentation practitioners engage in simultaneously: 'exploratory—the probing, playful activity by which we get a feel for things; move testing—any deliberate action undertaken with an end in mind; and hypothesis testing—experimenting to discriminate amongst competing hypotheses.' This multi-layered approach to reflection informed the post-performance analysis.

The process of reflection-in-action enables creative practitioners to rapidly expand or refine their work while developing tacit knowledge about which actions will likely yield fruitful outcomes in the future. Reflection-on-action broadens the body of knowledge by reflecting on the practice itself, ideally resulting in a more comprehensive theory informed by the numerous instances of reflection-in-action. Through reflection-on-action, practitioners strive to gain a deeper understanding of their experiences and learn from them, contributing to the advancement of their field (Scrivener, 2000).

Reflection-on-action, occurring after the presentation of the performance, provides an opportunity to step back and gain a holistic perspective on the research and the actualisation of the performing arts piece. This post-performance reflection will allow me to fully absorb the experience and critically assess various elements of the performance, particularly the interaction between the performer and the animation.

One significant reflection that emerged is the dynamic interplay between the performer and the animation. I contemplated how varying degrees of leadership between the two elements could create a captivating duet. For instance, in some moments, the performer takes the lead, guiding the narrative and dictating the animation’s responses, while in other instances,

the animation will assume a more dominant role, influencing the performer's actions. This choreography introduced an exciting dimension to the performance, amplifying the sense of collaboration between the live and digital elements.

Feedback gathered through interviews with both the performer and the audience also play a role in the reflection-on-action process. Conversations with the performer provided insights into their experiences, challenges, and creative preferences during the performance. This feedback illuminated aspects where adjustments enhanced the performer's engagement and artistic expression. Equally important was the feedback from the audience, as it shed light on their perceptions, emotions, and interpretations of the performance.

Comments and suggestions from the audience offered valuable perspectives on how the performance resonated with them and where it might benefit from refinement. This external viewpoint was instrumental in validating the effectiveness of the performance and uncovered potential avenues for future development. Assessing how well the narrative or storytelling in the performance was received by the audience as well as feedback on what aspects of the performance could be improved or refined were crucial in evaluating the effectiveness of the performance, as well as understanding the broader impact.

Furthermore, reflection-on-action allowed for a broader consideration of the performance's impact within the context of the research. It prompted contemplation of how this research contributes to the field of performing arts and the utilisation of machine learning-based motion capture. It also lead to thoughts on the broader implications of this technology in artistic expression and interactive experiences.

3.2.3 Research objectives

3.2.3.1 Explore models for pose detection

The objectives of this research are to explore the use of contemporary monocular and multi-view machine learning models for pose detection in the performance space. Monocular models are simpler to set up because they don't require camera calibration data, unlike multiple camera models. The models for this study needed to have the capability to output data that can be interpreted by animation software. I.e. an animated 3D mesh, or an

animated skeleton with joints representing the human form in an interchangeable format. They also require less equipment but cannot be as accurate as multiple camera models because only one angle of the subject is recorded. For these reasons, both models are investigated for the potential benefits they could bring to practitioners in the performing arts space.

3.2.3.2 Understand the current methods of choreographers and dancers

Gaining a comprehensive understanding of existing methodologies related to the collection and integration of motion capture data into performance employed by choreographers and dancers is essential for this study. Through qualitative interviews with practitioners in the field, the research aimed to uncover current methods, motivations and reasons for utilising specific approaches in their creative processes. This objective explored how machine learning-based motion capture integrated into (and potentially complement or extend) their current methodologies and potentially enhance their artistic work. The recordings from these interviews were analysed using thematic analysis (Braun & Clarke, 2006), which will be covered in Chapter 3.5 ‘Data Analysis’, dedicated to the interview analysis.

3.2.3.3 Verify that the identified models are applicable to performance by collaboration

An additional objective of this practice-based research was deploying and testing machine learning-based motion capture models with dancers, choreographers and directors in dance production settings. Working with directors of dance performances who use motion capture as animation into their productions will be a practical test-case scenario for the technology. This objective aimed to assess these models' performance, accuracy, and suitability in capturing and generating motion capture data that can be translated into animated meshes in a dance production setting. By involving dancers, choreographers and directors, the research gathered valuable feedback and insights from end-users, ensuring that the process of generating animated meshes meet choreographers' specific needs and requirements for use in their performances. The findings from this objective will provide insight into the value and functionality of machine learning-based motion capture for practical application in the performing arts domain.

Through the deployment of machine learning models across diverse artistic contexts and with various practitioners, insights extended beyond technical performance characteristics. Valuable information was gleaned regarding practitioners' preferences, as exemplified by the emergence of preferences for unexpected outcomes or glitches (mentioned in Chapter

4.3.7). Additionally, the iterative collaborative process became apparent, revealing how practitioners employ initial model outputs to inform subsequent steps in their creative process. Moreover, observations concerning performance dynamics, such as the failure of models in scenarios involving performers in positions or performance environments where models cannot accurately capture their poses, contribute to a comprehensive understanding of the implications and limitations of the technology in artistic practice.

3.2.3.4 Create a performance involving a live performer demonstrating the use of these models

The final objective focuses on the practical application and evaluation of the motion capture to animation pipeline in a performance setting. By implementing the entire pipeline, which encompasses capturing motion data, generating animation, and integrating it into a live performance with a dancer, this objective aimed to achieve several goals. These include assessing the feasibility of the process, identifying potential areas of improvement, and evaluating the utility of this approach for performance practitioners and choreographers. Throughout this process, a significant amount of knowledge was gained about how this new technology could influence the creative process, and where novel opportunities emerged that may not have been immediately obvious. Collaboration, planning, and choreography are also key elements of this integrated pipeline. The insights gained from this objective contributed to understanding the practical implications and potential benefits of using machine learning-based motion capture in performing arts.

3.3 Data Collection from interviews of practitioners using motion capture in performance

Conducting successful thematic analysis requires careful attention to data collection. (Braun & Clarke, 2006) highlight key considerations: It is not sufficient to merely compile extracts of interviews without a coherent analytic narrative. Extracts serve as illustrations of the researcher's analytical points, providing evidence to support a broader analysis that makes sense of the data. The data collection process should facilitate the identification of meaningful patterns and themes. "The 'analysis' of the material… is a deliberate and self-consciously artful creation by the researcher, and must be constructed to persuade the reader of the plausibility of an argument" (Foster & Parker, 1995). Claims should be supported by the data, and the extracts should align with the analysis. To establish a convincing narrative, the researcher must ensure that their interpretations and analytic points are consistent with the data extracts (Braun & Clarke, 2006). A robust thematic analysis

should demonstrate consistency between the interpretations of the data and the underlying theoretical framework. The researcher's analytical lens should be congruent with the conceptualisation of the subject matter. Data collection should be guided by the theoretical framework to ensure that the analysis captures relevant information and aligns with the broader theoretical perspectives (Braun & Clarke, 2006). Rigorous thematic analysis necessitates using a systematic method that aligns with the researcher's conceptualisation of the subject matter (Reicher & Taylor, 2005). By employing a structured and transparent approach, the researcher can enhance trustworthiness and credibility of their analysis (Braun & Clarke, 2006).

Booth, Colomb and Williams (2003) emphasize that successful data collection relies heavily on careful preparation and documentation. They note that "the more you sort out what you know from what you want to know, the more efficiently you will get what you need". This principle guided the preparation of interview questions and the overall data collection strategy.

3.3.1 Observations

Observations were made of the performer as they progressed through the stages of development of the performance piece. The performer was observed during rehearsals for creating the motion capture data. This stage involved embodying the movements and actions that were later translated into animation. The observations shed light on the performer's engagement with the creative process, their physical execution of movements, and their interaction with the technology used for motion capture. The performer had prior knowledge of what the final animation will look like and the environment The University of Technology Sydney's (UTS) data arena where they will be performing (details about the data arena are discussed in Chapter 7.2.1 below). Observing the performer embodying these aspects and the music was insightful.

One particular aspect was gaining insight into Cloé's (the performer) interpretation of the music while having the animation in mind. This observation offered a unique perspective on how she translated musical cues into movements and expressions, which in turn allowed me to envision how the animation synchronized and responded to her performance. Elements such as her poses, speed of her movements, and the spatial dimensions she occupied within the performance space came into focus. These observations not only enriched my

understanding of the interplay between the performer and the digital animations but also enabled me to offer feedback to Cloé.

Further observations were made of Cloé on the day of the performance. Although the animation was recorded offline, observations provided insights into how the performer interacted with the animation in a spatial and embodied context. Observing the performer's physical engagement with the animation in the data arena allowed an understanding of how the animated representation of their movements resonated with their own embodied experience. For example, how the performer reoriented themselves in relation to the projected animation and use visual cues in the animation when performing.

During the developmental stage of testing machine learning models for performance, observations were made during the rehearsal of performers in aerial slings[36]. The aerial sling is a versatile aerial apparatus consisting of a loop of fabric suspended from a single rigging point, creating a U-shaped configuration similar to a swing. This suspended fabric loop offers performers to explore movement both within and around its structure. Performers can utilise the sling in multiple ways: sitting or standing within its cradle, balancing on top of the fabric loop, or executing movements that flow between these positions(*Aerial Silks vs. Aerial Sling: What's the Difference and Which One Is Right for You?*, 2023). Observations during the rehearsal allowed for an in-depth understanding of how performers interact with aerial slings and their interaction with motion capture technology. Another critical aspect of the data collection process was obtaining verbal feedback from the performers after demonstrating examples of captured motion capture data converted into abstract animation. These informal conversations were insightful, as they will give some idea of how performers perceive their movements' representation in the abstract animation. It shed light on their reactions, emotions, and the extent to which the animation captured the essence of their performance. Figure 3.1 shows a performer using aerial slings in a performance.

[36] Aerial slings, https://www.verticalwise.com/aerial-silks-vs-aerial-sling/

Fig. 3.1 A performer using aerial slings.

Working closely with performance directors and choreographers to explore the integration of machine learning technology into their production process provided valuable insights into their creative approach, preferences, and workflow. The directors sent video clips of performances and request the machine learning model to interpret the motion by outputting an animated mesh. The animated mesh, after some processing, featured in their performances. Observing the selection of these clips provided insights into their artistic vision and the types of movements they considered significant for their production. These observations allowed for a deeper understanding of how the directors navigate the possibilities and limitations of machine learning technology. Observing the collaboration between the research, the technology, and the directors provided valuable insights into the dynamics of artistic vision and creative exploration. The observations revealed the iterative dialogue between all parties, highlighting the project's evolving nature. The observation process offered an understanding of how the directors balance their artistic intentions, the possibilities offered by machine learning technology, and the performers' embodied execution of the interpreted movements.

The approach to observational data collection was informed by established research principles that emphasize the importance of thorough documentation and careful notetaking.

As Booth et al. (2003) mention, complete and accurate recording of observations is crucial for maintaining research integrity and enabling thorough analysis later in the process.

3.3.2 Interviews

A challenge that can arise when conducting interviews is the risk of allowing the interviewer's assumptions to influence the research, and it is crucial to set aside any preconceived ideas beforehand (Anderson & Kirkpatrick, 2016). The goal was to approach the interview as exploratory, allowing the participants to freely share their experiences and insights. Simultaneously, the questions posed were designed to be open-ended, ensuring that they were tied to topics of interest and relevant to the core research questions and objectives. To obtain authentic data, an active engagement with the interviewees was required to explore their answers. This approach ensures that the data collected reflects the perspectives and experiences of the participants (Durkin et al., 2020). Open-ended and probing questioning techniques coupled with active listening were employed to facilitate a comprehensive exploration of the participants' perspectives. By allowing the participants to shape their narrative and giving them space to unfold their stories in their chosen manner, a deeper understanding of their experiences and insights was gained (Durkin et al., 2020; Serry & Liamputtong, 2013).

Interviews were conducted on three occasions during this research and are detailed below.

Practitioners using motion capture

A series of interviews were conducted with practitioners who utilise motion capture technology in their work. These interviews aimed to understand the types of technology currently employed and to discuss the pros and cons of these methods, with the goal of identifying the potential role of machine learning methods for motion capture in their workflow. In conducting these interviews, it was essential to employ effective interviewing techniques to ensure the authenticity and depth of the data collected. The interviews served as a means of gathering valuable insights and perspectives from individuals with diverse backgrounds in the field. Participants included academics and researchers, performers and choreographers, as well as professionals with experience in the commercial motion capture industry. The interviews were semi-structured, allowing for a flexible and organic conversation while ensuring that key topics and themes are explored. Each interview lasted

between 30 minutes to an hour, providing ample time for participants to share their experiences and opinions in depth.

To ensure the anonymity and confidentiality of the participants, their identities were anonymised. This approach was taken to encourage open and honest responses from the participants, fostering a safe space for sharing their thoughts and experiences. The interview questions were thoughtfully crafted to delve into the participants' experiences with motion capture. The interviews commenced by exploring the extent of their experience with motion capture systems and discussed their encounters with such technology. Participants were encouraged to share their likes and dislikes about the existing motion capture systems they have used, enabling a detailed understanding of their perspectives. Without mentioning machine learning to the participants, and understanding that machine learning methods have their own drawbacks, a key question posed to the participants was focused on the trade-off between the setup time, cost, and portability and the accuracy of motion capture data. The participants were then asked to reflect on whether they would be willing to trade off accuracy in exchange for the setup time, cost, and portability and if it would unlock any new possibilities or opportunities for their work. This question played a pivotal role in gauging the participants' openness to embracing machine learning methods for motion capture in their creative workflow. It aimed to ascertain if the potential benefits of new opportunities outweigh the potential compromise on the accuracy of the captured data. The insights gained from participants' responses to the questions contributed significantly to the understanding of where machine learning models could potentially fit into performance workflows. The interview questions are provided in Chapter 11.2 'Interview questions', below.

The performer

An interview was also conducted with the performer who participated in the presentation of the performance piece. The purpose of interviewing the performer was to get feedback about their experience of the process of using machine learning methods. The questions were structured to delve into the performer's past experiences with motion capture and understand their process when using conventional methods. This allowed for an understanding of the range of techniques and technologies they have been involved with in their performances. By gathering this information, the research seeks to explore the practices and challenges of a performer when incorporating motion capture in their artistic endeavours.

After the production of the performance piece, the performer was interviewed about her experience related to her previous experience with motion capture in a performing arts context. Once the performer's past experience with conventional methods was understood, the interview shifted focus to discuss how machine learning methods used in the performance piece differed from their previous practices. This section of the interview seeks to highlight the distinctions between traditional motion capture techniques and the innovative machine learning-based approach employed in the performance. By comparing and contrasting the two approaches, the interview aims to gain insights into the performer's perspective on the unique features and potential benefits of machine learning methods for motion capture. Understanding the performer's observations and perceptions was crucial in gauging how machine learning methods introduced new possibilities or addressed limitations they might have faced with conventional methods. The interview questions will be provided in Chapter 11.2 ‘Interview questions’, below.

The audience

The post-performance interviews with the audience offered valuable insights into their engagement. The purpose of these interviews was to gauge the audience’s overall experience, their perceptions of the performance, and gather their thoughts and impressions. A significant portion of the questions delved into the audience’s thoughts about the performance, exploring whether they had encountered similar experiences in the past and how this performance compared to their previous encounters. A key area of interest in the audience interviews was to gather feedback on whether they perceived the motion capture as live or not. As the motion capture and animation were pre-rendered offline, there was a particular curiosity as to whether the audience detected any indication that the motion capture data was not live during the performance. Insights from the audience’s perceptions regarding the live nature of the motion capture were helpful in evaluating the success of the performance in creating a seamless and immersive experience. Understanding whether the audience picked up on the offline rendering of the motion capture provided valuable feedback on the effectiveness of the technology in conveying a live and dynamic performance. If the audience frequently perceives a performance to have live motion capture when it's actually offline, it may indicate that motion capture processed offline can be done without the limitations of a live motion capture performance. Live motion capture often involves real-time processing, which may necessitate compromises like lower resolution or simpler graphics to maintain performance speed. In contrast, offline or pre-rendered graphics can be more complex and feature higher quality lighting and shadows, enhancing

the overall aesthetic quality of the animation. Audience responses in this context can inform creators about the potential for pushing the boundaries of motion capture technology to achieve visually complex and intricate results, even if not performed in real-time. This insight can be especially valuable in fields where aesthetics and visual impact are paramount. The interview questions will be provided in Chapter 11.2 ‘Interview questions’, below.

3.3.3 Self-reflection

Throughout the research process, self-reflection played a crucial role in understanding the purpose and significance of creating the performance piece. This introspective practice extended beyond the performance piece itself, encompassing various elements of the research, including the interviews with practitioners who utilise motion capture in performing arts, collaborative interactions with artists, choreographers, and directors, as well as the diverse stages of testing machine learning-based models for their applicability in the realm of performing arts. A key goal behind developing the performance piece and other collaborative artefacts was to introduce practitioners in the field of performing arts to machine learning technology, enabling them to explore its potential benefits and consider its integration into their creative practices. The performance piece also became a proof-of-concept where the entire pipeline from motion capture acquisition, collaboration with the performer, to final abstract animation presented as a performance could be explored. It is important to clarify that while the performance piece was an essential part of the research, it was not the primary output of the dissertation. Instead, the focus was on using the performance piece as a tangible example of how machine learning technology can be effectively employed in performing arts. By showcasing a practical application of machine learning in the creation of abstract animations based on motion capture data, an aim was to inspire non-technical artists who may not have extensive coding experience to recognise this technology's possibilities. Another key aim is to explore the impact of machine learning-based motion capture on the creative process and to explore its limits and benefits in a practical performance setting.

The use of open-source models was a deliberate decision, allowing the performance piece to be accessible to anyone interested in exploring machine learning in their creative endeavours. The research sought to democratise access to machine learning technology by utilising openly available resources, enabling artists from diverse backgrounds to engage with and benefit from this innovative approach. The performance piece presented a singular

illustration of how machine learning technology can be harnessed within performing arts, but its outcomes and potential applications are more general.

3.4 Data analysis

The framework for data gathering, inspired by Braun and Clarke's approach (2006), involves observations and interviews as primary data sources. The researcher plays an active role throughout the data analysis process by coding, sorting, and identifying themes within the data.

3.4.1 Familiarisation with the data

One of the foundational phases in thematic analysis is the process of data familiarisation. The researcher should immerse themselves in the data to grasp the breadth and depth of its content. Immersion typically involves repeated reading of the data, but it goes beyond mere passive reading. Researchers should actively engage with the data, searching for meanings, patterns, and nuances that could inform the subsequent analysis. Taking notes or marking ideas for coding during this initial immersion phase can be beneficial for later stages of the analysis (Braun & Clarke, 2006). Transcription plays a significant role in familiarising oneself with the data, and it is often considered a key phase of data analysis within interpretative qualitative methodology (Bird, 2005). The act of transcription should be recognised as an interpretive process, where meanings are created rather than simply being a mechanical act of transferring spoken sounds to paper (Lapadat & Lindsay, 1999). Therefore, the transcription should be approached with care and attention to detail to ensure a faithful representation of the verbal and, when applicable, nonverbal utterances. A rigorous and thorough "orthographic" transcript is essential, capturing a "verbatim" account of all verbal utterances. The goal is to retain the necessary information from the verbal account and maintain its original nature faithfully (Braun & Clarke, 2006). Proper transcription facilitates delving into the details of the data, gaining insights into the participants' words, expressions, and interactions.

The data is read and re-read, noting initial ideas, possible codes, and potential themes. This active reading and note-taking process sets the stage for the subsequent phases of thematic analysis, providing a solid foundation for generating initial codes and searching for themes.

3.4.2 Generating initial codes

The code generation process is where specific features of the data are identified and labelled, that appear interesting and meaningful in relation to the phenomenon under investigation. Codes are the most basic segments of elements of the raw data that can be assessed in a meaningful way (Boyatzis, 1998). The codes serve as the building blocks of analysis, allowing the researcher to organise the data into meaningful groups (Tuckett, 2005) that will form the basis of subsequent themes. Coding is an integral part of the analysis (Miles & Huberman, 1994) and is distinct from themes, which are typically broader and more conceptual in nature. The entire dataset is systemically analysed during the coding process, paying close attention to each data item. The objective is to identify interesting aspects and patterns in the data that may recur across the dataset, potentially forming the basis for themes (Braun & Clarke, 2006).

It is essential to ensure that all relevant data extracts are coded and then collated within each code, providing a comprehensive overview of the identified themes and patterns (Braun & Clarke, 2006). For the coding process of the interviews from practitioners, the software Nvivo[37] was employed, allowing for systemic and efficient data organisation. Many potential themes and patterns were coded as possible, as some aspects may not seem immediately relevant but could prove significant later in the analysis. Moreover, inclusivity is crucial when coding data extracts, meaning that relevant context should be retained whenever possible to avoid losing the broader meaning and context of the data (Bryman, 2001).

3.4.3 Searching for themes

During this phase, the researcher explores how different codes might come together to form overarching themes. The goal is to identify common patterns and connections within the codes and collate relevant data extracts that align with each identified theme (Braun & Clarke, 2006). As the analysis evolves, visual representation helps organise and sort the different codes into potential themes. Using Nvivo software, tables are created to sort them into themes, as outlined in the literature (Braun & Clarke, 2006). This phase invites the consideration of relationships between codes, themes and different levels of themes, such

[37] Nvivo, https://lumivero.com/products/nvivo/

as main overarching themes and sub-themes within them. By the end of this phase, a collection of candidate themes and sub-themes are gathered, along with corresponding data extracts that have been coded in relation to them. At this point, initial insights into the significance of individual themes may be noted. However, it is essential not to make premature decisions about the themes during this phase. The validation of the themes will come during the next stage, when all the coded extracts will be examined in detail. This cautionary approach ensures that nothing is abandoned prematurely, as a thorough examination of all data extracts will be necessary to determine whether the identified themes hold up to scrutiny. During the next phase, each data extract is closely analysed within the identified themes, refining, combining, separating, or possibly discarding themes as needed (Braun & Clarke, 2006).

3.4.4 Reviewing themes

Reviewing themes is where the candidate themes identified in the previous phase are refined and validated. Some themes are discarded if there is insufficient data to support them or if the data is too diverse to form a coherent pattern (Braun & Clarke, 2006). There are two levels of reviewing and refining themes. The first level is where the coded data extracts within each theme are reviewed. Reading all the collated extracts for each theme helps to assess whether they form a coherent pattern and if the data within each theme cohere together meaningfully. If the potential themes appear to hold up at this level, then the second level of this phase begins (Braun & Clarke, 2006).

At level two, the validity of individual themes in relation to the dataset is considered. This involves verifying whether the potential thematic map accurately reflects the meaning evident in the data as a whole. The goal is to ensure that the themes are not only coherent within themselves but also align with the overall story the data tells. This comprehensive review ensures that the thematic analysis captures the essence and intricacies of the data in a meaningful and accurate manner (Braun & Clarke, 2006). The entire dataset is reread for two purposes. Firstly, to ascertain whether the themes work effectively in relation to the dataset, validating their coherence and validity. Secondly, any additional data is coded that may have been missed in earlier coding stages (Braun & Clarke, 2006). By the end of this phase, a clearer picture is formed of the different themes, how they interconnect, and the overall story they collectively convey about the data. This in-depth review and validation process ensures that the final thematic analysis accurately reflects the depth and richness of the data, providing valuable insights and interpretations relevant to the research question and objectives (Braun & Clarke, 2006).

3.4.5 Defining and naming themes

This phase is where the essence of each theme is identified. It is essential to avoid overhardening a theme by attempting to encompass too much or making it overly complex. The goal is to define the themes by organising collated data extracts into a coherent and internally consistent account supported by accompanying narratives (Braun & Clarke, 2006). The collated data extracts for each theme are re-visited, where a detailed analysis is conducted. A thoughtful examination of what is interesting and significant about the data extracts and why they are relevant to the theme is required. By identifying the unique story each theme tells, a deeper understanding of the specific insights and patterns present with the data is gained.

Additionally, it is crucial to consider how each theme fits into the broader overall narrative of the research. The themes need to be assessed in relation to one another to ensure there is no unnecessary overlap and that each theme contributes distinct and complementary perspectives to the research question (Braun & Clarke, 2006). Sub-themes are also identified within each main theme, adding depth and nuance to the analysis. The recognition of sub-themes further enrich the narrative and provide a comprehensive account of the data's complexities (Braun & Clarke, 2006). Ultimately, the defining and naming of themes involves crafting a coherent and internally consistent account of the data that goes beyond simple summary. It requires a thoughtful and analytical approach to extract the essence of each theme and its contribution to the broader research narrative. The thematic analysis gains depth and robustness by considering the relationships between themes and potential sub-themes, providing valuable insights and contributing to a comprehensive understanding of the research topic (Braun & Clarke, 2006).

3.4.6 Producing the final report

The final report of the interviews with practitioners using motion capture in performance is in Chaper 4 'Interviews with practitioners in the performing arts who use motion capture'. It is the culmination of the thematic analysis process, where the complex story of the data is summarised and conveyed. It is intended to be concise, coherent, logical and interesting as it unfolds the story that the data tells across themes (Braun & Clarke, 2006). A well-constructed report ensures that the analysis is effectively communicated and the key insights and patterns within the data are highlighted. It provides a clear account of the themes and

sub-themes of the data. Data extracts are included to demonstrate the relevance of each, which also serve as concrete examples that illustrate the patterns and connections identified during the analysis (Braun & Clarke, 2006). By structuring the report, the reader is navigated though the analysis and can appreciate the interconnectedness of the themes. Throughout the report, a clear focus on the research question is maintained. By aligning the analysis with the original research objectives, the report ensures that the findings are relevant and contribute meaningfully to the broader field of study. The reader should be able to discern how the analysis addresses the research question and how each theme relates to the overall research narrative (Braun & Clarke, 2006).

3.5 Artefact Design

Johnston (2014) states that ontological and epistemological positions flow from establishing the core question: who is this research for? This research is not developing a new machine learning method but determining the place of machine learning-based methods of motion capture in a performing arts context. Therefore, this research is aimed at creative practitioners in the performing arts who are currently using motion capture technology or potentially interested in using it in the future. The main artefact, produced from this research is a performance piece where the performer dances alongside projected animation derived from their recorded motion. The artefact is intended to be displayed in the University of Technology Sydney (UTS) Data Area, which projects images in a circular room stereoscopically. There are other initial exploratory pieces that led to this final artefact as a culmination, discussed in Chapters 5 and 6.

There were a few key areas which needed to be met before the artefact was ready to be presented: To ensure the suitability of a machine learning model for generating motion capture data from a performer's movements, it's necessary to locate an appropriate model. This model should then undergo testing to validate its effectiveness in producing compatible animation data that can be applied within animation software. An additional facet of the artefact design involved experimentation with various animation concepts aligned with the performance. The integration of music in the performance served to unify the visual elements within the virtual environment and complemented the choreography, devised by the dancer, Cloé. As the artefact design progressed, and each component neared completion, iterative adjustments were made through reflection to enhance cohesiveness and achieve a visually appealing aesthetic.

3.5.1 The development of the artefact

In this section, we focus solely on the methodological framework and processes that guide artefact design. The specific details and implementation of our artefact will be explored comprehensively in Chapter 7. This separation allows us to first establish the foundational design principles and methodological approaches before examining their practical application.

The model

One of the main objectives for the development of the artefact is to find a suitable model capable of detecting human poses and generating motion capture data in a form suitable for abstract animation. The model should output a skeleton that can be manipulated and smoothed, facilitating the animation process. A comprehensive investigation was conducted using open-source repositories to source appropriate models that meet human pose detection requirements and output motion capture data in a compatible file format for 3D animation applications to process. Multiple models, including monocular and multi-camera approaches, were examined and evaluated based on their performance, compatibility and ease of integration. After careful consideration, the best performing model, when assessed against key metrics, was selected to underpin the development of the final artefact.

The animation

The other objective of developing the artefact is to look for inspiration for appealing animations that I can integrate with an animated mesh in the performance. Abstract animation was selected that is visually interesting and suited to being in a virtual gallery environment. Extensive exploration of existing animation that could be applied to animated meshes was conducted to identify animations that align with the desired visual aesthetic and the music chosen for the piece. Examples of animation were evaluated on animated human meshes to ensure their compatibility. The aim was to create a symbiotic relationship between the performer and the animation, where both entities enhance and harmonise with each other’s movements, blurring the line between the physical and the digital. It was decided to render the aminations offline to ensure optimal visual quality and synchronisation with the performer. This approach allowed for refinement and post-processing of the animations, resulting in visually polished content which seamlessly integrated with the live performance. The location of the performance was in a 360-degree room with projectors capable of displaying stereoscopic imagery, which also was a factor in determining that the

animation should be rendered offline. An art gallery environment was chosen as the environment to showcase the animations. Integrating visually appealing animations into the motion capture-driven artefact would enhance the immersive experience within the virtual gallery environment. It would create sensory experience, engaging the audience in a visually captivating and emotionally impactful manner.

The performance

Cloé, a professional dancer located in Sydney, was presented with an example of what the final animation would look like as part of the preparatory process, the accompanying music, and the virtual art gallery environment. This in-depth understanding allowed her to synchronise her movements with the visual elements, targeting in a cohesive and visually captivating performance. Cloé is introduced to the immersive 360-degree room (the Data Arena), which serves as the performance space. Within this environment, she can explore the spatial layout and envision her movements in direct relation to the animations driven by the motion capture data she generates. This familiarisation amplifies Cloé's sense of presence and engagement with the performance space.

To offer Cloé a holistic view of the final product, a stereoscopic preview of the animation is showcased within the room. This preview allows her to visualise the immersive and three-dimensional aspects of the performance, thus enriching her creative expression during the actual performance. Additionally, carefully positioned spotlights enhance her presence on the stage, while a multitude of animations surrounds her within the 360-degree room. Positioned symmetrically on both sides of Cloé, these animations remain synchronised with her movements, creating an interconnected visual experience. Table 3.1 illustrates the process of artefact development.

Artefact development process	Example of animation presented to performer, along with music and virtual art gallery environment
	Performer is introduced to Data Area performance space
	A stereoscopic 360-degree presentation of example animation presented to performer in the Data Arena
	Performer rehearses with the music with prior knowledge of the performance space and animation
	Performer performs with the music and motion capture is recorded
	Motion capture is processed and a mesh is output
	Abstract animation is generated from the animated mesh
	Gallery virtual environment with the abstract animation is rendered to a format compatible with the Data Arena
	Result projected in Data Arena with music

Table 3.1: Artefact development process.

3.6 Reflections on the conceptual decisions made

The creative and conceptual decisions evolved through three main phases: initial experiments, artistic collaborations, and the final performance. The initial experimental phase occurred during COVID restrictions, which imposed certain constraints that ultimately proved informative. Working with a performer remotely, using HD720 video footage that was handheld and had sub-optimal lighting, demonstrated that the system could function effectively even under less-than-ideal conditions. While these conditions weren't chosen deliberately, they helped establish the system's robustness and adaptability to real-world scenarios.

The collaborative phase with performing arts choreographers and directors provided valuable insights into how the technology could be integrated into existing creative practices.

Although the movement choices were predetermined by the directors, this phase validated the system's capability to serve actual production needs and built confidence in its practical applications.

The final performance phase involved several intentional creative decisions:

The selection of the performer was strategic, choosing a professional dancer with international experience and prior exposure to motion capture performances. This background enabled more nuanced exploration of the technology's capabilities and facilitated clearer communication about technical and artistic requirements.

The choice of the Data Arena as the performance venue was driven by its unique capabilities: 360-degree wraparound projection, stereoscopic display, and integrated sound system. These features created an immersive experience for the audience. While the space presented certain limitations for performer and audience movement, these constraints were incorporated into the choreographic process rather than viewed as obstacles.

The variety of animation styles was designed to create different levels of audience engagement. Some animations provided immediate, direct response to the performer's movements, while others introduced temporal delays or cumulative effects, creating varying degrees of connection between physical and digital movement. The addition of animated strands provided yet another layer of visual interpretation. The decision to set the animations within a virtual art gallery environment was both aesthetic and conceptual. By framing the animations as moving artworks within a gallery context, it provided audiences with a familiar framework for viewing and interpreting the work. The static, subtly lit background was deliberately understated to maintain focus on the dynamic relationship between performer and animations.

These creative and conceptual decisions were not made in isolation but evolved through practice, each phase informing and building upon the previous ones. The process demonstrated how technical capabilities, artistic vision, and practical constraints could be balanced to create a cohesive performance experience.

3.7 Ethical considerations

This research adheres to the rigorous ethical guidelines set by the University of Technology Sydney to ensure the ethical treatment of all participants involved. Before commencing the data collection, a comprehensive ethical review process was undertaken, including outlining the research aims, methodology, significance, location, number or participants, and potential risks. The review process involved peer evaluation to ensure the research's ethical standards were met and approved. All participants in this research were fully informed in writing of the research procedure and the implications of their participation before engaging in any data collection. This included the performer, audience members who attended the performance and answered questions, and the interviewees involved in the field of performing arts. Prior to any involvement, each participant received a detailed explanation of the research process, the data collection methods, and their rights as participants.

The ethical implications of research involving human subjects have become increasingly important in contemporary research practice. As Booth et al. (2003) note, researchers must be aware that studying people can potentially cause harm "not just physically but emotionally, by embarrassing them or violating their privacy". This understanding reinforced the importance of maintaining strict ethical standards throughout the research process.

Informed consent was obtained from all participants, emphasising their freedom to withdraw their involvement at any stage of the research without any question or repercussion. They were assured that their decision to participate or withdraw would not impact their relationship with the researcher or the University. The consent and information forms used to inform participants about the research process and obtain their consent are included in the Appendix section of this document. These forms outline the purpose of the research, the procedures involved, the rights of the participants, and the steps taken to ensure protection, confidentiality and anonymity (where applicable).
The project numbers listed below are related to the ethics approval for this research:

Performing arts practitioners' interviews

2022-4 UTS HREC ETH19-3452

Performer and interview

2023-3 UTS HREC ETH19-3452

Post-performance audience questionnaire

2023-8 UTS HREC ETH19-3452

3.8 Conclusion

This research methodology has been designed to produce new insight into the practical application and integration of machine learning-based pose detection systems in performance contexts, and is intended as a resource for practitioners looking to understand how and where such systems may deliver value from creative, efficiency and cost perspectives. An aspect of this practice-based research involves generating motion capture data from a dancer, transforming it into abstract animation, and presenting it as a performance in the University of Technology Sydney's Data Arena. The primary objectives of these tasks were to explore, evaluate, and demonstrate the capabilities, possibilities, and validity of machine-learning based motion capture systems in the realm of performing arts. Through reflective practice, incorporating reflection-in-action and reflection-on-action, the research aimed to develop a deep understanding of the subject matter and uncover valuable insights based on practical real-world application of the technology, interviews with practitioners and audience members, collaborative iteration on creative outputs, and the full production and demonstration of a centrepiece artefact.

The production of the artefact, a dynamic representation of the research's outcomes, emerged through a process of self-reflection and analysis. By drawing on interviews, observations, and self-reflection in conjunction with thematic analysis, insights from then was examined comprehensively to discern meaningful patterns and themes. The research ventured to understand the current methods utilised by choreographers and test the identified machine learning models with professionals in a production setting. Through interviews and observations, the project sought to gain insights into the experiences and perspectives of those working in performing arts, enabling a comprehensive evaluation of the machine learning-based motion capture systems' practicality and applicability.

4 Interviews with Practitioners in Performing Arts who use Motion Capture

4.1 Introduction

The convergence of technology and artistic expression has led to the integration of motion capture systems within the domain of performing arts giving rise to novel avenues for exploration and creative development. This chapter addresses the research question: ‘In performing arts, what are the characteristics creative practitioners look for in motion capture systems?’ (RQ1) Through a systematic examination of the experiences and insights of motion capture practitioners through interviews, this research seeks to elucidate the potential applications, challenges, and implications of employing machine learning-based motion capture systems in the context of performing arts.

To answer the research question above, the needs of practitioners in this field would have to be discovered first, which was the primary goal of these interviews. The interviews conducted for this study provide a platform for practitioners to articulate their insights into the technical, artistic, and logistical aspects of employing motion capture systems. Through these narratives, we gain a comprehensive understanding of the challenge’s practitioners face, the creative opportunities they pursue, and the potential barriers that may arise when utilising motion capture in their performances. The insights garnered from these interviews hold significance not only for individual practitioners but also for the broader discourse on the fusion of motion capture technology and performing arts. By unveiling the specific needs and requisites of motion capture practitioners, this research contributes to a more nuanced understanding of how machine learning methods of motion capture can fit into or improve current methods. The outcomes of this exploration have the potential to inform the design, development, and integration of motion capture systems that align more effectively with the artistic intentions and practical requirements of performers. The Figure 4.1 below serves as a visual roadmap of the thesis, showing both the research questions and their corresponding methodologies. The current section's focus is emphasised by highlighting the first research question and its associated method, while the remaining elements are shown in grey. Similar figures appear throughout the thesis to help orient readers, allowing them to track their progress through the document and understand how each section connects to the broader research framework.

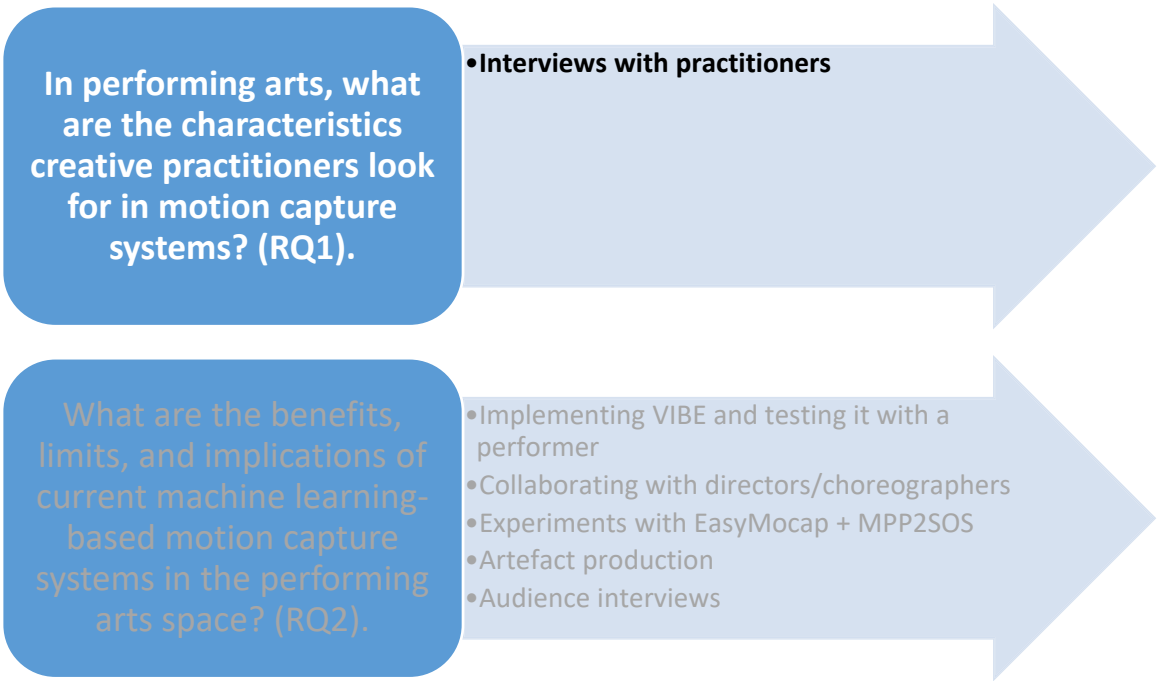

Fig. 4.1 Interviews with practitioners to address research question two.

4.2 Methodology

This chapter employs a qualitative research approach centred around semi-structured interviews to investigate the needs and experiences of motion capture practitioners in performing arts. The decision to employ semi-structured interviews as the primary research method stems from their ability to provide in-depth exploration of individual perspectives, insights, and experiences. This method allows for a rich and detailed account of the practitioners' encounters with motion capture systems, thus affording a nuanced understanding of their requirements and aspirations. The interviews were conducted with a diverse group of motion capture practitioners who possess firsthand experience in utilising motion capture systems within the context of performing arts. The interviews were semi-structured, allowing respondents the space to elaborate on their unique journeys, insights, and challenges. The information gathered during the interviews was subsequently transcribed verbatim to ensure accuracy in capturing the participants' perspectives. The methods for identifying themes are discussed in detail in Chapter 3.4 'Thematic analysis'.

Interview analysis

The software Nvivo was employed to analyse the qualitative data collected from the interviews. Nvivo facilitated a systematic and rigorous analysis process, allowing for the identification of recurring themes, patterns, and noteworthy insights within the dataset. The software aided in organising, coding, and categorising the transcribed data.

The interviews were conducted using Zoom[38], online video conferencing software, which provided a convenient platform for remote interactions with participants. This choice of platform allowed for flexible scheduling, overcoming geographical barriers, and accommodating the diverse locations of motion capture practitioners. During each interview, the video and audio were recorded to ensure accurate transcriptions and capture non-verbal cues that might contribute to a more comprehensive understanding of the participants' responses.

Selection criteria

Participants for this study were selected based on their substantial involvement in motion capture for performing arts. The participants consisted of professionals spanning from researchers delving into innovative technologies to choreographers and directors immersed in the creative process of performing arts productions. The diversity of participant backgrounds aimed to capture a broad view of the various dimensions, challenges, and opportunities surrounding the utilisation of motion capture in performing arts. Seven participants were engaged in the interviews, each possessing a unique vantage point and wealth of experience in the field.

- *Participant one* is a creative practitioner, possessing a diverse skill set that includes audio-visual art, software design, and musicianship. His body of work includes the development of audio-visual experiences for music visualisation and the creation of graphical presentations for interactive theatre performances.
- *Participant two* is an academic, specialising in performance studies. She has served as an advisor for a theatre and dance advisory board and her PhD thesis focussed on motion capture.
- *Participant three* is an academic with a notable record of presenting and performing artworks on national and international platforms. Her artistic pursuits have been

[38] Zoom, https://zoom.us/

dedicated to exploring movement-based approaches in the realm of wearable technology design.

- *Participant four* has roles as a dance director, choreographer, and performer. His experience spans a spectrum of performances, both on local and international stages. His work is often characterised by the integration of technology, notably motion capture.
- *Participant five* is an artist renowned for his collaborative work within contemporary performance contexts, extending across film, live performance, installation art and online spaces. He has received multiple awards and motion capture technology has played a pivotal role in many of his performances.
- *Participant six* is a researcher specialising in the development of technology to enhance motion capture techniques in performing arts.
- *Participant seven* is a prominent new media interaction artist with a specialisation in leveraging technology to enhance live musical performances, theatre productions, and interactive art installations. He has been involved in numerous performances that incorporate live motion capture technology.

4.3 Results

4.3.1 The selection of a motion capture system

This section presents the outcomes of the interviews conducted with practitioners using motion capture technology in the domain of performing arts. These interviews aimed to uncover the decision-making processes that inform the selection of motion capture systems based on their specific needs. Beyond a mere enumeration of technical specifications and practical considerations like cost and portability, the interviews aimed to delve into the multifaceted factors influencing practitioners' choices. This section analyses the participants' insights into the context, creative intentions, team dynamics, and technical limitations that collectively shape their decisions regarding motion capture technology. The aim is to offer a rich view into how selecting a particular motion capture system emerges as a product of the complex interplay of elements within the performing arts landscape. By dissecting these decision-making processes, this section seeks to provide a deeper understanding of how practitioners navigate the intersection of technology and artistic expression in their pursuit of leveraging motion capture systems for performing arts.

The selection of a motion capture system within the context of performing arts is a complex decision influenced by a myriad of factors, each uniquely shaping a project's artistic and technical trajectory. Beyond the foundational considerations of cost and portability, practitioners evaluate a range of aspects to determine the most suitable motion capture system for their specific creative goals. A pervasive sentiment mentioned in the interviews was the influence of context on the system selection. One participant stated, “And it depends, of course, for what purpose you use it”. The context in which motion capture is employed was cited as a pivotal factor that guided the choice of a particular system. This contextual awareness underscored practitioners’ ability to align technology with their artistic intentions, ensuring an integration that complemented the larger performance narrative.

Interestingly, the human element played a significant role in system selection. The availability of individuals with specific technological expertise and their familiarity with particular motion capture systems often influenced the decision-making process. Participants noted that the selection of a motion capture system sometimes hinged more on the capabilities of the team members than on a predetermined set of technical requirements. The sentiments were encapsulated in comments such as, “So it was kind of like about who was around” and “it tends to be circulated more around who’s in the room than it is around a specific set of skills”. The range of motion capture systems available introduced a layer of complexity to the selection process. A participant summed it up by stating, “To me, the big thing about MoCap is there’s no answer. There’s just picking the best thing for the situation.” The balance between system capabilities and the specific demands of the performance underscored the adaptive nature of motion capture integration.

A pragmatic acceptance of limitations was another recurring theme. Participants recognised that every motion capture system, regardless of its sophistication, possessed inherent limitations. This understanding led to a willingness to embrace imperfections and navigate challenges. As one participant expressed, “So we in general, try to redress that by accepting the limitations of technology at times and accepting some jankiness.”

4.3.2 The importance of accuracy in motion capture systems for performing arts

The topic of accuracy in motion capture systems emerged as a central theme during the interviews, sparking in-depth discussions about the significance of precision within the context of performing arts. While accuracy was a notable consideration for many participants, its importance was contingent upon the specific nature of their creative projects and the roles they fulfilled in performing arts. Among the seven participants, two individuals

highlighted the paramount importance of accuracy within their work. For these practitioners, the precise capture of movement nuances and intricacies was fundamental to achieving their goals. The motion capture system they employed was chosen primarily for its ability to faithfully represent the subtleties of human movement, ensuring a high-fidelity replication of their performances.

One participant was conducting research in Laban movement quality, where the motion capture data needed to be clear and of high quality to support the research objectives. The other researcher was focussed on utilising motion capture as an archival method. To ensure accuracy and reliability of the archival data, it was imperative to obtain precise representation of the performers, making accuracy a paramount consideration in both cases.

Conversely, the majority of participants acknowledged that their chosen motion capture systems might not be the most accurate available. This awareness stemmed from a pragmatic understanding of the trade-offs inherent in motion capture technology. While accuracy was recognised as a desirable attribute, participants were attuned to the broader considerations that shaped their decisions. Factors such as cost-effectiveness, portability, and adaptability to various performance settings were pivotal in determining the suitability of a motion capture system for their creative needs.

“(Accuracy) .. depends on the context.”

“I would say it (accuracy) depends on the context.”

“(Accuracy) depends on your goal.”

The participants exhibited an appreciation for the contextual nature of accuracy. They emphasized that the degree of accuracy required was contingent upon the specific demands of a given performance or outcome. The decision to prioritise accuracy depended on the intended outcomes and the artistic intent of the work. Some participants even expressed an affinity for the unique outcomes that lower accuracy systems could generate, introducing novel creative possibilities that might not be achievable with more precise technologies. Interestingly, even participants who utilised Vicon, widely considered the pinnacle of accuracy in motion capture, acknowledged that certain aspects of performance essence remained beyond its capture. The inherent complexity of capturing intricate muscle movements, facial expressions, and nuanced emotional cues posed challenges that

extended beyond the capabilities of even the most accurate and sensitive systems. This realisation highlighted the inherent limitations of technology in fully encapsulating the entirety of a performance, reinforcing the importance of a holistic understanding of accuracy within the context of performing arts.

“Like some distortion is, is kind of cool. You know, like sometimes some distortions kind of add to the creative, creative process.”

“.. when it's with facial movements, or hands movements, small movements, small motor movements, [it needs] to be pretty accurate. I guess also when you want to compare certain performances of choreography. When accuracy is not as important is when it's with like bigger movements.”

“it captures, like, skeleton position but not muscle use. So from a dance perspective, that's like, really like, it's much less interesting than if we could capture, like the use of muscles. But we're pretty far away from that.”

4.3.3 The significance of portability in motion capture systems for performing arts

While the participants predominantly discussed the topic of accuracy, portability of motion capture systems emerged as the second most prominent theme of the discussion amongst the practitioners. Unlike accuracy, which exhibited contextual relevance, portability was universally recognised as a crucial factor within performing arts. The ability to seamlessly transport and set up motion capture systems assumed paramount importance, contributing to performance efficiency, adaptability, and success. In the landscape of performing arts, the imperative of portability was uniformly acknowledged by all participants using motion capture for performance. This recognition was driven by the inherent dynamics of performing arts, where the ability to deploy motion capture systems in diverse settings swiftly was essential. The capacity to travel with the technology and establish a functional setup with minimal delays played a pivotal role in facilitating the integration of motion capture into performances. Evidently, high-end optical motion capture systems, known for their exceptional accuracy, were not inherently suited to the demands of performing arts venues. The impracticality of moving these systems and their associated technical intricacies led to an inherent acceptance of a trade-off between system accuracy and portability. As a result, some practitioners opted for sensor-based motion capture systems, allowing them to attach sensors to performers’ bodies, sometimes even beneath costumes, enabling motion detection without sacrificing the mobility and portability required for performances.

“it absolutely has to be portable, if it's not portable, if there's a Vicon system that has to be bolted in, you know, completely calibrated by someone, then it's just off the table.”

“It takes a bit to … set up. Sometimes it’s a bit annoying.”

“So it took time to set it up. But didn't take a crazy amount of time. You got to have someone sort of run around on the stage for a bit while you watched it, tweak some numbers.”

“The reason that we have predominantly stuck with the sensor based inertial one is that it allows the performance site to change so that you're not locked to being in a studio with all the cameras because that's a boring performance space.”

Beyond portability, the importance of set-up time emerged as an integral facet of motion capture integration in performing arts. The time constraints inherent in live performances, not including preparation time or rehearsing, necessitated swift and efficient system installation. Participants emphasised that the ease and speed of setting up a motion capture system directly influenced the practicality of its use within performance contexts.

“The repeated setup, eats into the rehearsal time and the tech is always the, the barrier, or it's always the thing that eats up the most time when we're working.”

The correlation between portability and set-up time further underscored the need for motion capture systems that strike a delicate balance between technical precision and logistical convenience. In addition to the technical and logistical considerations, the issue of data management and equipment proliferation was also discussed. Highly accurate systems will generate vast amounts of data which, while beneficial for analytical purposes, can prove cumbersome in performance settings. The quest for optimal portability often entailed a conscious effort to streamline equipment and data management to ensure a seamless performance.

“.. end up with this mass of MoCap data. And it's a lot of data to handle. And also, if you want to make the data legible, you have to do a lot of editing and processing. And yeah, you have to be very clear about what you want from that. So it might actually be too much”

“.. bandwidth of the data they produce and passing that around, it's like, quite restrictive, or you just need to think a lot about it, and manage that data. And that might be like, using multiple computers, or really powerful computers.”

4.3.4 The hardware limitations of motion capture systems

The adoption and integration of motion capture systems in performing arts, while offering a broad realm of creative possibilities, does bring inherent limitations. Understanding these limitations is vital for practitioners to effectively harness the potential of their chosen systems and produce effective performances. Additionally, an innate understanding of the flaws and limitations of these systems could indicate areas where machine learning motion capture systems could improve existing workflows. Throughout the interviews, participants illuminated various constraints that underscored the landscape of motion capture technology within the context of performing arts.

Lighting conditions and camera sensitivity

The impact of lighting conditions on motion capture system performance emerged as a significant limitation, particularly with 3D cameras and optical flow technology. These camera systems were found to be sensitive to specific lighting setups, potentially leading to inconsistencies or inaccuracies in motion detection. The interplay between lighting and camera sensitivity necessitated thorough consideration of lighting design to ensure optimal system functionality.

Camera resolution and distance impact

Motion capture system accuracy was observed to be influenced by camera resolution and distance from performers. When combined with increased distances, lower camera resolutions could lead to erroneous detection and compromised accuracy. This limitation necessitated a careful balance between system setup and performance space to optimise data capture.

Battery management

The reliance on battery-powered components introduced a potential limitation in motion capture systems. Inadequate battery maintenance and charging practices could lead to unexpected interruptions during performances, affecting data capture and overall system functionality.

Sensitivity to magnetic fields

Magnetic interference emanating from electronic equipment, metal structures, or nearby magnetic fields was cited as a potential challenge. Participants noted that these magnetic interferences could disrupt the tracking processes of some motion capture systems, resulting in inaccuracies or unexpected system behaviour. This sensitivity to magnetic fields

introduced an additional layer of complexity in the setup and execution of performances, necessitating a comprehensive understanding of the performance space and its potential magnetic influences.

Firmware updates and calibration interruptions

The unexpected impact of system firmware updates and recalibrations on system stability and functionality was acknowledged as a limitation. Participants highlighted instances where system settings were reset following updates or recalibrations, underscoring the need for careful system management and preparation to avoid disruptions during performances.

Calibration and accuracy degradation

The need for regular system calibration to maintain accuracy was recognised as a limitation for some motion capture systems such as Vicon. Failure to recalibrate could result in accuracy degradation over the course of a performance. Practitioners emphasised the significance of ongoing calibration to ensure consistent and reliable data capture.

The interviews indicated that every motion capture system has its own limitations, each requiring careful consideration and strategic management. The limitations outlined above highlight the multifaceted challenges that practitioners navigate as they employ motion capture technology to enhance their artistic expressions within the domain of performing arts. An awareness of these limitations, coupled with creative problem-solving and technical adaptability, empowers practitioners to optimise their chosen systems and maximise the potential for producing captivating and technically proficient performances.

4.3.5 The cost factors in motion capture systems in performing arts

The role of cost in motion capture system adoption within performing arts emerged as a pervasive and influential factor which was mentioned by all participants. Participants indicated the profound impact of budget constraints on their decision-making processes. The perspectives shared by these practitioners underscored the financial realities that shape the accessibility, experimentation, and potential of motion capture systems in the context of artistic endeavours. One participant aptly captured the financial challenge, remarking that “you just don’t have the budgets that are required.” The sentiment was echoed by others who reiterated that cost is a formidable consideration for most practitioners in the performing arts arena. The financial limitations were highlighted as an intrinsic aspect of the field, where artists often operated without significant financial backing. One participant noted, “We’re

artists, you know, we're not making huge amounts of money, we're not funded or have big corporate partnerships."

The notion of affordability and accessibility was expressed as a driving force for practitioners. The aspiration to engage with motion capture systems was counterbalanced by the desire to avoid investing in elaborate and expensive equipment suites. A participant emphasized this sentiment by stating, "the idea [is] that you don't have to purchase like a big suite of expensive equipment… for somebody who doesn't have a budget", when talking about the Vicon motion capture system. The juxtaposition between high-end systems costing hundreds of thousands of dollars and more accessible options priced within the reach of the individual was a recurring theme as was the desire for democratisation through affordability. Participants expressed the hope that lower costs could catalyse broader experimentation, collaboration, and artistic exploration. The sentiment that "it would be great if the systems were cheaper so that more people could experiment with them" underscored the potential transformative impact of reducing financial barriers to entry within the motion capture landscape.

4.3.6 The use of real-time systems in performances and monitoring movement

The notion of real-time functionality resonated prominently throughout the discussions. Participants described scenarios where real-time motion capture in the form of animated graphics was projected onto screens during performances, creating an immersive and synchronised fusion of movement and visuals. The ability to capture performers in one space and transmit data in real-time to another location opened up exciting possibilities for remote collaboration and cross-location performances. In interactive performances, real-time motion capture proved integral. The ability to generate immediate responses and visualisation based on performers' movements enabled dynamic synergy between technology and artistry. The utilisation of real-time visual feedback allowed performers to observe their movements from diverse perspectives, enhancing their self-awareness and contributing to the evolution of their performances. Participants acknowledged the value of visual feedback for performers to gauge the impact of their movements on the accompanying graphics or visual elements. Monitors strategically placed on stage provided performers with live feeds, enabling them to adjust their movements based on real-time observations.

".. capturing performers in another space and being able to process that, or present it in real-time, in a completely different location is really cool."

".. it's useful for them (performers) to see themselves from different perspectives, physically, like that's a really big bonus that they sometimes look like, they can perform a routine and then they can watch it from the top view or top down or something."

However, the concept of real-time interaction was not without its challenges. Latency, or the delay between movements and system response, was a concern. Despite the technical complexities of processing motion data in real-time, performers often expected a near-instantaneous response. The disconnect between expectation and actual latency could lead to frustration.

"… latency can be quite high. And it's not what the performers expect. They sort of want it to be a lot more still real, real-time considering all the processing that's going on."

While the desire for real-time interaction was considered, the reality of achieving it varied among practitioners. The recognition that not all instances necessitated real-time usage was also articulated, indicating that the mere presence of motion capture did not always dictate a real-time execution. There was recognition that the choice between real-time and pre-recorded systems was context-dependent. Artistic intent, technical considerations, and the desired impact on the performance often guided the decision to deploy one approach over the other.

".. doesn't mean that we've used MoCap in real-time, every time."

"(The animation was) very large size projected on screens. But these were pre-recorded."

4.3.7 The glitch artifacts found in the output of machine learning-based motion capture

In interviews with practitioners, an openness toward glitch artifacts was evident. Rather than requiring fully accurate motion representations, there was an interest in how machine learning-based motion capture models perceived and interpreted a performers movement, even if the output contained glitches. One interviewee mentioned that they enjoy the unexpected nature of pose detection models, and another remarked, "I love glitch art, and I love everything about the idea of that you learn from the glitch, I find that endlessly fascinating." Driven by practitioners' receptiveness and my own curiosity to control the glitch aesthetic, further investigations were undertaken to intentionally produce and manipulate glitches from machine learning-based motion capture models. The goal was to explore possibilities for exerting artistic control over these unintended artifacts.

As noted by prior work, these visual artifacts manifest for diverse technical reasons, including incorrect camera calibration, code defects, or user errors during processing. While initially deemed problematic, the creative potential in these glitches was recognised. As numerous artists have demonstrated, glitches provide valuable opportunities for experimentation and inspiration (Art et al., 2021; Betancourt, 2016; Cascone, 2001; Kane, 2019; Wallace & Martin, 2022). The glitch can be described as "... that which creates minor disturbances without actually damaging its major functioning. Glitches do not stop transmission: they merely make it scrappy, dirty or noisy" (Cubitt, 2017).

It is important to distinguish outright failures, where a model cannot run, from unexpected outputs where the results are simply surprising or unconventional. These rare occurrences may stem from errors in code, quirks of the AI model, or user mistakes. The original model authors likely encountered similar outputs but suppressed them as bugs, rather than sources of novelty. For artists, however, these accidental anomalies can be exciting revelations from otherwise deterministic systems. Though unintentional, they inspire new creative directions. By embracing glitches as more than just errors, artists find paths to innovation that conformist AI would never produce.

Examples of glitches

In testing motion capture machine learning systems, incorrect camera calibration often caused glitchy animations. Camera calibration is key for determining each camera's position in 3D space. With bad estimates of camera locations, the resulting animation floats unpredictably with a ghostly quality. The animated figure seems to take on a life of its own, drifting eerily around the subject. It sometimes mirrors the subject's motions briefly before floating away on a random path. While unintended, we found artistic merit in these faulty animations untethered from the original performance. The visual artifacts suggest supernatural themes of spirits detached from physical forms. Errors in machine learning can unlock alternative aesthetic possibilities beyond the system's intended output. Flaws in calibration spawn new directions for expression compared to error-free results. A video example is available here [vidStream_A]. In Figure 4.2, the image on the left displays the subject employed for motion capture, while on the right, the glitched output unveils a ghostly animation.

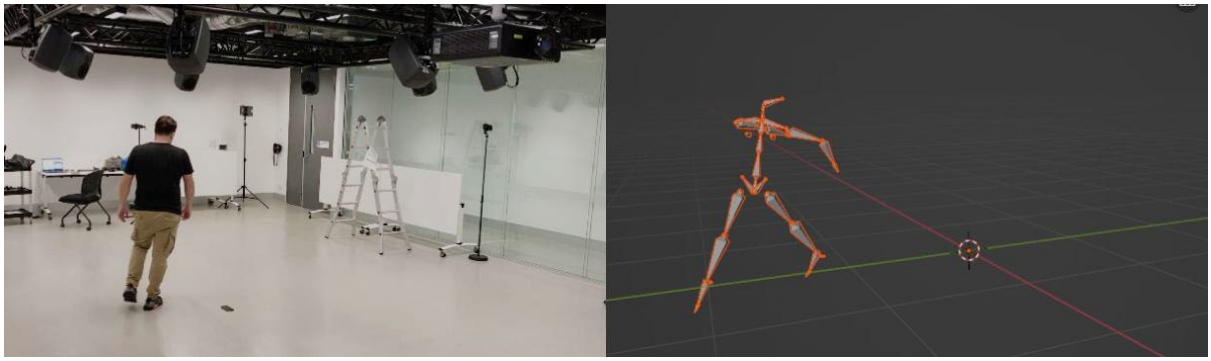

Fig. 4.2: Glitch compared to input video.

Recognizing the artistic potential of these glitched animations, experiments were conducted to intentionally induce calibration errors. This allowed the nature and degree of the resulting artifacts to be controlled. Through informed trial-and-error, the camera calibration parameters were varied and the impacts on the animated mesh observed. It was found that using more incorrect calibration data caused the mesh to become increasingly detached and ghostly. The more the calibration was flawed, the more the figure drifted aimlessly rather than matching the true subject. This method allowed the exploration of a range of glitch intensities and aesthetics. Artistically fruitful systems flaws can be intentionally triggered given proper understanding of a model's failure modes. A video example of the glitches merged with the correct pose detection output is available here [vidStream_B]. In Figure 4.3, the image on the left features the subject employed for motion capture, while on the right, glitches in the skeleton are generated by employing different degrees of inaccurate camera calibration data.

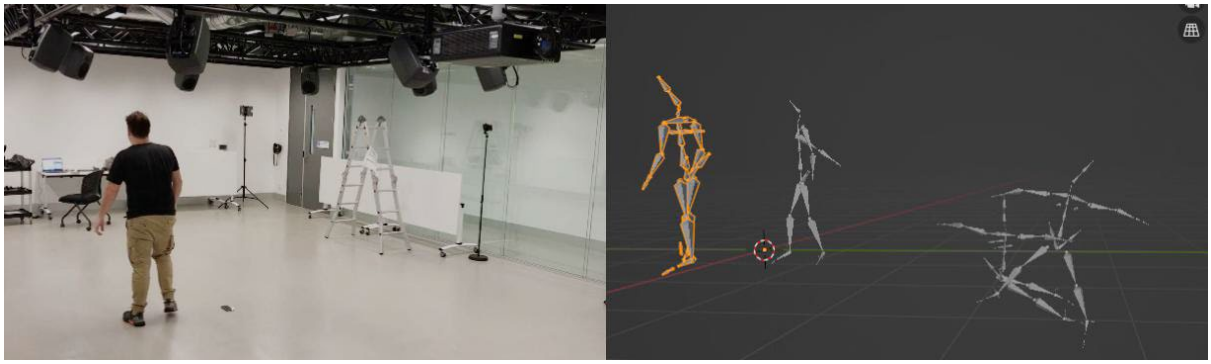

Fig. 4.3: Various glitches compared to input video.

Controlling the glitch

There is considerable creative potential in embracing the glitches and unexpected behaviours that can arise in machine learning models for motion capture. However, model creators rarely provide guidance on inducing or exploiting these glitch effects. There is merit in being transparent about the failure modes and boundary conditions where machine learning systems begin to exhibit anomalous outputs. Just as video game creators include hidden cheat codes to enable new modes of play, model developers could intentionally reveal areas of fragility where their systems glitch in artistically generative ways. This aligns with calls for algorithmic transparency and accountability. But beyond ethics, there are pragmatic benefits as well. Artists are innately experimental - by showing them where the edges of stability lie, model builders empower new directions for innovation and expression. Of course, glitches are often unintentional and difficult to predict. But even disclosing the types of errors that occurred during development would help users shape their approach. Creative exploration thrives on uncovering new possibilities. By pointing artists to the zones of unpredictability in machine learning, researchers can foster more spontaneous discovery and novel aesthetics. What initially appear as flaws can become wellsprings of creativity given the right mindset and information (Knight et al., 2023). A video example of where animation was applied to a glitched output of a machine learning model is available here [vidStream_C]. Figure 4.4 displays a single frame from an animation generated using the distorted output of a machine learning model.

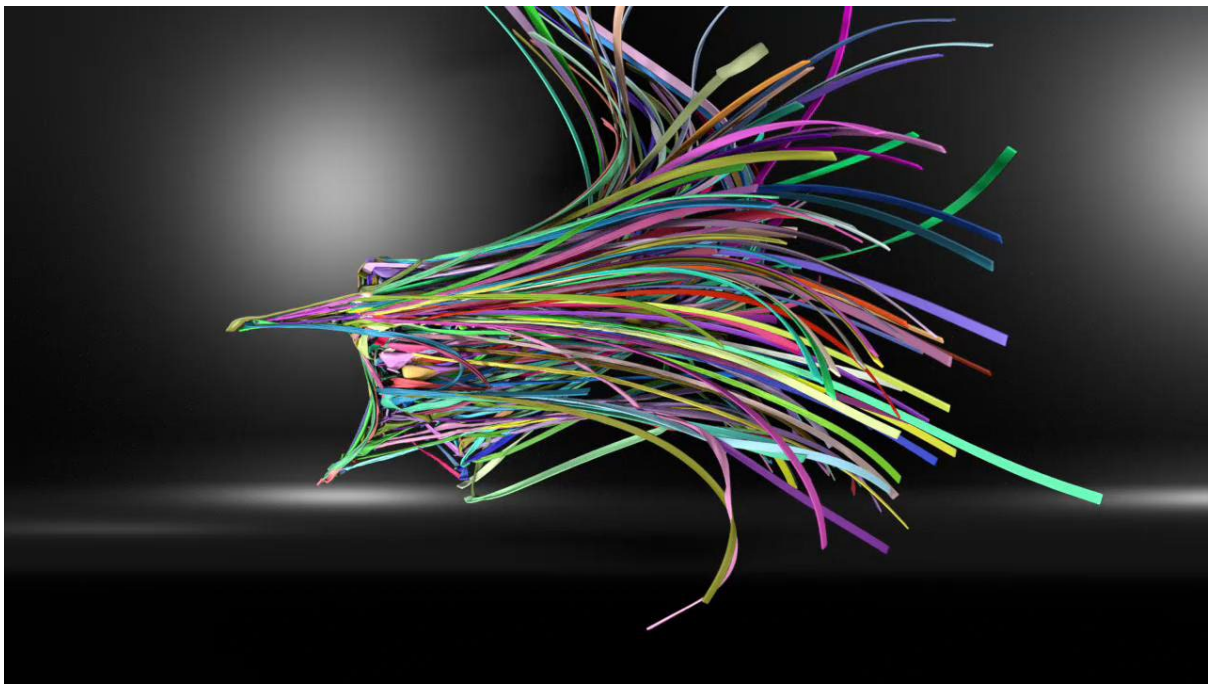

Fig. 4.4: Animation applied to glitched data.

4.3.8 The potential of machine learning systems in motion capture for performing arts

Among the participants, a subset of individuals had direct experience utilising machine learning models for motion capture. These practitioners embarked on an exploration that deviated from traditional marker-based approaches, opting for markerless solutions facilitated by machine learning technologies. Incorporating machine learning methods, particularly utilising the VIBE model, yielded a range of insights and reflections, encapsulating promise and limitations. In the words of one participant, the experience with markerless motion capture through machine learning was “really promising.” The departure from marker-based methodologies was recognised as a pivotal step toward innovation and experimentation within the area of motion capture. This transition resonated with the aspiration to engage in conversations that bridge technology and artistic outputs in novel and intriguing ways. The endeavour was described as multifaceted, characterised by the emergence of a new interaction between the machine and the resulting outputs. The interactive collaboration was perceived as both engaging and enjoyable. The interpretation of the animation against the performer was found to be exciting and fun.

This creative partnership between machines and artists highlighted the profound potential for artistic exploration and aesthetic discovery.

“.. having all this really interesting new offers and conversations between the machine and, and what the kind of outputs were. And it was really like a fun, amazing process watching those digital bodies kind of cope with everything we gave. So like I found that really like enriching and exciting and humorous and, like joyful”

The participants closely examined how the animated mesh interpreted and represented the human body, highlighting the interaction between the animated mesh and the specific characteristics that define the human form. A critical part of their engagement was reviewing the processed clips alongside the original video footage, allowing them to directly compare the original and processed versions. This visual comparison enabled the practitioners to see what aspects of the movement were accurately captured by the technology and what was missed or inaccurately represented. This process gave them better insight into the current capabilities and limitations of the motion capture technology. However, limitations were acknowledged as well. The absence of real-time processing capability was noted as a primary constraint when outputting complex animated shapes, particularly within the context of live performances. Despite this, the practitioners appreciated the process and its outcomes, highlighting the integration of the processed mesh into their performances.

4.4 Discussion

In this section, we delve into a discussion that builds upon the insights garnered from the results section. The findings serve as the foundational framework for the ensuing discourse. The discussion aims to provide a deeper understanding of the implications, challenges, and potentials of integrating machine learning methods within the landscape of performing arts. Drawing from the insights provided by the participants, we examine key themes that emerged from the interviews. These themes encompass a range of critical aspects, including accuracy considerations, portability factors, hardware limitations, cost implications, and the dynamic between real-time motion capture systems. By extrapolating from these themes, we aim to uncover the nuanced relationship between the participants' experiences, needs, and aspirations and the capabilities and limitations presented by machine learning-based motion capture systems. These discussion points address the research question: 'In performing arts, what are the characteristics creative practitioners look for in motion capture systems?' (RQ1)

Motion capture system selection

The examination of the motion capture system selection of practitioners in the context of performing arts reveals a multifaceted process interwoven with various considerations. The choice of motion capture systems is not a one-size-fits-all decision but rather one that emerges due to the complex interplay between contextual factors, technical expertise, and artistic intentions. In this context, the role of machine learning methods in motion capture system selection becomes an intriguing aspect that warrants discussion. Integrated with motion capture systems, machine learning methods hold a significant place in the selection process. These methods can offer enhanced capabilities for capturing movements and generating outputs contributing to the overall performance experience. However, the decision to employ machine learning methods is closely intertwined with the intended outcomes of the performance and the feasibility of integrating these methods.

It is noteworthy that the presence of technical expertise also plays a pivotal role in motion capture system selection. In some cases, the availability of individuals with expertise in a particular system heavily influences the choice. This observation also extends to machine learning methods, where practitioners might opt for systems with machine learning capabilities if someone on the team possesses the requisite technical know-how. Machine learning models, while requiring a level of technical proficiency, are accessible to a broad

range of practitioners beyond computer scientists. This democratisation of access is gradually becoming a defining characteristic of modern technological advancements, allowing artists, performers, and choreographers to harness the capabilities of machine learning models to enhance their performances. The potential integration of machine learning methods is not solely restricted to specialised domains.

Accuracy in traditional motion capture systems compared to machine learning methods

Participants' discussion of accuracy with motion capture systems revealed that while accuracy is a consideration, its relevance is inherently contextual. The sentiment among participants suggests that their goals are often not centred on attaining the highest level of accuracy but rather on achieving a level of accuracy that suits the creative demands of their chosen performance medium. It is important to acknowledge that the discussion on accuracy inherently entails a spectrum of possibilities. While machine learning-based methods might not currently match the detection capabilities of high-end systems like Vicon, they have the potential as an attractive alternative in terms of accuracy when it comes to outputting skeletons and meshes for the creative process. For machine learning models to be practical in representing artistic expressions in the performing arts, they need to align with the varied goals of performers and artists. Additionally, the choice between using monocular or multiple cameras adds complexity. Whether a monocular or multi-camera setup is chosen affects the level of accuracy achieved. The intended creative outcome and specific needs of the performance drive which setup is selected. The accuracy of the machine learning models should be viewed as just one factor to consider when choosing between monocular and multi-camera motion capture systems, within the broader context of the artistic goals.

The evolution of machine learning methods within the motion capture landscape invites the anticipation of continued advancement. The continual development of this technology is likely to lead to advancements in accuracy. As machine learning models mature, there is an expectation that the accuracy gap between established systems and emerging technologies will gradually narrow.

Portability of motion capture systems compared to machine learning methods

Portability is a key requirement for motion capture systems in the performing arts. Modern machine learning models are well-suited for this because they can be built from and work effectively with data from low-cost, consumer-grade video sources like mobile phone cameras. The practitioners emphasised the importance of portability as a central factor when choosing and using motion capture technology for performances. Portability allows the equipment to be easily transported and set up in different performance environments, which is crucial for achieving the desired artistic outcomes.

Machine learning models align well with this need for portability. They can rely on accessible cameras like smartphones or compact devices on tripods, which are easy to transport and set up. This minimalistic approach with unobtrusive equipment integrates seamlessly into performing arts settings. Monocular systems are particularly portable, with streamlined setups using minimal equipment that could potentially be operated handheld. The widespread availability of camera-equipped devices among most people also adds to the appeal of machine learning approaches.

Hardware limitations of motion capture systems compared to machine learning methods

Traditional motion capture systems can be subject to a range of hardware limitations. These include lighting conditions, camera resolution, battery management, sensitivity to magnetic fields, firmware updates, and calibration intricacies. In contrast, machine learning motion capture methods, while not entirely exempt from hardware limitations, demonstrate distinct advantages. These methods exhibit a low-maintenance nature, minimising hardware limitations. For instance, machine learning models have shown promise in operating effectively under uncontrolled or natural light conditions, mitigating the constraints posed by challenging lighting setups. The flexibility to adjust video capture resolution caters to expansive capture spaces, optimising accuracy and coverage.

Battery management is a concern for some systems, including machine learning systems, because they use devices that need to be powered. This issue can be addressed by supplying the devices with power connected to an outlet, reducing the risk of disruption for long capture periods. Moreover, machine learning models are not sensitive to magnetic

interference, providing a significant advantage over traditional systems such as sensor-based motion capture systems that grapple with this limitation.

Cost factors of motion capture systems compared to machine learning methods

The economic constraints inherent in motion capture systems for performing arts are evident as emphasised by the participants. Budget limitations significantly influence decision-making within the industry. In this context, machine learning methods present a pragmatic alternative, aligning cost-efficiency with the creative objectives of practitioners. Machine learning methods offer distinct advantages by sidestepping the need for costly equipment. Notably, their hardware requirements are minimal, often relying on common devices such as mobile phone cameras. For enhanced accuracy, multi-camera systems demand only a few accessible devices with modest resolutions situated on basic tripods. The cameras do not need to be specialized, as some motion capture systems necessitate, providing practitioners with the flexibility to employ various devices. The integration of reasonably priced action cameras, known for their portability, offers a further cost-effective option, making machine learning methods viable.

Beyond hardware considerations, machine learning methods reduce costs by eliminating the need for specialized marker-laden suits. The reliance on body shape in machine learning systems negates the expense associated with marker-equipped attire. This practicality would resonate with budget-conscious practitioners seeking efficient yet cost-effective solutions.

The use of real-time motion capture systems compared to non-real-time machine learning methods

Integrating real-time motion capture systems has become a pivotal aspect within the performing arts domain, enabling dynamic performances and facilitating immediate feedback from performers. The participants' narratives highlight the extensive use of real-time systems to enhance performances and monitor real-time movement. Notably, some practitioners adopt a hybrid approach, combining live motion capture during performances with pre-recorded animation, blending real-time immediacy with meticulously crafted animations. In contrast, machine learning methods exhibit a current limitation in real-time functionality. While real-time performance is achievable with a potent graphics card, it is accompanied by limitations. The processing demands of generating detailed animations impede the real-time output that practitioners seek. However, the participants' insights highlight awareness of

these constraints, reflecting an understanding of the current capabilities of machine learning methods.

Despite the current limitations, participants indicate that they work within the confines of technology's capabilities. This acknowledgement accentuates the adaptability of practitioners, who navigate the existing limitations to create performances that align with the available technological frameworks. There is prevailing awareness that machine learning models are still evolving and their potential will likely expand.

Glitches in the output of machine learning-based motion capture systems

Through both accidental discoveries and intentional interventions, the artistic potential in these system flaws was analysed. Experiments with erroneous camera calibration demonstrated how even basic technical errors can unlock alternative aesthetic possibilities. By embracing glitches as more than just mistakes, new directions for creativity and expression can emerge. However, fully leveraging this potential requires transparency from model developers. It is argued for revealing areas of fragility where machine learning systems begin to glitch in generative ways. Enabling artists to probe the edges of unpredictability fosters innovation. Of course, many glitches evade prediction or control, but disclosing common failure modes would empower more intentional creative exploration. Flaws need not be seen as merely problematic - with the right approach, they become wells of originality.

Machine learning methods in performing arts

The affirmative feedback voiced by these practitioners highlights the potential viability of integrating machine learning motion capture methods into performing arts. The application of these methods in a real-world production context sheds light on the tangible benefits of this innovative approach. The practitioners' exploration serves as a testament to the potential of machine learning models to enrich and augment the creative process within performing arts.

4.5 Conclusion

This exploration into the experiences of motion capture practitioners within performing arts reveals promising potential for machine learning methods to emerge as a viable option in

this domain. While traditional systems continue to play a role, the interviews highlighted key areas where machine learning models hold distinct advantages in addressing the needs of practitioners. Factors like portability and cost were frequently cited as major considerations that shaped motion capture adoption. Machine learning methods align closely with these priorities through minimal equipment requirements, with modern approaches performing effectively even when using consumer-grade mobile phone footage as input.

The capacity to transport and swiftly set up accessible cameras offers practitioners the mobility their performances demand. Additionally, negating the need for specialised equipment or attire, means that contemporary machine learning approaches provide a pragmatic solution for budget-conscious artists. Beyond portability and affordability, modern machine learning models demonstrate capabilities in overcoming limitations like lighting and space constraints, which were flagged as key issues for existing motion capture setups by interview participants. There was an openness toward glitch artifacts. Driven by this receptiveness, intentional glitch production and manipulation triggered by machine learning could represent a new avenue for creative expression. Finally, participant feedback suggests that improved machine learning performance in real-time capture and display environments would open up a greater range of production use cases.

This exploration reveals optimistic indicators that machine learning-based motion capture can serve as a viable option for performing artists seeking portable and affordable systems. These methods can potentially become a transformative technology that grants practitioners access to capabilities for producing creative performances unfettered by physical and financial constraints. As machine learning methods progress, their integration promises to reshape possibilities at the intersection of technology and artistic expression.

5 Experiment: Monocular Pose Detection for Performance

5.1 Introduction

In this chapter, the results of the investigation into the utilisation of a monocular machine learning-based motion capture system for performing arts are presented. This research aims to explore the feasibility and potential applications of machine learning-based motion capture methods in performing arts, specifically in capturing and analysing human motion. The investigation into a monocular machine learning motion capture system directly ties into the research question guiding this project. The overarching goal is to determine if machine learning methods can enhance or improve current motion capture techniques in performing arts contexts. Monocular systems require only a single camera, so equipment costs and setup complexity are reduced compared to traditional multi-camera rigs. The accessibility, affordability and portability of monocular machine learning solutions promises to address key barriers that may limit the adoption of motion capture for performers and choreographers. This study explores whether such a promise is corroborated through practical testing and use of a modern monocular machine learning system.

The primary aims of this study are to examine:

Feasibility of machine learning for performance motion capture:

The investigation will ascertain whether machine learning approaches could successfully capture and analyse human motion in a performing arts context. The workflow and outputs of the model will also be assessed for their relevance in performing arts.

Integration of monocular detection for accessibility and portability:

The goal was to investigate the advantages of monocular machine learning-based motion capture, specifically if it provides a cost-effective and accessible solution to artists, performers, choreographers, and directors.

In the following sections, the details of the machine learning model are discussed in relation to testing the model on a dancer's movement. Additionally, collaborative efforts with choreographers and directors are discussed. Their artistic expertise played a vital role in assessing the feasibility of using the model in a production setting.

Figure 5.1 below shows how research question two will be addressed through practical implementation and testing of the VIBE machine learning-based motion capture model with a performer.

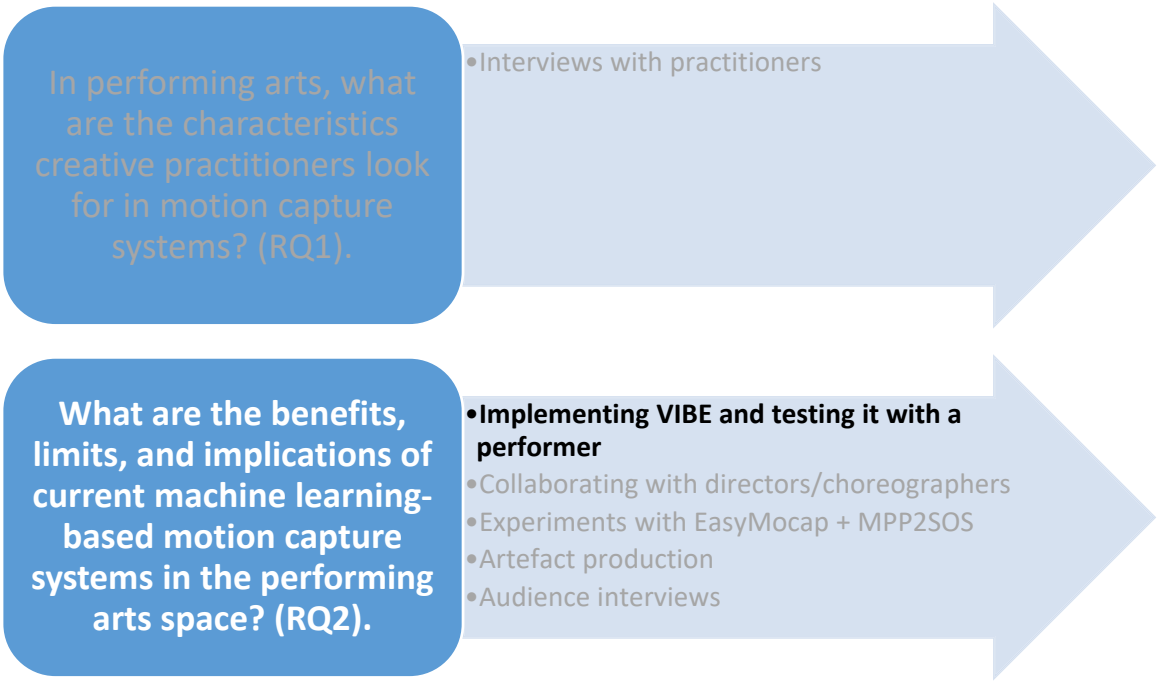

Fig. 5.1 Implementing VIBE and testing it with a performer to address research question two.

5.2 VIBE

5.2.1 Motivations for model selection

After examining and running a number of monocular machine learning-based motion capture models, VIBE (Video inference for Human Pose and Shape Detection) (Kocabas et al., 2020) was selected. Key considerations included the ability to use consumer-grade equipment, portability, ease of setup, relatively low cost, and minimal hardware requirements. The VIBE model satisfies these criteria, making it a suitable choice for such applications.

VIBE leverages a large-scale 3D motion capture dataset called AMASS (Mahmood et al., 2019), which has dance movements in its dataset used for training. This encourages the

model to generate human motion that closely resemble real-world movements, making it more suitable for capturing performing arts motions authentically. The model uses video sequences to train, benefitting from temporal information, allowing it to capture motion dynamics and continuity (Goodfellow et al., 2014). This is particularly advantageous in performing arts, where smooth and coherent motion can be crucial for artistic expression. Equally importantly, VIBE is trained with footage captured in a variety of environments (including indoors and outdoors) and with a variety of video devices, increasing the likelihood that the model will perform well outside of tightly controlled Vicon stages, increasing the number of use cases for performance. This diversity of the dataset helps the model to learn from various real-world human movements across diverse contexts, including those that might not be present in indoor 3D datasets used in other models. The abundance of in-the-wild data enriches the model's understanding of human poses and motions in a performance art context (Kocabas et al., 2020).

Another benefit that the model brings is that VIBE is designed to detect and estimate poses for multiple people in a single video sequence. This capability is especially valuable for performing arts scenarios that involve group performances or choreographed interactions between multiple performers (Kocabas et al., 2020). VIBE is optimised to run on both CPU (processor) and GPU (graphics card) hardware. This flexibility ensures that users with various computational resources can benefit from the motion capture capabilities of the model. VIBE has fast inference speeds, allowing for near real-time processing depending on the GPU used. On powerful GPUs, such as the RTX2080Ti, the model can achieve up to 30 frames per second (fps) (Kocabas et al., 2020). These speed measurements do not account for the time required to film the subject, transfer the recorded footage to a computer, generate the 3D mesh output, or apply the animation data to the mesh.

VIBE has demonstrated state-of-the-art results on popular 3D human pose estimation datasets (Kocabas et al., 2020). Its accuracy and robustness across diverse and partially occluded poses make it a suitable candidate for performance art applications. A distinctive feature of the model is its ability to output the animated human mesh in .fbx file format. This output format is compatible with popular 3D animation software used in the entertainment industry. Artists and animators can easily import the .fbx files into their preferred animation tools for further editing and refinement for their artistic projects (Kocabas et al., 2020).

An advantage of the VIBE monocular machine learning system is that it is markerless and does not require specialised motion capture suits. The model relies solely on standard RGB video footage as input, rather than physical markers placed on the performer's body or reflective dots on tight-fitting suits. This markerless approach liberates the performer to move freely without wearing any attachments or constricting outfits.

5.2.2 Setup and data collection

After selecting a monocular pose detection model, the objective was to investigate its utilisation and performance with a performer. The focus of this experiment was directed towards employing accessible hardware, such as utilising the camera on a smartphone, and operating in sub-optimal space and lighting conditions. The purpose of the capture process is to generate an abstract animation that mirrors the contours of the dancer, incorporating a motion trail effect. This abstract animation will subsequently be applied to the mesh produced by the model, resulting in the animation depicted in Figure 5.3 below.

Due to the limitations imposed by the COVID-19 pandemic, the performer could not be recorded in person. Instead, the performer engaged in a remote collaboration, capturing a video of her dance choreography in her living room. This video footage served as the primary dataset for evaluating the VIBE monocular model. The choice of the living room as the capture environment introduced variations in lighting conditions, challenging the model in an environment with uneven and dynamic lighting. The presence of a bright window intensified lighting at one end of the room, gradually diminishing towards the other end, resulting in notable variations in illumination levels. The video recording utilized the built-in camera of an iPhone 8, with no reliance on additional camera equipment such as high-resolution cameras or professional-grade lenses.

Given that the camera was handheld during video capture, there were subtle shifts in the camera position throughout the performance, leading to minor changes in perspective and framing. The deliberate choice not to re-shoot was strategic, aiming to assess the robustness of pose detection within the model under sub-optimal capture conditions, where lighting, framing, and device capabilities are of lower quality than typically encountered in professional performance settings. This strategic test allows for an exploration of the extent to which ML-based motion capture expands the range of settings viable for capture,

addressing known constraints related to cost, environment, and flexibility as highlighted by interviewees. It provides a valuable examination of the model's adaptability to less-than-ideal conditions, thereby contributing insights into its potential applications and effectiveness in diverse capture environments.

Lab setting vs. real-world environment

Some motion capture datasets are recorded in a controlled lab setting, where the performer is placed in an empty room with minimal background distractions (Ionescu et al., 2014, (Tsuchida et al., 2019). This controlled environment allows detection algorithms to focus solely on detecting the keypoints of the human subject without interference from any background shapes or objects. In contrast, this experiment involved capturing video footage in a real-world uncontrolled environment. As a result, the background of the capture area was cluttered with furniture and miscellaneous items, presenting a more complex and challenging scenario for the motion capture model. The cluttered background in the video acquisition serves as a test of the VIBE model’s robustness. If the model can successfully detect the human pose amidst the background clutter, it indicates its ability to handle real-world scenarios effectively. Robustness in cluttered environments is crucial for practical applications of motion capture in performing arts, where the capture space may contain various objects or scene complexities. By successfully detecting human poses in the presence of clutter, the VIBE model showcases its ability to focus on the key elements of motion – the performer's body – despite potential distractions from the surroundings.

Machine learning methods have the potential to streamline the motion capture process, such as by reducing the need for highly controlled environments, and/or broaden the range of spaces where data can be collected. For instance, this could enable capturing motion data in environments not typically used for motion capture, like outdoor settings or home environments.

Video resolution and capture space

The video footage captured for this experiment was recorded at a resolution of 1280 x 720 pixels, commonly known as HD720 (*4K Support on Smartphones - Playback, Recording & Display*, n.d.), at 30 fps (frames per second). An increase in video resolution to 1920 x 1080 (HD) or beyond, such as 4K resolution (3840 x 2160 pixels), could offer advantages in

motion capture. Higher-resolution videos provide greater clarity and detail, making the dancer sharper in the frame. Despite the potential benefits of higher resolutions, the HD720 resolution used in this experiment proved adequate for effectively capturing the performer's movements. The dancer remained clearly visible in the frame, enabling successful keypoint detection and motion analysis. The capture space was limited to the performer's living room size but was ample for this experiment. The area covered was about 3 x 4 meters. However, if the capture space needed to be larger to accommodate more extensive performances or choreography, adjustments to the camera setup would be required. Expanding the capture space would necessitate situating the camera further away from the performer to encompass the increased area. Alternatively, the use of a wide lens could provide a broader field of view without increasing the camera distance. However, both options may have implications for the recorded video and detection accuracy. A larger capture space or increased camera distance could potentially result in the dancer appearing smaller in the frame. The visible reduction in the dancer's size may present challenges for pose detection algorithms, introducing the possibility that the relative proportions and finer details of the performer's movements could become less discernible. However, it is anticipated that, despite this reduction in size, the performance of the model will remain reasonably robust. A higher video resolution could provide more detail at the expense of the model running slower due to the increased data.

Camera calibration not needed

Some machine learning-based motion capture systems, particularly multi-camera systems, require camera calibration (mentioned in Chapter 6.2). The VIBE model operates without relying on camera calibration, simplifying the setup process and making motion capture data acquisition more accessible to users without expertise in the camera calibration process. This user-friendly and time-saving aspect broadens the use of motion capture technology to a wide range of users, including artists, performers, and directors who may not have specialised knowledge in camera calibration.

5.2.3 Review of outputs

After the performance is recorded, the camera is connected to a computer using a USB cable to transfer the video for further processing. The video file, in .mov or .mp4 format, serves as the input for the VIBE model. The model operates on the video file directly and does not require an additional step of converting the video into individual images per frame.

The first output produced by the model is the detection of humans present in the video. Once the human detections are made, the model proceeds to estimate the poses of the detected humans in each video frame.

Conversion into a workable format

The VIBE model's pose detection output is stored in a .pkl (Python Pickle) file (Kong et al., 2020). This file contains the data for each keypoint of the subject detected in each video frame. The .pkl file records the spatial coordinates and orientation of the detected keypoints, providing a comprehensive representation of the subject's pose over time. The .pkl file is not readily usable for animation purposes in 3D applications. To import the motion capture data into 3D animation software, it needs to be converted to a .fbx file format.

The conversion process requires the user to install the Blender[39] API (application programming interface) for the Python programming language. Blender is an open-source 3D graphics and animation program that supports the .fbx file format and can import animations contained within it. Using the Blender API, the .pkl file can be processed and transformed into a .fbx file, making it compatible with various 3D animation applications. The .fbx format preserves the keyframe animation data and skeletal hierarchies, required for accurate animation representation. Once the .fbx file is imported into Blender, it contains the animated mesh of the dancer along with all the bones that connect to the joints of the mesh. Each bone is associated with keyframes, representing the spatial transformations of the joints over time. The user can manipulate these keyframes to edit and refine the animation.

Keyframe smoothing

During the post-processing stage, it was observed that some degree of smoothing is essential to enhance the quality of the amination. Without smoothing, the animation suffers from jittery and abrupt movements of the joints, leading to a less visually appealing result. By applying smoothing techniques, the animation's temporal flow becomes more natural and visually pleasing, providing a more cohesive and coherent motion. To achieve the desired level of smoothness, a filtering approach was employed to process the keyframes obtained from the VIBE model's output. Specifically, an averaging filter was applied to the keyframes,

[39] Blender, https://www.blender.org/

considering the values of the keyframes over a short time window of five frames. The temporal averaging filter computes the average value of each keyframe by considering its neighbouring keyframes over the specified time window. This process effectively reduces the influence of individual noisy or erratic keyframe values, resulting in a more stable and continuous animation. A video example of the comparison between a mesh with temporal smoothing applied is available to view here [vidStream_D]. Figure 5.2 illustrates a contrast between the initial detection (blue) and the mesh with applied temporal smoothing (grey) across the recorded video. Table 5.1 outlines the four steps of VIBE, spanning from video acquisition to the generation of a temporally smoothed mesh output.

Fig. 5.2: Initial detection and temporal smoothing comparison.

In summary, the motion capture and animation steps from acquisition to keyframe smoothing can be broken down into four main steps:

Video acquisition: The performer's dance routine is recorded using a camera, and the video is transferred to a computer for processing.

Running the VIBE model: The VIBE model is run on the video, which detects human subjects and estimates their poses in each frame.

Conversion to .fbx format: The output data from the VIBE model is converted into the .fbx file format using the Blender API, making it suitable for animation purposes.

Smoothing the animation: The animation is smoothed by applying a temporal averaging filter to the keyframes. This step refines the motion and ensures a visually pleasing and coherent animation.

Figure 5.3 below shows the steps when using the VIBE machine learning-based motion capture system.

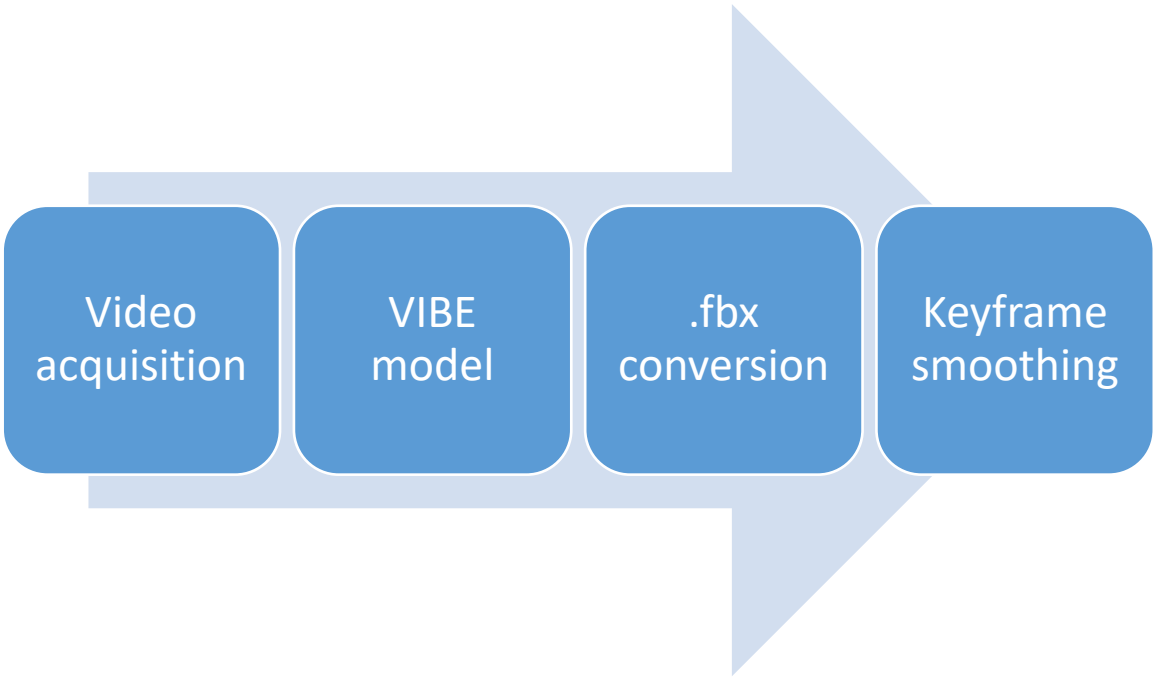

Fig. 5.3: VIBE steps.

The video used in the motion capture process consisted of 2335 frames, capturing the dancer's performance over time. During the analysis, it was observed that approximately 1.3% (30 frames) of the video resulted in no detected pose. The lack of detected pose on these frames was primarily due to the dancer's fast movements, causing the detector to miss keypoint information. However, VIBE correctly detected subtle movement, as illustrated during the early frames where the performer was essentially stationary. This observation indicates that the VIBE model effectively detects keypoints even during periods of subtle motion. During the conversion process from the .pkl data file to an animated .fbx file, frames with no detected pose had no associated animated mesh. As a result, the animated mesh gradually drifted out of sync with the actual performance, leading to a temporal misalignment.

A solution to this issue is to shift any keyframes up the timeline where the detection failed to detect a pose. This brought the animation in sync with the performance. In cases where frames lacked animation due to skipped detections, the missing frames were interpolated based on the keyframes before and after the gap. This interpolation process generated smooth transitions between adjacent frames, ensuring the animation appeared visually consistent and continuous. Figure 5.4 provides visual depictions of key stages in the VIBE model, including the unprocessed video input of the dancing subject (left), the superimposed

mesh generation of the VIBE model on the video before applying temporal smoothness (centre), and a frame featuring the abstract animation applied to the resulting mesh.

Fig. 5.4**:** VIBE stages.

Interestingly, the presence of anomalies, such as undetected frames and jittering, did not negatively impact the animation's outcome. As the animation is abstract and does not closely resemble the contours of the human form, the smoothing is unnoticeable. In traditional animation and motion capture, the pursuit of flawless precision is often paramount. However, in performing arts, the incorporation of anomalies can be intentional and purposeful. In some instances, performers collaborate with technology, allowing anomalies to emerge organically. The unpredictability and responsiveness of glitches can turn technology into a creative collaborator. As detailed in Chapter 4.3.7, which specifically examines the purposeful exploration of glitches within the artistic process, the discussion highlights situations where intentional glitching and anomalies were intentionally incorporated as transformative elements.

5.2.4 Final abstract animation

The process of creating the abstract animation from the motion capture data involved an iterative and creative exploration of various animation styles. Over the course of a few weeks, multiple experiments were conducted to find the most captivating and visually appealing representation of the captured motion.

Testing animation styles

Initial animation tests utilised a sample motion captured animation obtained from the Mixamo[40] website. Mixamo offers a diverse collection of motion captured animation data, including samples of various dance styles. Throughout the animation creation process, a diverse range of dance styles and animation combinations were tested to explore the various creative outcomes that could be achieved. The objective was to find a unique animation style that complemented the performer's movement while providing a visually captivating representation. It was decided that a trailing effect would be employed for the animation. The selection of the trailing effect is independent of the model employed to generate the animated mesh for the performance. The effect created can be applied to an animation or any mesh. This effect involved creating an elegant and flowing motion trail that followed the dancer's movement as they performed. This choice was influenced by the nature of the dancer's motion, which exuded grace and fluidity. A video example of the animation juxtaposed with the performer is available here [vidStream_E]. Figure 5.5 depicts a singular frame derived from the abstract animation produced by the VIBE model. A corresponding frame from the input video is superimposed in the bottom-left corner.

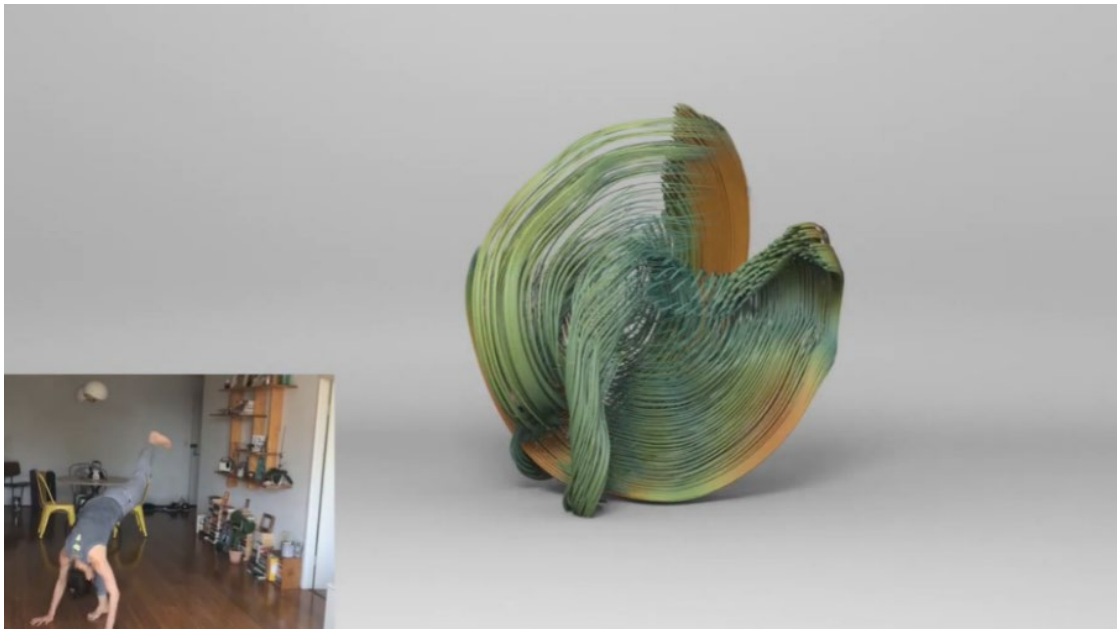

Fig. 5.5: Abstract animation with input video.

[40] Mixamo, https://www.mixamo.com/

The trailing effect

The abstract animation, derived from the motion capture data and VIBE model's output, was created using the 3D application SideFX Houdini[41]. Following a reference from the Entagma[42] website, the animation achieved the desired trailing effect, enhancing the artistic feel of the final output. To create the trailing effect, specific points on the mesh geometry were isolated. These points were strategically selected from areas of high curvature on the mesh to give the trails a more organic appearance. This deliberate selection helped avoid revealing the entire human form and contributed to the abstract nature of the animation. The duration of the trailing points was set to trail for a period of 1.5 seconds before gradually fading off. This duration was chosen to strike a balance between highlighting the dynamic motion and allowing the trails to dissipate smoothly, avoiding abrupt endings. The speed of the animation determined the colour of the trails. Areas of the mesh that exhibited fast movement were assigned a specific colour, such as orange while slower-moving sections were coloured green. This colour coding added an extra layer of visual dynamics and complexity, accentuating the varying motion speeds in the animation. The trails were further enhanced by applying subdivision techniques to their geometry, providing them with a smooth appearance. To achieve a more appealing aesthetic, the trails were edited to appear thicker at their centre and gradually taper off towards the beginning and end. This design choice added an elegant visual flourish to the animation.

Rendering the 3D scene to produce an image sequence involved substantial computational resources and took approximately 5.5 hours to complete. The machine used to render the animation had a 3.6GHz processor, 16 GB of RAM, and a GeForce GTX 1650 GPU. The final animation consisted of 2335 frames rendered at a frame rate of 30 fps, resulting in a dynamic 1-minute and 17-second animation.

5.2.5 Discussion

The experiments conducted in this study have shed light on the potential of utilising machine learning methods in a performing arts context and its integration in creating abstract animation. The results have highlighted the effectiveness of this method in producing visually expressive animation. One of the key advantages of the VIBE method is its simplicity and

[41] SideFX Houdini, https://www.sidefx.com/
[42] Entagma, https://entagma.com/

accessibility, offering a single smartphone solution and eliminating the complexities associated with marker-based systems. VIBE outputs an animated 3D mesh in the .fbx format. The mesh is connected to bones forming a child-parent relationship, each linked to a joint driven by the root (pelvis) joint. This hierarchical structure ensures that the animated mesh accurately mimics the performer's movements, capturing the essence of the performance. The .fbx file output provides artists and animators with added flexibility. Each joint and bone can be further smoothed temporally, or its animation can be precisely adjusted. This versatility empowers animators to edit or re-animate the captured movement, creating diverse animations and movements to suit the artistic vision.

A notable observation is the resilience of the VIBE method to uneven lighting conditions and a handheld camera setup. Its ability to handle these conditions by detecting the human pose (with only minor anomalies) testifies to its ability for real-world challenges. One anomaly observed in the VIBE method was the occasional failure of the pose detector to detect a human on frames where the performer was moving rapidly. Another aspect that came to light was the presence of jitter in the animated mesh, where the motion of the mesh did not precisely match the smooth movement of the dancer. This jitter effect could have led to distractions in the final animation. Keyframe smoothing can be applied to mitigate the impact of such jitters and proved sufficient in this experiment.

This initial experiment suggests that the VIBE method streamlines the motion capture workflow, offering a rapid process for generating animations. With minimal manual processing, it allows artists to focus on their creative vision rather than technical complexities. This workflow is particularly advantageous for performance settings, where artists can have a quick turnaround of motion capture data to experiment with.

The successful application of machine learning methods in this study suggests that low-cost markerless motion capture is not only feasible but also effective for performing arts scenarios. By eliminating the need for physical markers and specialised camera setups, the VIBE method liberates performers from the constraints of traditional motion capture systems, unleashing creative possibilities.

Key outcomes from this study relate to the research question: 'What are the benefits, limits, and implications of current machine learning-based motion capture systems in the

performing arts space?' (RQ2) Specifically, this initial experiment suggests that modern machine learning-based motion capture may be useful in:

- Enabling the development of motion-based expressive and abstract animation.
- Increasing motion capture accessibility by lower cost of entry for video capture hardware (e.g. by producing acceptable results through recordings from consumer-grade mobile phones)
- Expanding the range of environments in which motion capture data is recorded, including resilience to challenging capture conditions like uneven lighting and moving cameras
- Streamlining workflows, including the rapid generation and tuning of animation
- Liberating performers from constraints of physical markers worn in traditional motion capture

5.3 Collaboration with choreographers and directors

5.3.1 Introduction

Collaborations were conducted with two experienced dance and theatre practitioners – Harrison Hall and Dr Sam McGilp - to explore the potential of monocular machine learning motion capture for performing arts applications. Harrison Hall[43] is an acclaimed choreographer and performer whose work situates contemporary dance within experiential art environments. His recent projects aim to heighten embodied experience in mixed media performances by blending digital and live worlds. Harrison has toured globally with leading Australian dance companies. Dr Sam McGilp[44] is an award-winning director and artist whose diverse practice encompasses film, theatre, installation and online spaces. Their expertise spans various facets of technology-integrated dance, theatre and media arts.

Collaborating with accomplished practitioners like Harrison and Sam offers significant advantages for this research exploring machine learning motion capture for performing arts. Given their extensive experience interweaving digital media and live expression, they provide invaluable firsthand perspectives on the needs of choreographers and directors in

[43] Harrison Hall, https://harrisonhall.com.au/About-Harrison-Hall
[44] Dr Sam McGilp, https://thesubstation.org.au/artist/sam-mcgilp

this space. They are well positioned to assess if and how emerging tools like machine learning-based motion capture could support new modes of digitally-enabled performance. Their feedback will help determine whether machine learning-based motion capture addresses key limitations of current motion capture approaches for creative practitioners. Hands-on experimentation with the monocular system in their own practice will reveal if machine learning techniques offer meaningful benefits in real-world performing arts contexts.

Figure 5.6 below shows how research question two will be addressed through collaborative work with performing arts directors and choreographers.

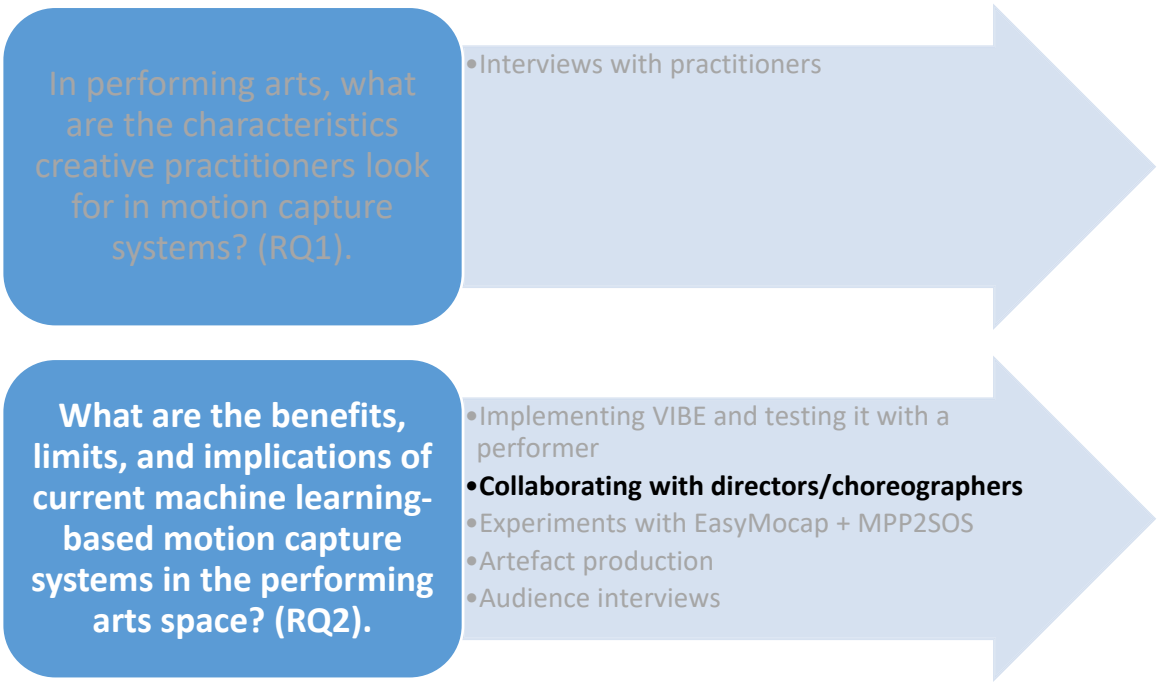

Fig. 5.6 Collaborating with directors and choreographers to address research question two.

5.3.2 Project Overview

The collaborative project with Harrison and Sam, called *Running Machine*, involved integrating monocular machine learning-based motion capture into a mixed-media stage production. Their envisioned performance combines live dance with other forms of motion tracking, like sensor-based and LiDAR systems. To enhance the visual experience, they wanted to overlay pre-rendered animations of the dancers generated from machine learning pose detection. The concept is to project the machine learning driven character meshes onto

the stage environment with live choreography. This layered effect serves both an aesthetic and informative role for the audience by visualising the underlying motion in a stylised, illustrative manner. My role was to capture dance sequences using the monocular model and convert the motion data into animated assets for insertion into the final performance. This presented an opportunity to evaluate the model's capabilities for an actual creative application. It also offers a platform to investigate the influence of the machine learning model on collaboration within this domain. Additionally, it presents an avenue to examine the responses of experienced practitioners to both the process and outputs derived from the machine learning-based approach.

Working remotely

An additional opportunity presented by this collaboration was further evaluating the monocular machine learning system in a remote workflow. The preceding chapter (5.2) also occurred in a remote setting. Significantly, however, this iteration involves a more extended and iterative form of remote collaboration. I was based in Sydney, while Harrison and Sam were located in Melbourne. We conducted project discussions over Zoom, and they shared video clips of rehearsals via online file transfer. I would then process the clips with the monocular model to generate animated meshes, which I returned for review. This remote pipeline provided a test case for the portability and accessibility of the machine learning approach without needing to be physically present for motion capture data. The ability to collaborate from a distance highlights the flexibility the single-camera machine learning system offers. It also demonstrates the potential for choreographers and directors to adopt motion capture techniques without geographic constraints, opening new creative pathways.

Objectives and goals

From a motion-tracking standpoint, Harrison and Sam were interested in exploring the visual potentials of the monocular machine learning-based system. Having seen my previous test examples of generated mesh animations, Harrison and Sam had already decided they wanted this raw, glitchy aesthetic before we began processing their own footage. They were specifically drawn to the unfiltered, glitchy quality of the mesh outputs and had planned from the start to incorporate this style into their production. After viewing some initial examples, they were drawn to the unfiltered, jittery nature of the model's pose estimations. For the desired aesthetics of their production, they preferred seeing the sporadic contours of the human form untouched by temporal smoothing or editing. This aligned with their artistic

intent to expose the model's interpretive representation of the human form. To further this idea, Harrison and Sam purposefully captured footage that would challenge the model. Their goal was to elicit unexpected and glitch-like effects that occurred when detecting human pose from video where the subject is purposefully obscured. Rather than accurate and smooth motion replication, they sought to uncover inventive visual artifacts generated by the model's attempt to reconstruct erratic inputs. This experimental approach allowed assessing the model's capabilities beyond reliable tracking, opening up new artistic possibilities at the boundaries of machine learning based motion capture capability.

".. we are often interested in stressing the technology to create unusual effects or artifacts .. there are likely to be artifacts and glitches in the quality of the capture. But we are quite comfortable with that, because we think they look cool. We are able to make more interesting performances by not just being dictated to by the constraints of the technology."

Dr Sam McGilp

".. I was really interested in the way that kind of glitches and I'm really interested in the attempt that the machine tried to make on these bodies, the image of these bodies, and particularly in the attempt when we were purposely feeding the video that we fed was really quite abstracted..."

Harrison Hall

Green screen ambiguity

Various batches of video were provided for processing using the machine learning-based motion capture model (VIBE). The first batch of videos captured an intriguing scenario where two performers interacted in a creative and imaginative manner in front of a green screen. One of the performers was dressed in a green suit, while the other performed in regular attire (Figure. 5.7). Throughout the performance, the performer in the green suit would lift and manipulate the other performer, resulting in visually captivating and gravity-defying movements. The choreography of the performance was designed to give the illusion that the two performers were connected, forming a single body. In some instances, one performer wore the green screen suit on the upper half of their body, while the other wore it on the lower half, creating the illusion of a complete and unified body in their collaborative movements. An additional step involved removing the green screen background from the footage, which had the effect of making the performer wearing the green clothing less visually prominent.

This caused the detection of the regular-dressed performer to increase in accuracy while making the green suit-wearing performer less noticeable.

A fascinating aspect that emerged during this phase of the project was when the performers were obstructing the camera's view of themselves, either fully or partially, the model sometimes struggled to accurately detect and track their movements.

The intentional ambiguity in the performance, with one performer partially covered in green, adds an intriguing layer of unpredictability to the aminated mesh generation. Harrison and Sam appreciated this effect, as the model would occasionally switch between detected performers. The integration of machine learning-based motion capture, the removal of the green screen, and the intentional ambiguity in the performance resulted in an unexpected and delightful fusion of experimental dance and digital video. The project embraced the model’s occasional confusion as an artistic element, embracing the spirit of experimentation and pushing the boundaries of traditional performing arts. A video example of the mesh output of the VIBE model is available here [vidStream_F]. In Figure 5.7, the top-left corner shows the unprocessed video with two performers in front of a green screen. In the top-right, the green colour is removed to make the performer in the green suit less visible. The bottom-left corner illustrates the pose detection switching between performers, while the bottom-right corner demonstrates the pose detection mesh aligning more closely with the visible performer after removing the green in the video.

Fig. 5.7: VIBE detections.

Running machine

The next set of clips captured various performers on a treadmill, also against a green screen backdrop (Figure. 5.8). The need for the green screen backdrop was not particularly necessary, although it did provide some separation of the performer from the background for easier detection. These tie into Harrison and Sam's upcoming *Running Machine* production. The videos showed the performers running, jogging, walking and dancing on the machine. Successfully tracking the cyclic limbs and strides could generate animated assets for both the live performance and promotional materials. One example of promotional material made for the production was a 360-degree immersive video featuring the view of a camera flying through a point cloud landscape. As the camera flew through the point cloud, the generated meshes by the model would be visible in particular areas, enhancing the experience. The treadmill added a dynamic element to the performance, with the performer's movements

embodying the essence of motion and speed. The VIBE model exhibited relative ease in detecting the performer on the treadmill compared to the first set of videos examined with Harrison and Sam. The clear and unobstructed framing of the performer within the video footage facilitated more accurate detections than previously.

Two versions of the animated mesh were generated to explore artistic possibilities. The first version remained unfiltered, showcasing the raw machine learning interpretation of the performer's movements on the treadmill. This unfiltered approach allowed Harrison and Sam to witness the model's initial detections and explore the untamed essence of the performer's motion. In contrast, the second version underwent a keyframe smoothing process to remove the inherent jitter from the mesh. The keyframe filtering technique involved smoothing the animation by averaging the keyframe values over several keyframes, resulting in a temporally smooth sequence. While they found value in comparing both versions, they ultimately preferred the original animation since it showed the model's authentic output without any post-processing alterations. A video example of the VIBE mesh output with the performer on a treadmill is available here [vidStream_G]. An interactive 360-degree video example of the promotional material for *Running Machine* is available here [vidStream_H]. In Figure 5.8, a performer on a treadmill is presented against a green screen, with an overlaid generated mesh. In Figure 5.9, a singular frame is depicted from the 360-degree immersive promotional video for the *Running Machine* production, where the VIBE model detected the pose of the performer.

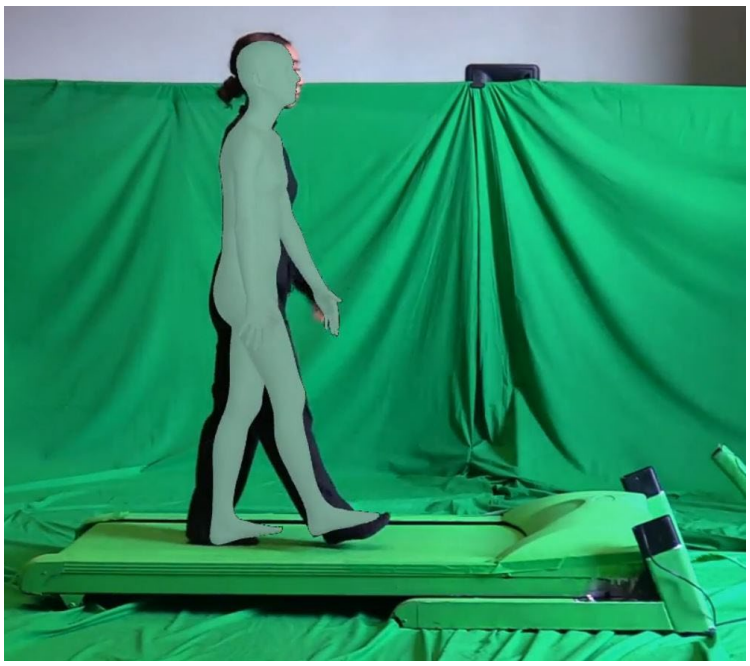

Fig. 5.8: VIBE mesh over input video

Fig. 5.9: *Running Machine*.

The model’s response to unconventional motion

Harrison and Sam aimed to test the model's capabilities by capturing unconventional movement and body shapes, along with intentionally obscuring a portion of the subject. The video footage showcased one of the performers, who was physically disabled, confidently manoeuvring on a skateboard, wearing a gorilla mask. The directors were interested in exploring how the VIBE machine learning model would interpret and animate a body that deviated from the norm, moving in ways that were distinctive and non-conventional. As the VIBE model is primarily trained on people without disability, it was anticipated that interpreting the movements of a performer with distinct gesture and body dynamics may pose a challenge to the model. In addition to variations in body shape, the complexity arose from the subject's movement on a skateboard, wearing a mask, occasional total obscuration when skating behind a pillar. These factors presented additional challenges for pose detection. As observed in the video [vidStream_I], the mesh exhibits significant erratic behaviour as it attempts to track the subject's pose, including inaccuracy and temporal irregularity. It was interesting to observe how the VIBE model interpreted the movements of the physically disabled performer in the video. The erratic nature of the animations highlighted the potential for unique artistic expression that emerges when machine learning encounters diversity.

A video example of the performer with a disability on a skateboard wearing a gorilla mask is available here [vidStream_I]. Figure 5.10 illustrates a singular video frame featuring the performer on a skateboard wearing a gorilla mask. The input video is on the left, while the right shows the VIBE mesh superimposed over the video.

Fig. 5.10: VIBE detection over input video.

5.3.3 Outcomes

The collaboration between Dr Sam McGilp and Harrison Hall and the implementation of the VIBE machine learning model has yielded interesting insights into the performance and value of ML-based pose detection when deployed in unconventional environments that largely lay outside of the types of training examples the model would have seen. This fusion of artistry and technology has opened new avenues for artistic expression, offering a glimpse into the potential of machine learning-based motion capture in the world of live performances. One of the significant outcomes of this project was the ability to harness machine learning-based motion capture to create an animated mesh derived from motion capture data to be used in a performance. This collaborative project provided valuable insights into the potentials and limitations of monocular motion capture for creative applications like performing arts. Experimenting with the system in rehearsals and productions revealed meaningful benefits that this method could offer. One of the most notable benefits is the minimal setup and equipment required. Unlike traditional motion capture systems that demand an array of cameras and intricate calibration procedures, the VIBE model proved to be efficient with just a single camera. The model also provides utility in that it can be applied post-performance capture. This accessibility meant that performers could be captured in various settings, even during the COVID restrictions, without the need for specialised equipment or elaborate setup. The simplicity of a single-camera setup can empower artists to focus on their creativity without being bogged down by technical

complexities. Traditional marker-based motion capture systems can be exorbitantly expensive, presenting a significant barrier to artists and performers with limited resources.

Harrison and Sam approached this work with the intention of challenging conventional motion capture paradigms. While traditional motion capture typically focuses on normative bodies performing expected movements, they deliberately sought to push boundaries by working with atypical bodies performing unconventional movements. Their goal was to interrogate the model's underlying assumptions through practice - specifically, to reveal and examine how the model interprets bodies and movements that fall outside its expected parameters. This provocative approach aimed to explore the limits of digital performance by making visible the model's embedded assumptions and limitations rather than trying to work within them. Through this process, they questioned the standard practice of motion capture which often prioritizes capturing 'average' bodies performing predictable actions.

While I, as the researcher, set up and operated the system in this study, it's important to clarify that the technology's accessibility operates on different levels. The performers and choreographers themselves may not directly implement the technical setup, just as they typically don't operate lighting or sound systems in traditional performance contexts. Instead, the accessibility of these systems is particularly relevant for small performance companies who can integrate this technology into their productions with relatively modest technical support.

Much like how performance companies employ technical specialists for lighting, sound, and stage management, they could similarly incorporate machine learning-based motion capture without requiring extensive infrastructure or resources. This represents a significant shift from traditional motion capture setups, which often require specialized studios and expensive equipment. Small companies can achieve motion capture effects with consumer-grade cameras and computing equipment, even if they need someone with technical expertise to implement the system.

It was evident that the models struggled with accurate and temporally coherent results when handling body types and movements that deviated from the training data. However, this apparent limitation unexpectedly opened up intriguing artistic opportunities, enabling directors to explore the boundaries and unique characteristics of machine learning in this context. An interesting aspect is that, within an artistic realm, errors need not be viewed as

detrimental or deal-breaking. This represents a stark contrast to other contexts, such as pose detection for medical purposes, where accuracy is of utmost importance.

However, with the adoption of the VIBE model, the costs associated with motion capture are drastically reduced. By eliminating the need for expensive cameras and suits with reflective markers, the VIBE model provided a more affordable and accessible solution for capturing human motion in a performing arts context. This cost-effectiveness can enable directors to allocate their budget more creatively, investing in other artistic elements to enrich the performance. Figure 5.11 displays the poster for the *Running Machine* production, where the VIBE model was used to detect poses used in the production.

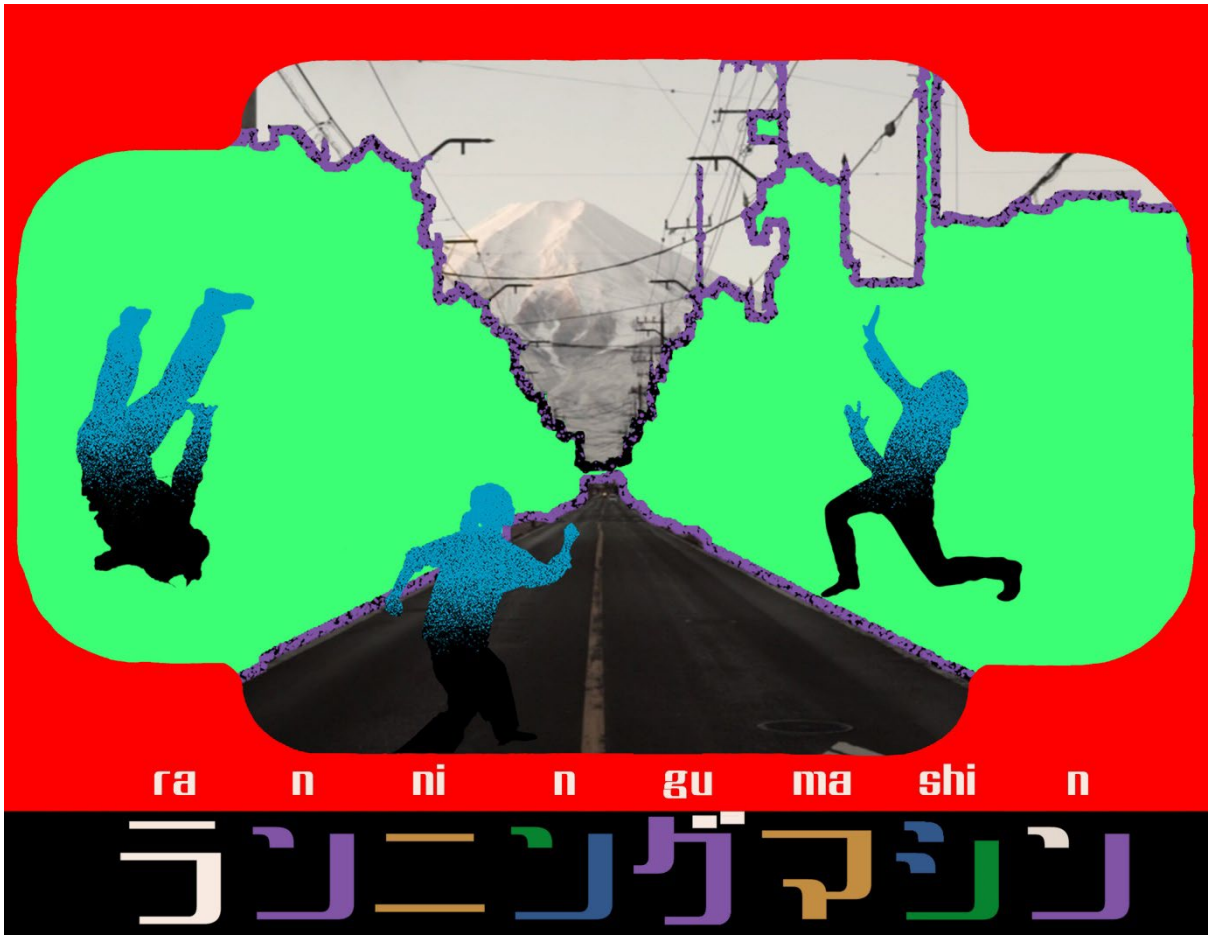

Fig. 5.11: *Running Machine* poster.

Although Harrison and Sam often work with real-time technologies in their production, they found meaningful creative potential in the VIBE system even without current real-time capabilities. The ability to generate animated meshes after capturing rehearsal footage

provided options for stylised visual overlays, and the system's temporally-coherent motions enabled new techniques for motion-driven stage visuals. Once real-time applications of monocular machine learning models that create meshes are achieved in the future, the incentives and benefits for performance workflows will become even stronger. Real-time pose tracking and animation will unlock more interactive and responsive performances powered by these models. However, in its present form, VIBE already offers valuable generative capacities for digital dance tools. The collaboration provided promising indications that machine learning can enrich choreographic possibilities, which stand to expand greatly as machine learning motion capture is improved. The project emphasised that despite some limitations, VIBE possesses substantial artistic applications, which will only increase as the technology progresses.

The promise of machine learning-based motion capture systems to enrich choreographic possibilities stems from their reduced technical barriers. Unlike traditional motion capture approaches that require specialised suits and equipment, these systems can operate with just a single video source. This simplification creates new opportunities for artistic exploration in several ways:

First, performers can move naturally without being encumbered by motion capture suits, allowing for more spontaneous and unrestricted movement exploration. Second, the minimal technical requirements - needing only a camera rather than a specialised studio setup - means practitioners can experiment more freely and frequently with the technology. This increased accessibility enables more iterative creative development and testing of ideas.

The removal of physical markers or suits also opens up possibilities for capturing movement in various contexts and settings that might be impractical with traditional motion capture systems. This flexibility allows choreographers and performers to explore movement in different environments and situations, potentially inspiring new creative directions.

These factors combine to make motion capture technology more available for artistic experimentation and integration into the creative process, rather than being limited to specific technical sessions in specialized facilities. This accessibility encourages more widespread artistic exploration of the technology's creative potential.

"I could say that there was a financial and technological barrier using inertial motion suits for a more democratised motion capture kind of environment or community. So I was just really excited because, you know, yeah, the fact that it (machine learning-based motion capture)

could just run off video, which is something that a lot of people have access to this day, and age was really exciting.." Harrison Hall

"I think it's evident that pose estimation, machine learning approaches will be the most accessible, predominant method of motion capture in the future, because it is low cost. It requires the least additional infrastructure to be imposed on the performer. And so, we're excited to work with it as a technology" Dr. Sam McGilp

5.4 Conclusion

This chapter has presented the results of an investigation into monocular machine learning motion capture for performing arts applications. Through practical experiments and collaborations with experienced choreographers, the feasibility and potential of machine learning techniques in this creative domain were explored. The findings demonstrate that the VIBE model offers notable advantages in accessibility, affordability, and portability compared to traditional motion capture systems. Its accuracy in capturing diverse movements, compatibility with animation workflows, and resilience to suboptimal conditions also show advantages. However, limitations arose with the occasional missed detections on fast motions and inherent jitter in raw mesh animations.

More significantly, it consistently encountered challenges in detecting keypoints associated with unconventional movements, body configurations, and orientations. Notable areas include instances where the performer was manipulated by an individual clad in green attire against a green screen, resulting in the model's struggles to accurately discern poses and produce a mesh with irregular motion. Another illustration arises in situations where the model grapples with the identification of atypical bodily structures in videos featuring individuals with disabilities. The inaccurate outputs generated in cases of erratic detection, however, were shown to also present creative opportunities for artistic exploration.

Collaborating with innovative directors revealed meaningful creative potential in machine learning-based motion capture, even in its current offline form. The generated visual assets opened new possibilities for stylised overlays and abstractions of movement. The artifacts and ambiguity from challenging the model yielded intriguing results showing machine learning's interpretive role in capturing movement.

The collaboration with choreographers and directors in performing arts field addresses the research question of: 'What are the benefits, limits, and implications of current machine

learning-based motion capture systems in the performing arts space?' (RQ2). Findings indicate that monocular machine learning-based motion capture:

- Offers accessibility through single-camera solutions and cost-effectiveness by eliminating expensive cameras and suits.
- Enables remote collaborations and workflows without geographic constraints.
- Provides resilience to suboptimal conditions like dynamic lighting or partially obstructed views.
- Creates intriguing aesthetic effects, such as glitches, from unconventional inputs.
- Opens new creative directions in visualisation and abstraction of movement.
- Accelerates innovation in technology integrated digital performance.
- Acceptable accuracy and temporal coherent animation when working in settings that are poorly represented in training sets.

Overall, this research provides evidence that monocular machine learning techniques can enhance motion capture for performing arts by lowering barriers to access. Additionally, training sets could be created that encompass a broader spectrum of artistic performances and encompass a more diverse range of body shapes and movement dynamics, including individuals with disabilities. With a focus on creative workflows, machine learning-driven motion capture can become an empowering asset for innovation in technology-integrated stagecraft. The collaborative insights compiled here can help guide the development of machine learning tools tailored for the arts. By embracing the spirit of experimentation, machine learning can accelerate the evolution of digital performance practices.

6 Experiment: Multi-camera Pose Detection for Performance

6.1 Introduction

Multi-camera pose detection using machine learning has emerged as a promising approach for a more accurate method than monocular systems. Multi-camera setups provide more observation views of the subject and scene, allowing the model to overcome occlusion challenges and ambiguity in single-view detection. Although multi-camera systems require more equipment and potentially greater effort and know-how to manage camera calibration, they promise improved accuracy and robustness in pose estimation.

The primary aim of this chapter is to investigate the feasibility and effectiveness of utilising a multi-camera system for motion capture in the context of performing arts. This exploration stems from the recognition that traditional optical motion capture methods using markers and suits provide very accurate pose estimation but require dedicated studio spaces and expensive equipment, lacking portability. Single-camera machine learning methods are more accessible and affordable but, as seen in the preceding chapter, have limitations in accuracy when images are occluded. Multi-camera pose detection aims to strike a balance, gaining enhanced accuracy from multiple views while remaining portable and cost-effective compared to traditional optical systems. A series of experimental tests were conducted in various settings to gain a comprehensive understanding of the practical implications and performance of the multi-camera pose detection model. These tests were designed to provide valuable insights into the feasibility, accuracy, and overall effectiveness of the multi-camera system in comparison to the monocular setup.

6.2 Camera calibration

Camera calibration is a foundational process that establishes a mapping between the real world and the camera's image plane. It rectifies distortions and aligns multiple cameras to ensure accurate and consistent pose estimation across different viewpoints (Sadekar & Mallick, 2020). While camera calibration introduces an element of technical complexity, its successful execution is paramount for many multi-camera pose detection systems. Calibration involves intricate steps such as capturing calibration patterns, calculating intrinsic and extrinsic camera parameters, and performing distortion correction (discussed in Chapter 6.2.2 and 6.2.3).

However, it is noteworthy that the current body of literature lacks comprehensive guidance on the optimal sizing of checkerboard patterns during camera calibration, discussed in Chapter 6.2.1. This absence may require a measure of trial-and-error during the calibration process, potentially leading to delays in the initiation of experiments or production.

Furthermore, the need for camera calibration introduces challenges related to reconfiguration in different capture spaces. This aspect may complicate the application of multi-camera motion capture technology in scenarios involving performances with movements between stages or distinct areas. The recalibration when transitioning between spaces adds a layer of complexity that artists need to navigate, impacting the practicality and efficiency of employing this technology in diverse performance settings.

6.2.1 Chess pattern

At the core of achieving accurate and reliable multi-camera pose detection lies a critical preliminary step: the calibration of cameras using a carefully designed chess pattern. The pattern consists of alternating black and white squares arranged in a grid pattern. This high-contrast arrangement facilitates accurate feature detection and extraction, which is essential for precise calibration. The physical dimensions of the squares need to be precisely measured and recorded, as these measurements are used in the calibration code to compute intrinsic and extrinsic parameters. The chess pattern is affixed to a hard, stable surface to ensure accurate and consistent calibration results. Ideally, the surface should be rigid and flat, such as a thin piece of wood. A surface that does not bend or flex is essential, as any deformation could introduce inaccuracies into the calibration process. The stability of the surface is paramount to maintain the alignment of the chess pattern and to ensure that the cameras capture the pattern accurately from different angles (Zhang, 2008).

While the size of the chess pattern can vary, there are some general guidelines to consider based on the experiments I conducted. While the literature does not specify a precise size for the chess pattern, its dimensions are contingent on factors like the size of the capture area and the distance between cameras. It’s imperative that each camera can easily discern the chess pattern, otherwise the calibration process may fail. Initially, an A4 paper size (210 mm x 297 mm) was chosen for printing the chess pattern. This size proved visible to all cameras when employed in smaller capture spaces, typically spanning around two square

meters. However, when conducting tests in more extensive capture spaces, the A4-sized pattern became less discernible due to the increased distance to the camera. Consequently, a larger A2 size (420 mm x 594 mm) was adopted, ensuring its visibility and effectiveness in these larger settings. In Figure 6.1, a chess pattern is presented on an A4-sized sheet for camera calibration on the left, while an A2-sized chess pattern is displayed on the right.

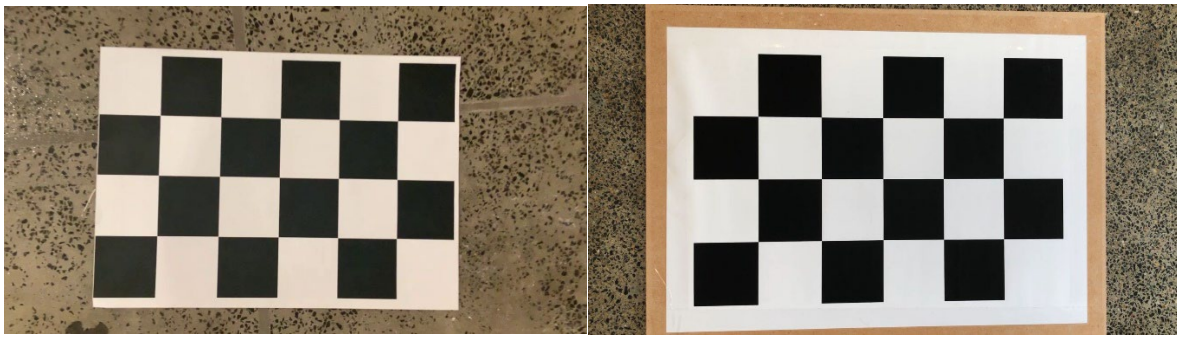

Fig. 6.1: A4 and A2 chess patterns.

6.2.2 Calculating camera intrinsic parameters

Accurate and reliable pose estimation hinges on understanding and mitigating the nuances introduced by lens distortion. Every camera lens possesses a certain degree of distortion that can impact the fidelity of image capture. Intrinsic calibration emerges as a pivotal step, addressing this distortion paves the way for more precise pose estimation. Optical distortion is a consequence of the inherent curved shape of the camera lens. The curved design, while essential for focussing light onto the sensor, introduced distortions that can significantly affect the accuracy of image capture. The distortion originates from the differential magnification across the lens's field of view, leading to subtle yet perceptible alterations in the appearance of captured scenes. The lens curvature accentuates magnification towards the centre of the lens, yielding an effect where the centre of the captured image is slightly magnified compared to the edges. Consequently, seemingly straight lines in the real world appear to curve or bend along the edges of the image. This curvature, often referred to as barrel or pincushion distortion, adds an element of complexity to the mapping between the 3D world and the 2D image plane (Grigonis, 2023). Figure 6.2 displays various types of lens distortion. On the left is the absence of distortion, in the middle is barrel distortion, and on the right is pincushion distortion.

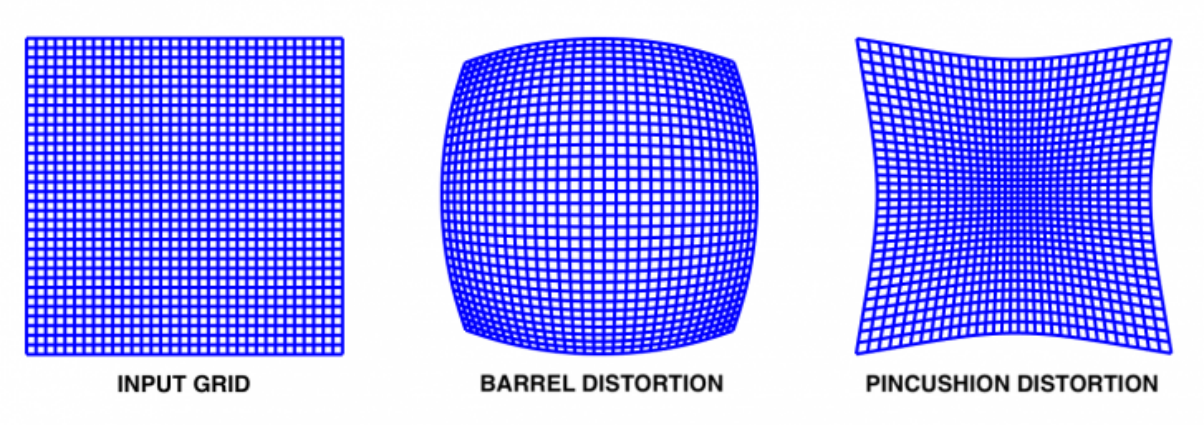

Fig. 6.2: Types of distortion (Dulari Bhatt, 2021).

To comprehensively account for lens distortion, the pattern is observed from diverse viewpoints, ensuring its presence across different regions of the image frame. By capturing the pattern at various distances, orientations, and angles of view, the calibration process captures a range of distortion-induced alterations. The diversity in the appearance of patterns enhances the calibration data, enabling it to effectively model and rectify distortions (Zhang, 2008). Sophisticated calibration algorithms leverage the captured video footage and the observed pattern variations to iteratively estimate the intrinsic parameters of each camera. The derived intrinsic parameters lay the foundation for distortion rectification, ensuring that the subsequent pose estimation process accounts for and compensates for the inherent lens-induced distortions (Sadekar & Mallick, 2020).

6.2.3 Calculating camera extrinsic parameters

Extrinsic parameters calculate where each camera is positioned in space to ensure accurate pose detection (Dulari Bhatt, 2021; Zhang, 2021). Extrinsic parameters are calculated using the same chess pattern as the calculation of intrinsic parameters. Unlike intrinsic parameters that focus on lens-specific distortions, extrinsic parameters delve into the physical arrangement of the cameras in relation to one another and the scene they capture. The chess pattern acts as a spatial reference point, enabling the system to triangulate the position and orientation of each camera in the 3D scene. Each square within the pattern is a known entity, acting as a geometric anchor that facilitates the correlation between the captured images and the spatial reality (Anwar, 2022).

The chess pattern is ideally positioned on the floor of the capture space so that it is stationary. Once the cameras are set into position to record the performer, the pattern is placed in a position where each camera has an unobstructed view of it. Through the camera's lens, the pattern serves as a canvas of distinctive features to be decoded and translated into meaningful spatial coordinates. Utilising pattern detection algorithms, the software calculates the precise locations of the corners of the squares within the field of view of each camera, based on the image data captured.

Simultaneously, corresponding corners of the pattern are recorded by every other camera, forming a web of interconnected data points. This intricate network of correlated square corners allows the calibration algorithm to discern the spatial interplay between the cameras, facilitating the determination of their relative positions and orientations (Anwar, 2022). Figure 6.3 illustrates a chess pattern positioned on the floor of the capture space, visible to all cameras. The green highlights indicate that the chess patterns were detected successfully.

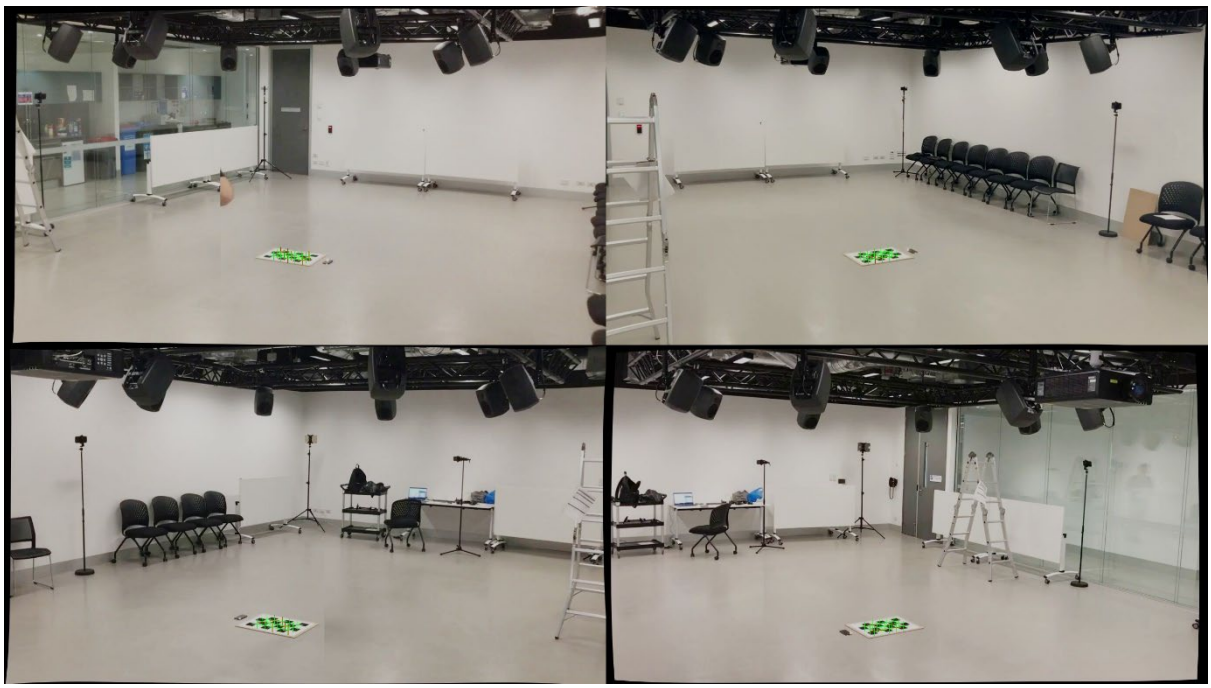

Fig. 6.3: Chess pattern visible to all cameras.

The choice of the chess pattern size is intrinsically linked to the dimensions of the capture space and the intended field of view for each camera. Striking the right balance between pattern size and capture area ensures the pattern is optimally visible within the camera

frames. In the camera calibration process for multi-camera pose detection, it was observed that the size of the chessboard used for calibration is a critical factor. The visibility of corners to each camera is enhanced with a larger-sized chessboard. Conversely, smaller-sized chessboards, such as A4 dimensions, prove inadequate in larger capture spaces exceeding 10 square meters, as they appear too small in the camera's field of view, making the chess squares indiscernible.

Once the cameras have recorded the chess pattern and initiated the extrinsic parameter calculation, stability and immobility of the cameras and tripods become paramount. Any movement or displacement of the camera's post-pattern recording can introduce errors in the calculated extrinsic parameters. Even the slightest shift can disrupt the spatial relationship between cameras and the chess pattern, rendering calibration invalid. Therefore, ensuring that the cameras and tripods remain securely fixed in their positions throughout the calibration process and subsequent performance scenarios is essential. The cameras' immobility extends beyond preventing physical displacements. Environmental factors such as vibrations, shocks, or accidental bumps can also jeopardise the integrity of the extrinsic calibration. These disruptions can go unnoticed yet profoundly impact the calibration's accuracy. Practitioners should remain vigilant and implement measures to safeguard against any inadvertent disturbances that may compromise the calibration's precision. This requirement carries implications for the applicability of the technology in performance art spaces, where a stable and immovable platform or stage is typically utilised.

While automated pattern detection algorithms streamline the calibration process, occasional challenges may arise. Factors such as lighting variations, pattern size, camera focus, or other unforeseen conditions can lead to failed pattern detection attempts. In such instances, manual intervention becomes necessary. The LabelMe application (Marois & Syssau, 2008) offers a solution, allowing practitioners to input the positions of square corners within the pattern manually. This manual entry serves as an alternative data source to ensure accurate correlation and calculation of extrinsic parameters, even in scenarios where automated detection encounters difficulties. LabelMe is an open-source annotation tool that offers a user-friendly interface that allows manual pixel coordinate recording. The application loads the relevant image, presenting it in a visually intuitive format allowing easy interaction. A crosshair tool serves as the practitioner's virtual pointer, assisting in selecting a specific pixel for recording. The application records the pixel's X and Y coordinates, which can be used to create the extrinsic camera parameters.

However, it is important to acknowledge that the manual labelling process introduces considerations related to speed, accuracy, and reliability. While manual labelling allows for precision, it tends to be time-consuming and may not be suitable for real-time applications, impacting the speed of the calibration process. Automated detection, on the other hand, is faster but may compromise accuracy under challenging conditions. The choice between automated and manual approaches involves a trade-off between the efficiency of the calibration process and the robustness of result. Further exploration and consideration of these trade-offs are necessary for practical implementation in dynamic performance environments. In Figure 6.4, an instance of incorrectly detected chess pattern is depicted, demonstrating inaccuracies in detecting the corners of the pattern. The coloured markers on the chess pattern indicate where the corners of the pattern were detected.

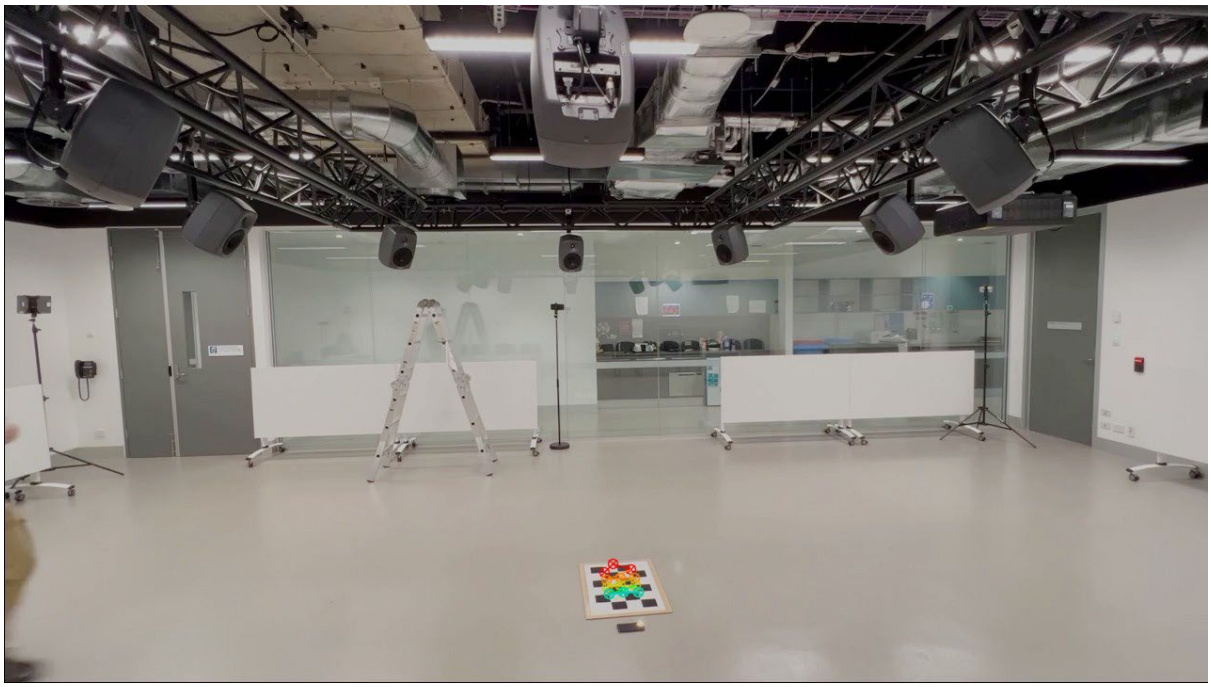

Fig. 6.4: An incorrectly detected chess pattern.

6.2.4 Conclusion

Camera calibration is a necessary cornerstone for multi-camera pose estimation. The calibration process rectifies lens distortions and establishes the spatial interplay between cameras by accurately determining intrinsic and extrinsic camera and spatial parameters. This foundational step is crucial for translating performers' dynamic movements and

expressions into a cohesive digital representation that resonates with the essence of live performance.

While camera calibration may appear complex and daunting to some, cultivating and understanding its mechanics empowers performing arts practitioners to navigate this technical landscape confidently. Acquiring the know-how to calculate intrinsic parameters and discern the significance of extrinsic calibration lays the groundwork for harnessing the full potential of multi-camera technology with machine learning.

The intricacies of camera calibration are not devoid of challenges, especially in dynamic studio environments where dancers bring the stage to life. The need for precise setup, immobility, and stability post-calibration can indeed pose logistical hurdles. Variations in lighting conditions, pattern size, and camera focus may result in failed chessboard pattern detection attempts. Such challenges may necessitate manual labelling of the chessboard, introducing further complexity and preparatory overhead to the process.

The choice between automated and manual labelling methods more generally presents practitioners with a decision regarding the trade-off between speed and accuracy. While automated methods offer faster calibration, they may compromise accuracy. Practitioners need to carefully consider these trade-offs based on their specific requirements and constraints in the context of dynamic performance spaces.

Furthermore, the lack of comprehensive documentation on optimal checkerboard pattern sizes for calibration introduces uncertainty and may require a trial-and-error approach, potentially delaying the initiation of experiments or production. In smaller capture spaces, A4 chess pattern sizes prove sufficient for calibration purposes. However, in larger capture spaces, larger chess sizes, such as A3 or A2, become preferable due to their increased dimensions, enhancing visibility from each camera's viewpoint.

Finally, the need for stationary cameras during calibration has implications for the practical use of multi-camera pose detection in performance spaces. In performance art scenarios where stability is paramount, any movement of the camera platform may negatively impact the accuracy and reliability of the motion capture data output. Practitioners should consider the implications of these requirements when planning and executing performances in dynamic spaces.

6.3 Easymocap

6.3.1 Motivations for model selection

After investigating numerous models, EasyMocap (Shuai, 2021), a versatile machine learning model,was selected. EasyMocap stood out because of its comprehensive feature set and its popularity on the GitHub platform. EasyMocap is a multi-camera human pose estimation system that can track body, hand, face, and finger motions. The software uses multiple video feeds as input to estimate 3D skeletal poses. Beyond body tracking, EasyMocap provides facial animation and finger tracking capabilities. However, this study focused solely on evaluating EasyMocap's performance for full-body motion capture and did not comprehensively assess its facial and finger tracking features as it was not within the scope of this thesis.

One of the benefits of EasyMocap's toolkit is the SMPL (Bogo et al., 2016) module, which stands for "Skinned Multi-Person Linear" model. This module transforms the traditional concept of pose estimation by creating a 3D mesh derived from the detection data. SMPL leverages the keypoints generated by the multi-camera setup, utilising these data points to create a 3D mesh that embodies the performers' motion. Creating an animated mesh based on the estimated skeletal motion facilitates compatibility with animation software. Having a skinned mesh, rather than just the raw skeletal data, enables direct integration into standard 3D animation pipelines. This streamlines the process of translating the captured movement into digital assets suitable for applications like animated films, video games, VR experiences, and beyond.

The EasyMocap model is configured and executed within the Anaconda[45] virtual environment, harnessing the power of contained configurations. This encapsulated environment is safeguarded against conflicts with existing software and libraries, fostering a controlled setting conducive to seamless experimentation. The virtual environment served as a workshop where the EasyMocap model could run without external disruptions or interfering with other models on the same computer. While the devices recorded movement, a dedicated app orchestrated rhythmic flashes of light, acting as a synchronised metronome on a device in view of all cameras. This method ensured the captured videos were initiated

[45] Anaconda, https://www.anaconda.com/

from the same frame, creating a reference point that could be translated by the EasyMocap model.

Figure 6.5 below shows how research question two will be addressed by experimenting with machine learning-based motion capture models including EasyMocap and MPP2SOS.

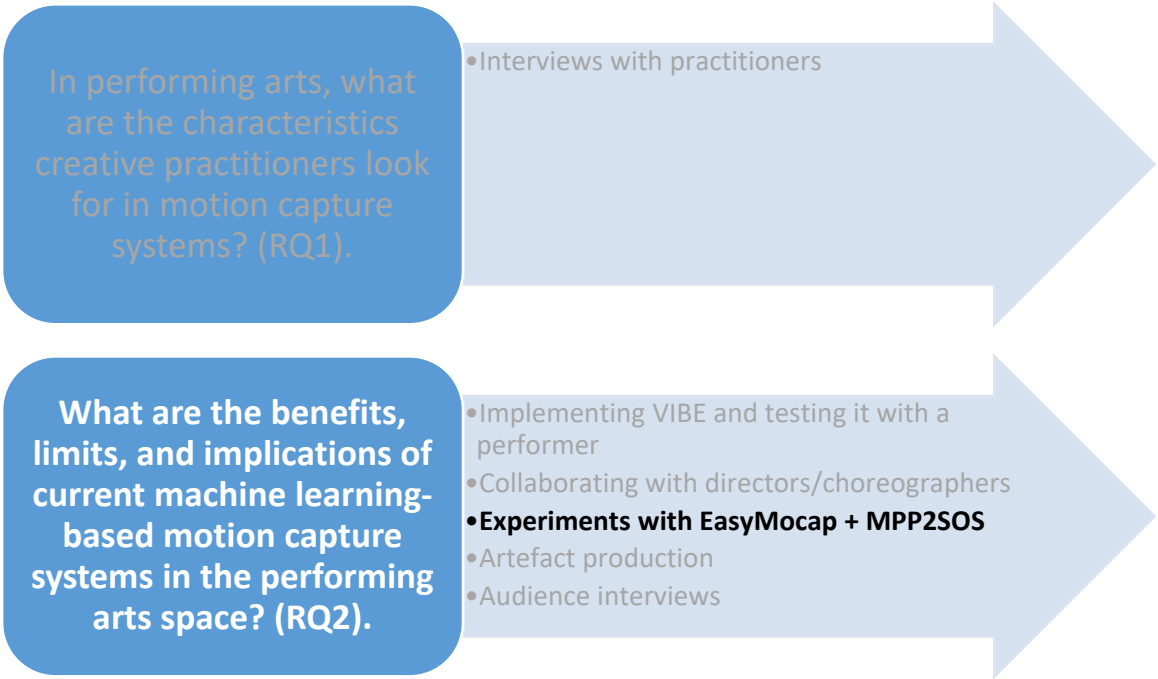

Fig. 6.5 Experiments with EasyMocap and MPP2SOS to address research question two.

6.3.2 Setup and data collection

Location one:

Testing the multi-camera machine learning model was based on real-world scenarios, spanning two distinct locations that mirrored environments commonly encountered in a performing art setting. The first location included real-world challenges that performers might encounter. Sub-optimal lighting conditions including a reflective floor. The presence of glare on the camera's lens from the lights in the scene and a cluttered background further compounded the possibility of inaccurate pose detection. In this context, a modest A4-sized camera calibration pattern was utilised. The capture space spanned approximately 1x2 meters, encompassing a compact area that mimics a limited space that may be available in particular performance settings. The selection of this size enabled the emulation of the physical constraints some performers may face, where movements and expressions unfold within defined boundaries. The choice of a smaller calibration pattern size mirrored the

practical considerations of real-world scenarios, testing the model's adaptability to diverse performing arts contexts. Figure 6.6, shows the lighting conditions at the first location.

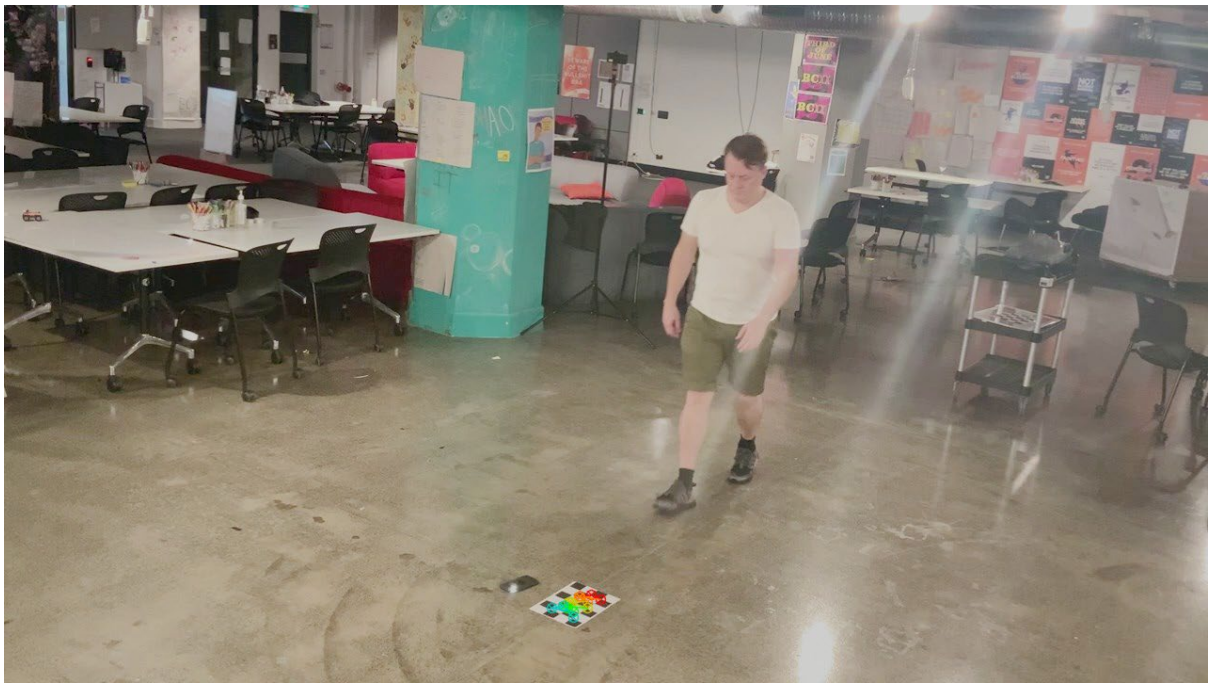

Fig. 6.6: Location one.

Camera layout

Four devices were used to capture performer movement: an iPad Pro, iPhone X, iPhone 7, and an iPad Air facing towards the centre of the capture area. The devices were mounted on tripods approximately 160cm from the ground and positioned in the corners of a space of approximately 6 x 5 meters. The devices recorded the movement simultaneously in landscape mode at HD720 resolution and 30 fps. This setup aimed to emulate the challenges, available equipment and intricacies that performers may encounter on stages marked by sub-optimal lighting and confined spaces. A video example of the mesh output of the EasyMocap model superimposed over the video is available here [vidStream_J]. Figure 6.7 illustrates the arrangement of cameras at the first location. Figure 6.8 displays the first location, presenting the overlay of the mesh on the input video utilising the EasyMocap model.

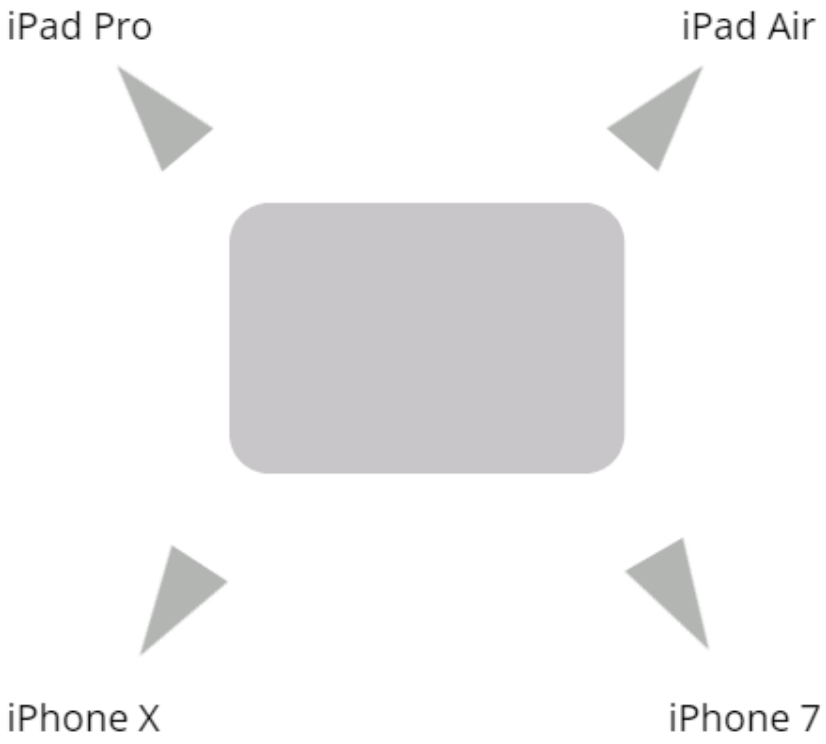

Fig. 6.7: Location one camera layout.

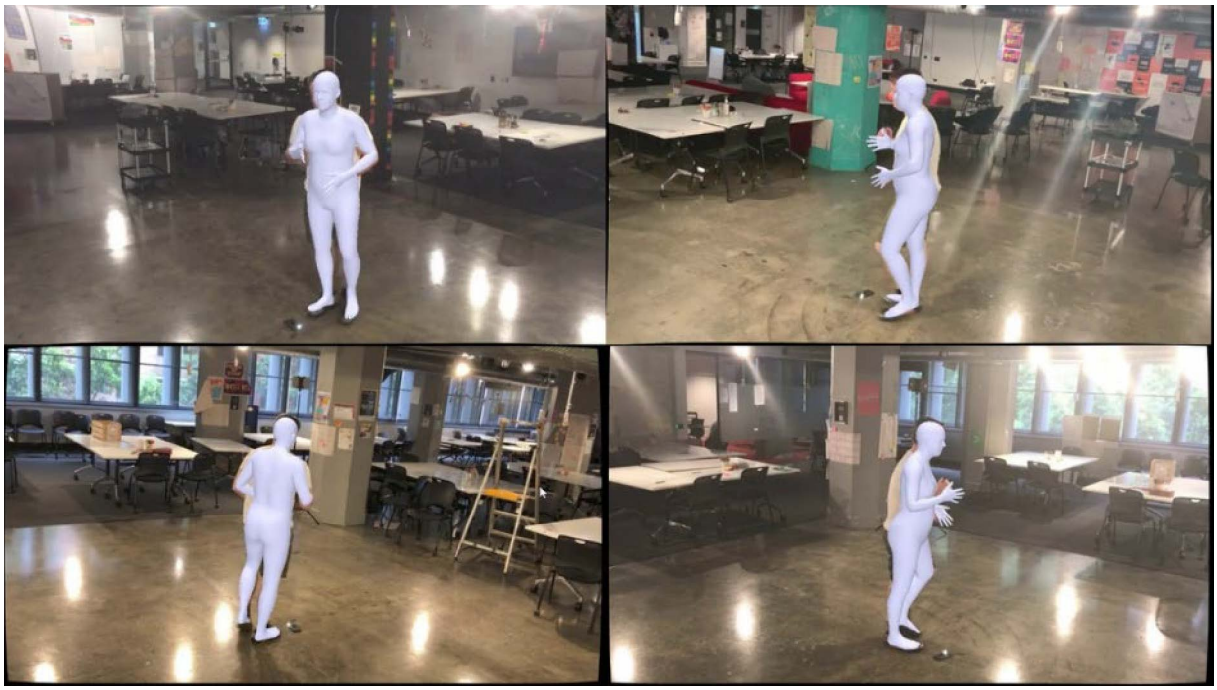

Fig. 6.8: Location one, with the mesh overlay.

Location two:

The second location had improved lighting conditions, featuring a less reflective floor that minimised glare. The absence of external sunlight contributed to a controlled environment, ensuring consistent lighting throughout the data collection process. The capture area expanded to approximately 3 x 4 metres, mirroring a larger performance stage that allowed performers to command a more extensive physical canvas. In alignment with the expanded spatial dimensions, a larger camera calibration pattern was employed at this location, and there was considerably less clutter in the background. The heightened size of the pattern accounted for the increased capture space, enabling enhanced accuracy in capturing feature data. With the expansive canvas of this location, a distinctive feature emerged in the form of a glass wall that bordered one side of the performance arena. The other side of the reflective glass wall would occasionally have people walking by. The EasyMocap model showed resilience, despite the intermittent presence of people moving by the reflective glass, the model focussed on the movement it was intended to capture.

Camera layout

The second location for testing the EasyMocap model used six devices. The devices included an iPad Pro, a Huawei Mate 20 Pro, an iPhone 6S, an iPhone 7, an iPhone 13 and an iPad Air 2. They also stood approximately 160 cm from the ground on tripods. The cameras were in a space of approximately 10 x 7 metres, where the devices painted a more expansive backdrop. Similar to the first location, this setup also recorded in landscape mode at HD720 resolution and 30fps. A video example of the EasyMocap mesh output on the input video is available here [vidStream_K]. Figure 6.9 illustrates the arrangement of cameras at the second location. Figure 6.10 displays the second location, presenting the overlay of the mesh on the input video utilising the EasyMocap model. Table 6.1 presents a comparison between the two testing locations for evaluating the EasyMocap model.

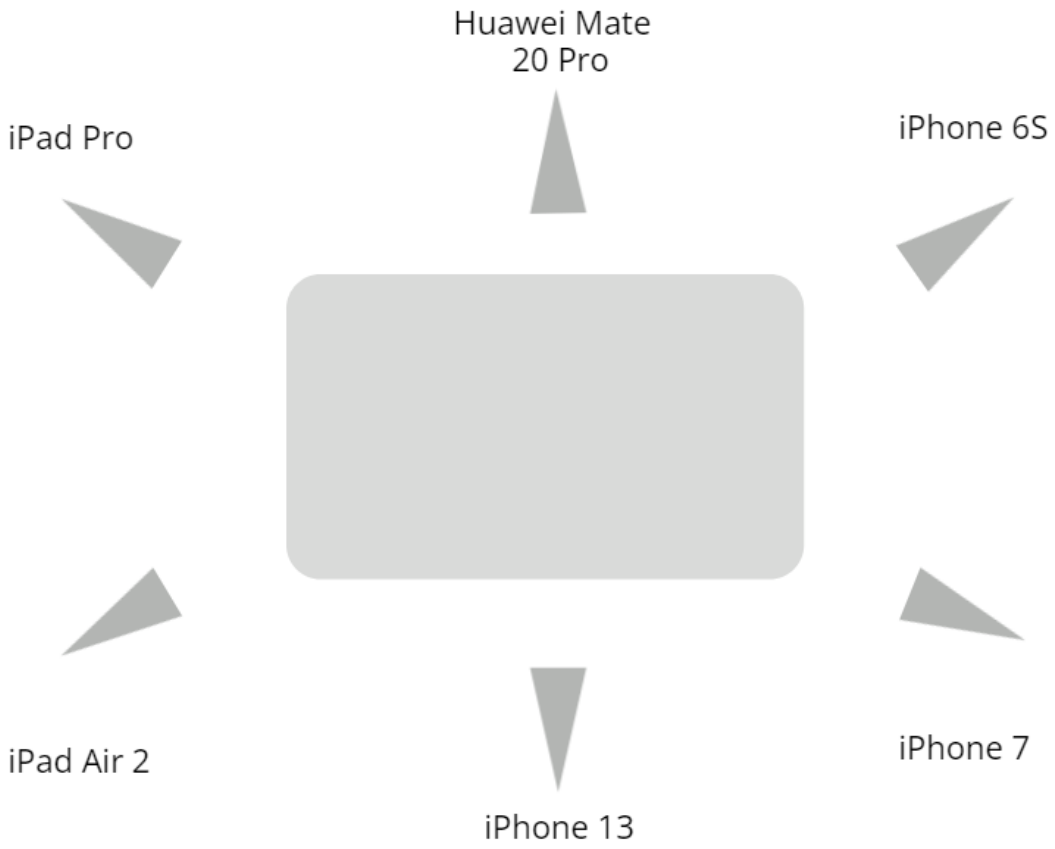

Fig. 6.9: Location two camera layout.

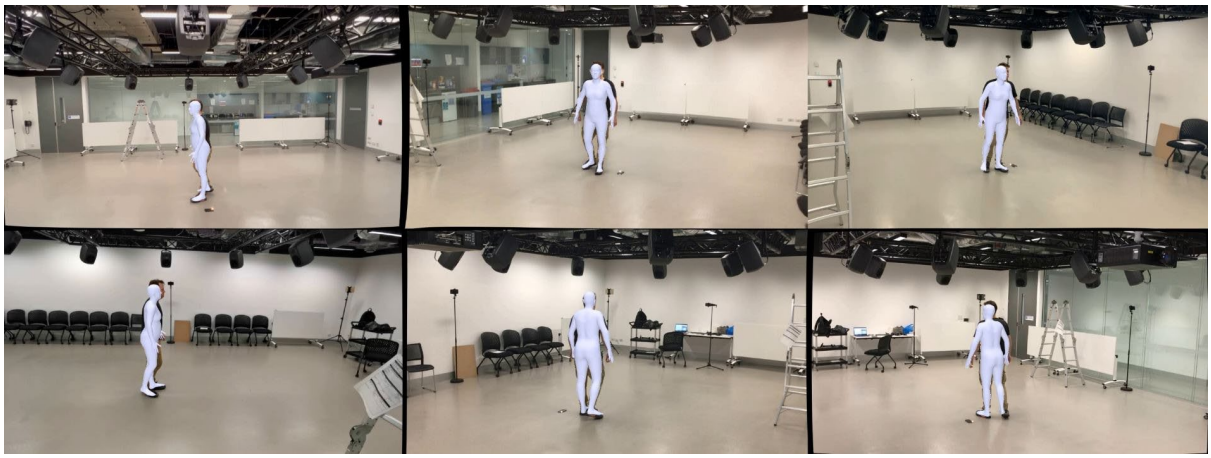

Fig. 6.10: Location two, with the mesh overlay.

Aspect	Location one	Location two
Number of devices	4 (iPad Pro, iPhone X, iPhone 7, iPad Air)	6 (iPad Pro, Huawei Mate 20 Pro, iPhone 6S, iPhone 7, iPhone 13, iPad Air 2)
Device height	Approx. 160 cm	Approx. 160 cm
Capture area dimensions	Approx. 1 x 2 metres	Approx. 3 x 4 metres
Working area dimensions	Approx. 6 x 5 metres	Approx. 10 x 7 metres
Recording resolution	HD720	HD 720
Recording frame rate	30 fps	30 fps
Lighting conditions	Sub-optimal	Improved
Floor reflectivity	Reflective	Minimally reflective
Light interference	Glare from lights, bright external light	No external sunlight or glare
Approximate time from video analysis to mesh creation	Approx. 45 minutes	Approx. 1hr 15 minutes

Table 6.1: Location comparisons for EasyMocap testing.

6.3.3 Review of outputs

The pattern detection mechanism faltered when calculating the extrinsic camera calibration at both locations. The reason for the failure in the chess pattern detection mechanism remains unclear. Manual detection emerged as a viable solution where automation failed. On select camera angles, the positions of key features on the chess pattern were manually recorded using the LabelMe application.

There was no meaningful difference in the capture performance observed across our two locations. The EasyMocap model was robust in the presence of sub-optimal lighting, reflective surfaces and across varying capture stage dimensions. The animated mesh generated by EasyMocap resulted in a substantial enhancement in temporal smoothness compared to VIBE. EasyMocap had a noticeable increase in accuracy, as evidenced by the animated mesh that followed the movement of the captured video.

Worth noting, however, was the appearance of a sporadic glitch in EasyMocap's outputs. This glitch manifested as occasional rotations within the generated mesh, seemingly unrelated to the detected motion. While the motion appeared accurate, these rotations were a puzzle requiring investigation. Initially, attention was directed towards the camera calibration process to identify any potential sources of error. However, an exploration of GitHub forums revealed that this glitch was not an isolated occurrence. Other users had encountered similar issues and engaged in discussions regarding its origins and potential solutions. The authors of the code acknowledged the glitch and indicated their ongoing efforts to address it. This glitch underscores the complex nature of software development when creating state-of-the-art models and the challenges inherent in creating them. In the context of this study, this glitch does not invalidate the overall findings but rather highlights the ongoing refinement process intrinsic to machine learning. A video example of the glitch presented by EasyMocap is available here [vidStream_L].

Due to a desire to ensure glitch-free results for later artefact production, a decision was made to explore alternative multi-camera pose detection models. Practitioners seeking to harness such technologies may grapple with similar challenges. Navigating such challenges can be both frustrating and discouraging, particularly for practitioners who may not possess the technical expertise to troubleshoot or identify alternative solutions. The spinning glitch serves as a reminder that the landscape of machine learning-based mocap is – at least at present - marked by nuances and intricacies which demand careful engagement by end users beyond the mere execution of code.

6.3.4 Conclusion

This chapter delved into the application of EasyMocap, an open-source multi-camera machine learning model available on GitHub. The model can generate a detailed mesh that accurately represents human movement, surpassing the accuracy of monocular models, such as VIBE. Testing was conducted on two distinct locations, with varying conditions and setups, to assess the model's performance and versatility. In the first location, which simulated real-world challenges, sub-optimal lighting and a smaller space were considered. Despite these obstacles, EasyMocap produced results that were visibly similar to the second location with optimal conditions. This suggests that the model works in distinct settings, even where sub-optimal conditions are encountered.

A notable glitch involving the occasional spinning of the mesh unrelated to detected motion highlighted the complexities of software development. While the model's developers acknowledged this glitch, it prompted consideration of alternative models to ensure glitch-free results. The challenges of camera calibration and manual data input were also identified, revealing potential hurdles practitioners may face when employing multi-camera systems.

While limitations existed in this version, EasyMocap shows potential as a viable option for performing arts in the future once the glitch is resolved. The following exploration focuses on evaluating a glitch-free multi-camera pose detection model for motion capture tailored to creative needs.

6.4 MPP2SOS

6.4.1 Motivation for model selection

MPP2SOS (Barreto, n.d.) is an amalgamation of three distinct models – MediaPipe[46], Pose2Sim (Pagnon et al., 2022a) and OpenSim[47]. MPP2SOS is designed to fuse the capabilities of these three models to realise advanced motion capture. MediaPipe serves as the foundational framework, enabling pose estimation. Its capacity to recognise intricate human movements sets the stage for robust motion capture data. Pose2Sim steps in as the intermediary, by triangulating the poses from all cameras, filtering the animation and configuring OpenSim. OpenSim adds the final layer by allowing the generated skeleton to be manipulated in 3D applications.

While EasyMocap provided a direct animated mesh output, the MPP2SOS model took a different approach. Rather than generating a complete 3D character mesh, it produced a skeletal structure consisting of joints and bones. To achieve a comparable animated character for performance needs, an additional step of skinning is required after using MPP2SOS. This skinning process involves creating a 3D mesh and binding it to the joint

[46] MediaPipe, https://developers.google.com/mediapipe
[47] Open Sim, https://simtk-confluence.stanford.edu:8443/display/OpenSim/Welcome+to+OpenSim

skeleton to achieve a movable character. Although skinning requires some familiarity with 3D software, it is a straightforward process for experienced 3D animators.

Blender, widely used 3D software in research(Gu et al., 2019; Mills et al., 2017; Pagnon et al., 2022b; Su et al., 2023), is favoured for its open-source nature and robust tools for visualising 3D data. Additional features, known as add-ons, can be developed to expand Blender's 3D capabilities or refine its functionality. MPP2SOS was integrated into Blender as an add-on for convenient implementation by Carlos Barreto.

Unlike the glitches encountered with EasyMocap, the initial phases of testing MPP2SOS exhibited promise. Using MPP2SOS resulted in fewer glitches compared to EasyMocap during initial testing, and it was more user-friendly because of its integration into the Blender platform. Additionally, MPP2SOS offers a streamlined process that simplifies the execution of motion capture. The steps required to set up and run the model are relatively straightforward, alleviating concerns for those who may not be well-versed in technical intricacies.

6.4.2 Setup and data collection

Initial tests of the MPP2SOS model were undertaken using a camera setup similar to EasyMocap, but with the inclusion of five cameras. Using MPP2SOS resulted in fewer glitches compared to EasyMocap, and it was more user-friendly because of its integration into the Blender platform. Instead of manually writing lines of code, the process was initiated with a simple button press in the Blender user interface, after which the various processes were automated and executed with minimal intervention. Figure 6.11 illustrates the camera layout when trialling the MPP2SOS model. A video example of the output of the MPP2SOS model compared to the input video is available here [vidStream_M].

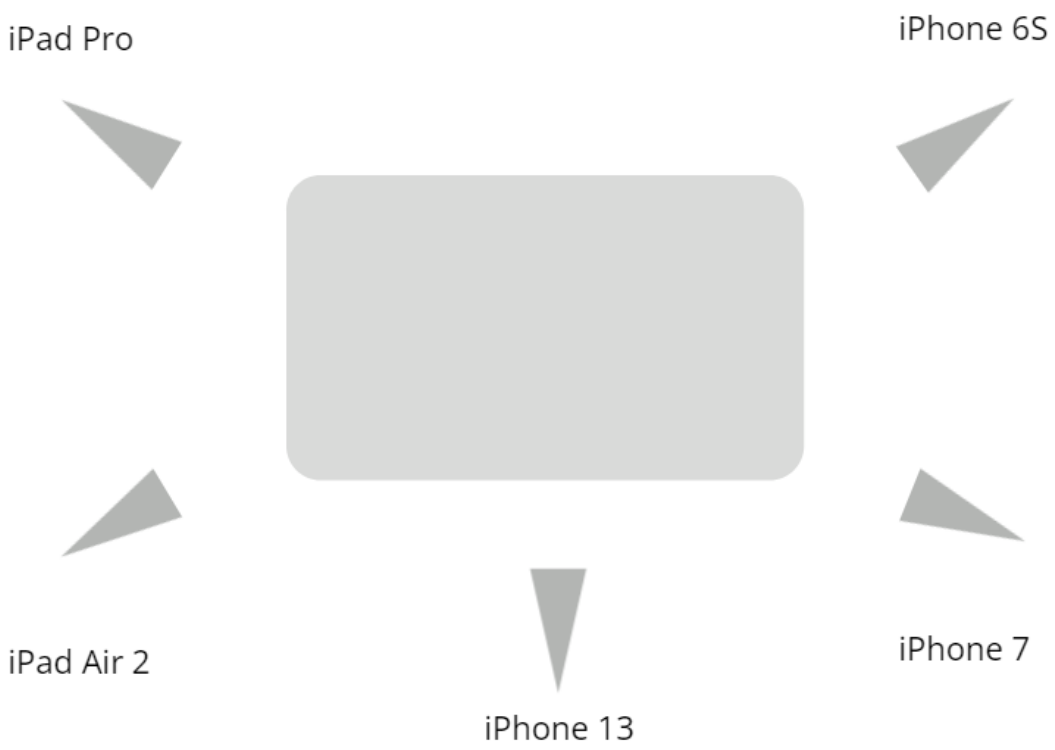

Fig. 6.11 The camera configuration for trialling the MPP2SOS model.

With successful preliminary trials of the MPP2SOS model complete, the journey ventured into more authentic realms. The goal was to assess the model's capabilities in capturing the dynamic and intricate movements of performers in a real-world setting – a pivotal step in evaluating its practical applicability in performing arts. An opportunity emerged through collaboration with the dance production company "Box of Birds[48]", helmed by director David Clarkson. At the time, the company was rehearsing a performance that embodied a blend of ground-based expressions and suspended movements on aerial slings. The fusion of gravity-defying actions and earthbound gestures presented an ideal stage to examine MPP2SOS's capabilities in capturing diverse forms of movement. This experiment aimed to capture the performance with multiple cameras, process the movement, and apply an abstract animation to it. The experiment took place in a large hall, with a capture area spanning approximately 25 meters by 15 meters. Four cameras were employed to record the

[48] Box of Birds, https://www.boxofbirds.net/

performer's movements, including iPhones and iPads. Due to the large capture space, a high-definition recording resolution of 1920 x 1080 pixels was utilised.

With MPP2SOS, the processing time for all cameras would be in excess of two hours, which while suitable for the production of high-quality outputs overnight, had lower utility during live rehearsal, where rapid feedback is valuable. As such, the lightweight VIBE model was used for quick previewing during rehearsal, and MPP2SOS was used for producing final polished animations. VIBE would take between approximately fifteen minutes to process a one-minute video clip from a single camera.

During a break in the rehearsal, VIBE was run on the footage from one of the devices. This allowed for the application of pre-prepared abstract animations onto the mesh derived from the captured footage. The use of VIBE and its fast processing capability served two primary purposes. First, it provided performers with prompt visual feedback on their movements, enhancing their understanding of how their choreography translated into animations. Second, it offered a platform to explore the creative potential of abstract animations as a complementary artistic medium. After a few minutes, the performers could view their performances imbued with abstract animations. This swift turnaround offered a tangible representation of how their movements transformed into visualisations and underscored the viability of using such tools in a performing arts setting, as it enabled faster exploration of a wider range of creative ideas. A video example of the abstract animation driven by the VIBE model is available here [vidStream_N]. To expedite the production of the animation, a preview render was generated with low-quality settings to facilitate a prompt review, as depicted in Figure 6.12.

Fig. 6.12: Low quality preview render.

6.4.3 Review of outputs

As the performance evolved to encompass suspension in aerial slings, the VIBE and MPP2SOS systems encountered challenges. While ground-based detection remained relatively stable, the models struggled to accurately detect poses once the performers left the ground and were suspended in aerial slings. This struggle was particularly evident in the animated skeletons' behaviour, which exhibited uncontrollable movement throughout the 3D space. The inability of both models to effectively detect and capture poses in the context of suspension in aerial slings shed light on their inherent limitations. It became apparent that the models were not trained to handle such scenarios, leading to erratic behaviour in their outputs. This test highlighted the boundaries of the models, showcasing their inability to adapt to unique and unconventional performance scenarios. While the review of outputs unveiled limitations in the VIBE and MPP2SOS systems, it also provided a deeper understanding of their capabilities. The successful ground-based detection and swift animation processing of VIBE demonstrated the potential for these models to enhance traditional performance settings. Yet, the challenges encountered in suspension scenarios underscored the need for further research, testing, and improved training data to accommodate a broader spectrum of performing arts scenarios. A video example of the glitchy detection of the performer on aerial slings is available here [vidStream_O]. In Figure 6.13, a frame is presented wherein the MPP2SOS model fails to generate an accurate pose detection of the performer on aerial slings.

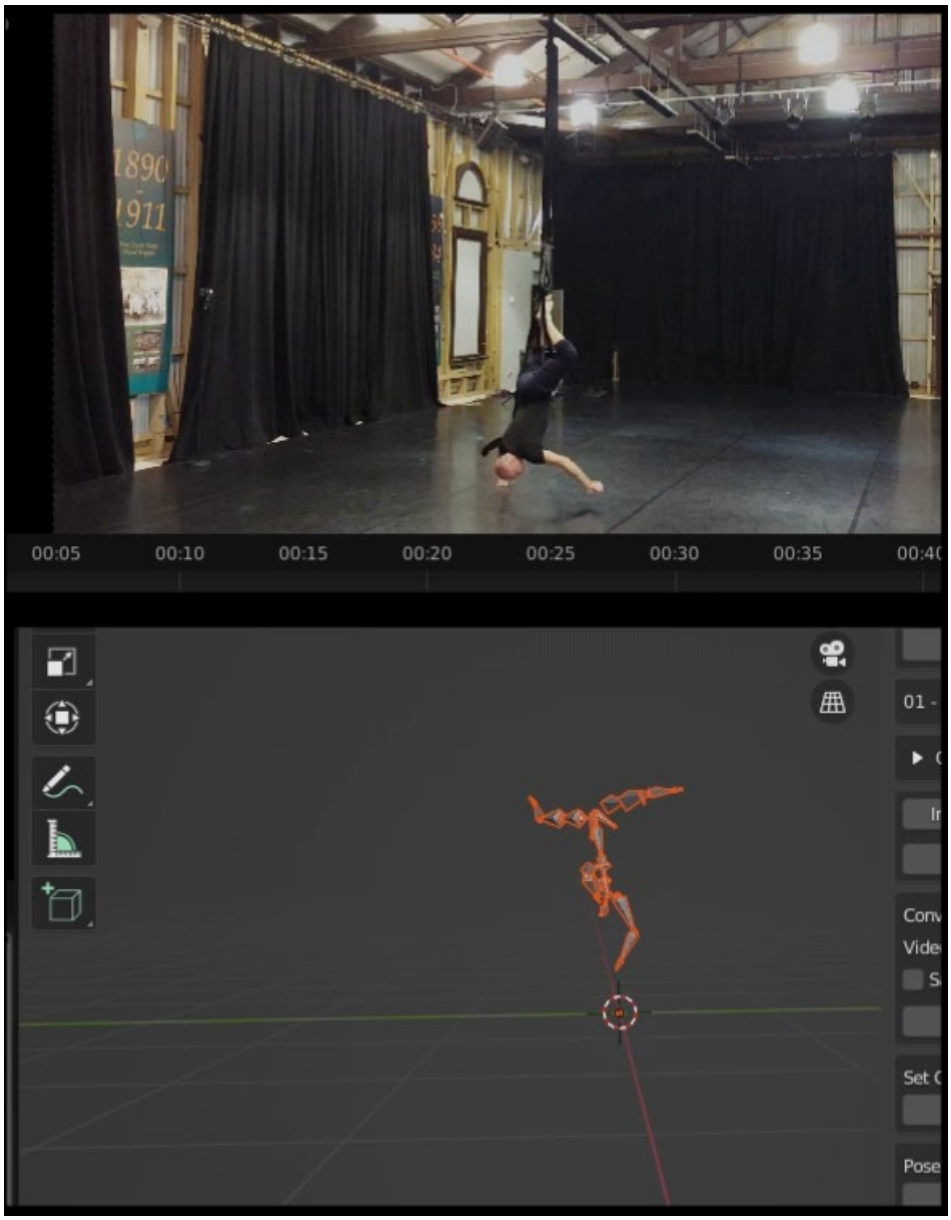

Fig. 6.13: MPP2SOS model detection glitch.

6.4.4 Summary of MPP2SOS findings

The exploration of MPP2SOS spotlighted the potential of this multi-camera pose detection system while revealing some limitations. For common ground-based movements, MPP2SOS demonstrated accurate motion capture capabilities, demonstrated in vidStream_M above. The modular architecture leveraging MediaPipe, Pose2Sim, and OpenSim provided robust body tracking suitable for many performance scenarios. However, MPP2SOS struggled to maintain tracking fidelity when presented with uncommon poses. The uncontrolled outputs for aerial movements highlighted the need for expanded training data and research to

handle unconventional cases. While showing promise for traditional stage performances, MPP2SOS requires further refinement to accommodate diverse contemporary styles.
The raw skeletal output of MPP2SOS necessitates additional skinning to achieve an animated character model. This presents an extra step compared to automated tools like EasyMocap. But with standard 3D workflows, the skeletal data can be readily adapted for creative needs. Overall, MPP2SOS exhibits strengths as a multi-camera pose detection system for many performing arts applications. With expanded capabilities to handle edge cases, it can become an even more versatile tool for creative motion capture. This exploration provided valuable lessons on pragmatic implementation within the realities of live rehearsals and productions.

6.5 Conclusion

The exploration of multi-camera pose detection provides valuable insights into the capabilities and limitations of current machine learning approaches for motion capture in performing arts. The investigation analysed two models, EasyMocap and MPP2SOS. Real-world testing under diverse conditions revealed the robustness but also the boundaries of these systems. Both models showed accurate body tracking for common movements in standard stage settings. However, limitations emerged for unconventional aerial poses, indicating the need for expanded training data.

EasyMocap's integrated character animation streamlines motion capture but encountered glitches requiring alternative models. In contrast, MPP2SOS provides raw joint data enabling customisation but requires additional post-processing. Camera calibration and manual interventions proved essential to maximise accuracy. While increasing complexity, proper calibration and supplemental data entry with camera calibration enabled quality results.

Experiments with the EasyMocap and MPP2SOS models highlight benefits and limitations for those who wish to integrate machine learning-based motion capture into live performance, addressing the question: 'What are the benefits, limits, and implications of current machine learning-based motion capture systems in the performing arts space?' (RQ2)

Benefits:

- Reliable capture of common human movements and poses for use in performances. Both models showed robust body tracking capabilities, aside from the glitches generated by EasyMocap.
- Potential for rapid visual feedback turn-around to performers by applying animations and visualisations driven by the captured motions. While MPP2SOS, being a multi-camera model, demands considerably more processing time compared to VIBE, employing both models together showed that VIBE's swiftness provided performers with a rapid preview, sparing them from enduring the slower processing time of MPP2SOS. This was demonstrated during the Box of Birds rehearsal.
- Customisable output skeletal data from MPP2SOS allows adaption and skinning for different performance needs.

Implications:

- Limitations capturing unconventional or aerial poses indicate a need for expanded training data and algorithm development.
- Glitches like those encountered with EasyMocap highlight unresolved intricacies in state-of-the-art models.
- Post-processing like skinning for final meshes and animations requires some technical expertise. Simplified tools would benefit non-technical practitioners.
- Proper camera setup and calibration is essential for quality results but adds complexity for end users.

This research showed that multi-camera systems appear to be viable for many performance contexts, but open challenges remain around handling atypical poses and simplifying workflows for non-technical practitioners. As algorithms evolve with larger datasets, multi-camera methods show strong potential for accessible yet accurate motion capture. The findings showed that to expand the value of these machine learning approaches for performing arts contexts, not just larger datasets, but a far greater diversity of training videos is likely to be required, including videos capturing distinct types of performance, movement, and body shapes.

While creating a custom training corpus for aerial and unconventional floor work would indeed be valuable, it represents a different research direction than the one pursued in this project. The deliberate choice to work with existing machine learning-based motion capture systems allowed us to examine and reveal their current limitations and assumptions through practice. Rather than attempting to create a better-trained system, this project focused on understanding and making visible how existing systems interpret non-normative movement patterns.

7 Results: Artefact Production

7.1 Introduction

This chapter explores a creative project that combines machine learning techniques and embodied artistic expression. The focal point of this enquiry is a collaborative performance, in which a dancer's movements are captured and translated into a visual narrative, made possible through the application of the multi-camera machine learning motion capture model MPP2SOS. This undertaking, grounded in the practice-based research framework, aims to scrutinise the process that navigates from the conceptualisation of the performance to its live execution while also analysing the perceptions of the audience and the performer, Cloé.

This investigation's initial phase involves conception and planning of the performance. This section delves into the interaction between the technical capabilities of the MPP2SOS system and the creative intentions of the performer. It outlines how the coordination of the choreography, music, and digital animations was managed in the pursuit of a seamless fusion of artistic elements. This portion emphasises the importance of interdisciplinary collaboration in shaping the overall vision of the performance.

Moving forward, the chapter shifts its focus to the live performance itself. Detailed insights are provided into the integration process, where motion capture-driven animations are synchronised with the dancer's movements and the accompanying music. The chapter analyses the harmony of these components, which produces a multi-dimensional artistic experience. The discussion illustrates how machine learning-driven motion capture can be used to create unique animations that compliment live dance performance. The study also addresses the audience's engagement with the performance. This section examines feedback from the audience members, shedding light on their interpretations and emotional responses to the performance.

Cloé's reflections provide a personal viewpoint on their experience with the performance. Cloé delves into their experiences with motion capture and compares them with the current performance involving machine learning methods. This comparison provides insights into their perception and offers a first-hand account of a deeper understanding of machine learning-based motion capture's effects on the Cloé's creative process and engagement.

Figure 7.1 below shows how research question two will be addressed through the creation of an artistic artefact.

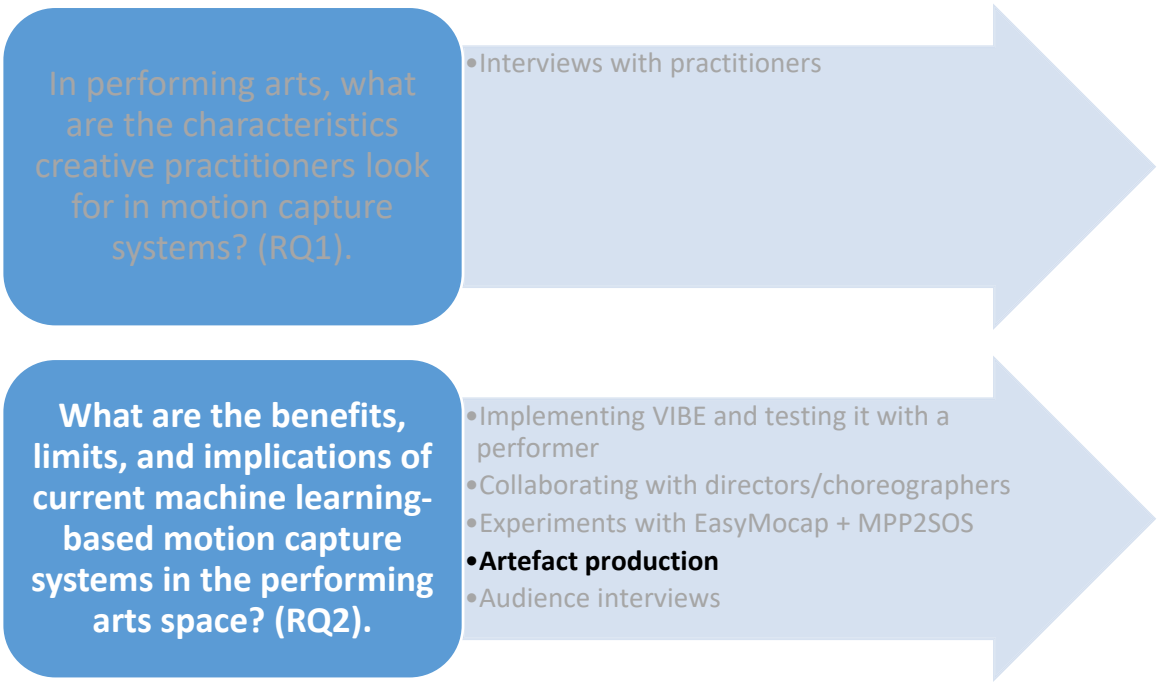

Fig. 7.1 Artefact production to address research question two.

7.2 Performance planning, preparation and execution

7.2.1 The abstract animation

The animation process began with experimentation, exploring various animation styles derived from motion capture data acquired from Mixamo[49] before it was captured from the performer. This exploratory stage allowed for the generation of diverse animation possibilities. The subsequent refinement of these styles became more defined as the performance's overarching theme and virtual environment was established. Houdini[50], versatile 3D animation software known for its procedural capabilities, emerged as a key tool in the animation production process. Its procedural nature provided an optimal platform for integrating abstract animations with the animated mesh of the performer.

[49] Mixamo, https://www.mixamo.com/
[50] SideFX Houdini, https://www.sidefx.com/

The procedural capabilities of the Houdini 3D application refer to its ability to generate and manipulate content algorithmically. This means that instead of creating objects or effects manually, users can define rules and parameters that govern their creation and behaviour. Houdini's procedural approach allows for the creation of complex and dynamic effects that can be easily modified and updated throughout the production process.

Four abstract animations were situated on plinths within a virtual art gallery environment, and each animation unfolded as a distinct visual story, closely intertwined with Cloé's movements. The animation style employed in the performance was influenced by both personal design preferences and external sources of inspiration. This selection process aimed to ensure a seamless integration between the animations, the virtual environment, and performer's movements, fostering a cohesive and captivating narrative. Inspirations from Entagma[51] and CG Artist Academy [52] contributed to the animation's technical intricacies and artistic nuances.

The choice of slowly evolving abstract sculptural animations was particularly well-suited to the characteristics of machine learning-based motion capture. Unlike traditional marker-based systems that provide highly precise point-based tracking, machine learning-based motion capture offers a more holistic but potentially less granular capture of movement. This technical consideration influenced the artistic direction, leading to the development of animations that emphasise fluid, continuous transformations rather than rapid, precise changes. The procedural nature of Houdini complemented this approach, allowing for the creation of animations that could gracefully adapt to and harmonise with the captured motion data's level of detail. The decision to work with abstract, sculptural forms also provided flexibility in interpreting and responding to the motion capture data, enabling an artistic expression that worked with, rather than against, the technical characteristics of the machine learning-based system. This synergy between technical constraints and artistic choices ultimately contributed to the aesthetic of the performance, where the animations appear to organically respond to and evolve with the performer's movements.

The conceptual foundation of the project began with the vision of creating a virtual art gallery within the unique circular projection environment of the UTS Data Arena. The choice of art gallery as environment is discussed in section 7.2.3. This setting influenced my creative approach, inspiring me to design an immersive experience that would blend physical performance with digital art. The spatial constraints of the Arena led to a creative decision to limit the performance area to roughly two square meters. This constraint allowed enough

[51] Entagma, https://entagma.com/
[52] CG Artist Academy, https://www.youtube.com/c/cgartistacademy

space in the Data Arena to the performer to dance, the audience to gather in one area and the animations in the virtual art gallery to be visible without being cramped.

In developing the animations, a conscious choice was made to move away from traditional keyframe animation in favour of procedural animation techniques. This decision was rooted in the desire to create a more organic and dynamic interaction between the dancer and the digital elements. One of the key creative concepts was the gradual reveal of a trail-like animation. This effect was envisioned as a way to surprise and engage the audience, creating a visual narrative that would unfold over time. The animation begins subtly, with just the outline of the dancer, and gradually builds into a kinetic sculpture, offering viewers a evolving visual experience that rewards continued attention.

Colour played a crucial role in the aesthetic decisions. For the trail animation, a colour scheme was developed that was directly linked to the speed of the dancer's movements. Slow movements produced blue trails, medium speeds resulted in purple to magenta hues, and fast movements generated yellow trails. This visual language was designed to help the audience intuitively interpret the performance, adding an additional layer of meaning to the dancer's movements. In creating the digital fabric animation, an artistic choice was to eliminate the effects of gravity. This decision was aimed at enhancing the dancer's movements while creating a smooth, flowing aesthetic that would complement the chosen music.

For the animations featuring spherical and cuboid shapes, the creative process involved finding a balance between abstraction and representation. A conscious choice was made to connect these shapes to various areas of the dancer's digital mesh, but deliberately made this connection less literal. By offsetting some shapes from the mesh, a more intriguing visual experience was created that wasn't immediately obvious in its relation to the dancer's movements. This decision adds depth to the performance, inviting the audience to discover the connections between the physical and digital elements over time.

In designing these animations, there was a focus on creating a sense of contrast and harmony. The juxtaposition of thin, elongated cuboids with spherical shapes was a deliberate choice to add visual complexity and interest. However, a cohesive colour scheme

was maintained across all animations, ensuring that while individual elements might contrast, the overall visual composition remained harmonious. Throughout the creative process, collaboration with the dancer was key. I shared my vision through test animations and music, which allowed the dancer to develop choreography that integrated with the digital elements.

The decision to work with a single performer against a static backdrop was deliberate and served multiple purposes. First, it allowed for a clear demonstration of the relationship between performer and animation - one body 'driving' the visual elements. This one-to-one connection made it easier for audiences to understand the interplay between physical movement and digital response. Multiple performers would have created a more complex visual environment that might have obscured this fundamental relationship.

The choice also had practical considerations. The Data Arena's limited space made working with a single performer more manageable. Additionally, using one performer helped maintain a relatively straightforward experimental process, allowing for focused development and refinement of the system. While introducing multiple performers could be an interesting direction for future work, establishing the effectiveness of the system with a single performer was an important first step in proving the concept's viability.

While the Data Arena offered 360-degree projection capabilities, the decision to maintain a single perspective was intentional. The multiple visualisations of the performer's movements, projected around the space, already provided different interpretations of the same movement data. Adding multiple viewing angles of these animations could have overcomplicated the visual experience and potentially confused audiences about what they were observing. The single perspective approach helped maintain clarity and allowed viewers to focus on understanding the relationship between the performer and the various animated interpretations of their movement.

1) Rectangular forms

Rectangular cuboid shapes attached to the mesh generated by the MPP2SOS machine learning model formed the foundation of this animation. As the Cloé moved, the cubes responded dynamically, their rotations and shifts synchronised with the dancer’s kinetic

expressions. This interplay suggests a fusion of the organic and the geometric, encapsulating the harmonious coexistence of divergent visual elements.

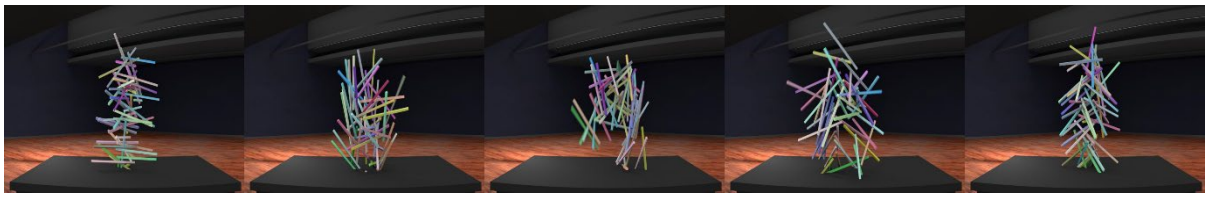

Fig. 7.2 Five sequential frames of animated cuboid shapes.

2) Motion trails

This animation emphasised Cloé's motion by creating trails left by the mesh's movement. These accumulating traces gradually coalesced into a static motion sculpture by the performance's conclusion. This portrayal over time via accumulated motion served as a representation of kinetic energy crystallised in a visual form.

The trails animation presents a distinct visualisation approach, functioning more like an evolving motion sculpture than a real-time animation. Unlike the other animations which respond instantaneously to the pre-recorded movements of the performer, the trails animation accumulates movement data over time, creating a more holistic view of the performance. This creates an interesting temporal contrast - at the start of the sequence, when the performer is relatively still, only their basic outline is visible. However, around 20 seconds into the performance, the accumulated motion trails become apparent as the system builds up the visualization of their movement pathways. This delayed, cumulative approach results in a slower, more contemplative visualization that reveals the entire journey of the movement rather than just its immediate state.

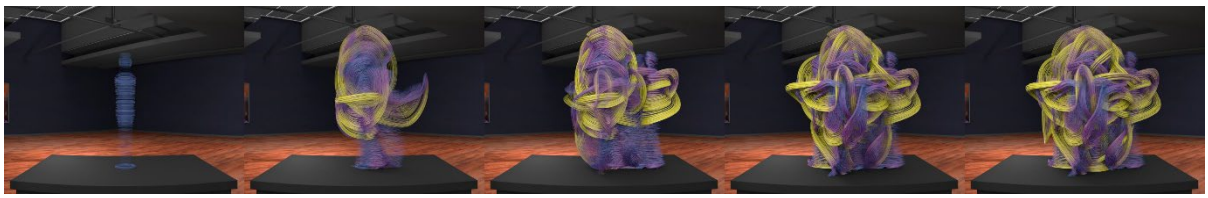

Fig. 7.3 Five sequential frames of animated motion trails

3) Floating strands

This animation manifested as delicate strands attached to Cloé's mesh. Suspended in the air, these strands followed the mesh's movements, evoking a sense of fluid connectivity. The interaction between the dancer's motions and the floating strands formed an intricate choreography in the virtual realm, highlighting the interplay between physical and digital forms.

Fig. 7.4 Five sequential frames of animated floating strands.

4) Orbiting spheres

Spheres, animated with subtle rotations, were interwoven into the mesh, punctuating the performance with a captivating motion. The harmonised dance of the spheres and the performer underscored the integration of motion and rhythm.

Fig. 7.5 Five sequential frames of animated orbiting spheres.

A video example of all of the animations together, including the performer is available here [vidStream_P]. In Figure 7.6, an illustration showcases four distinct types of abstract animation featured in the performance. These include rectangular forms (top left), motion trails (top right), floating strands (bottom left), and orbiting spheres (bottom right), with Cloé situated on the right.

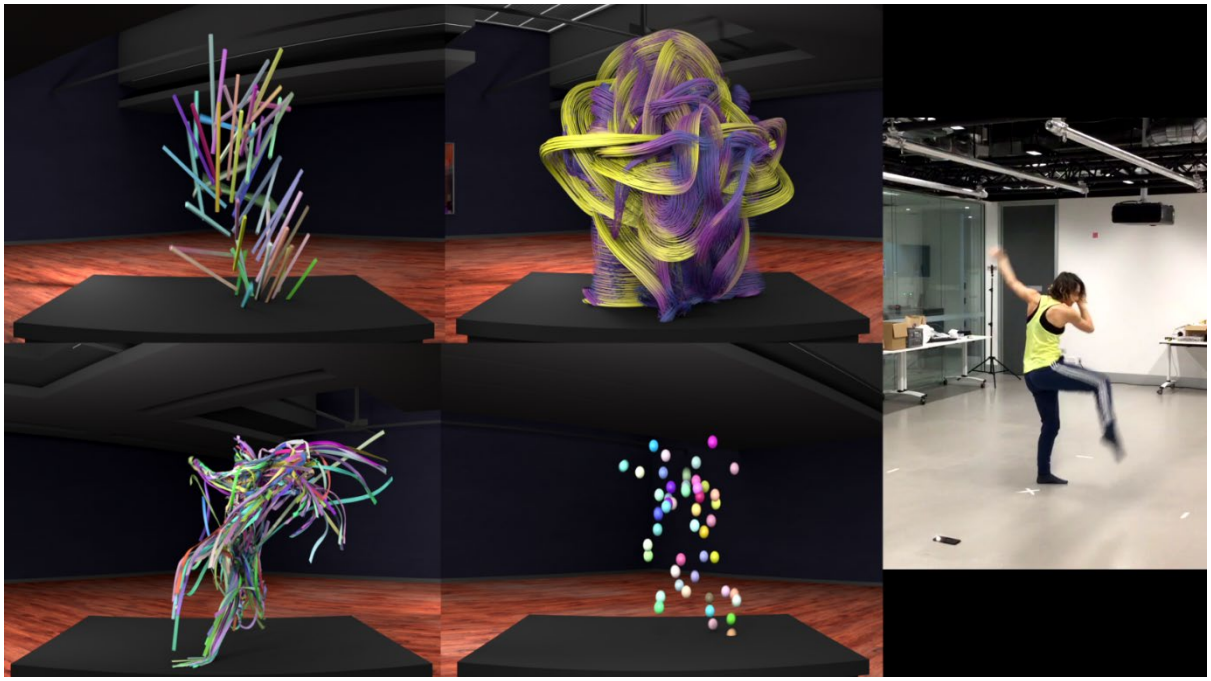

Fig. 7.6: Animation examples with performer, Cloé.

7.2.2 The Data Arena

The choice of venue is important in shaping the performance's overarching aesthetic. The University of Technology Sydney's (UTS) Data Arena, an innovative 360-degree interactive data visualisation facility, was selected as the backdrop for the performance. This cylindrical space, with a height of four meters and a diameter of ten meters, provided an immersive canvas that enveloped the viewers within a visually stimulating environment. The high-performance computer graphics system employed within the Data Arena drives a network of six 3D-stereoscopic projectors, edge-blended to produce a continuous three-dimensional panorama. Each audience member is equipped with active-shutter glasses to fully harness the stereoscopic effect, presenting distinct left and right views to achieve a stereo-visual effect. Figure 7.7 depicts the interior of the UTS Data Arena, where images are projected onto the surrounding screens.

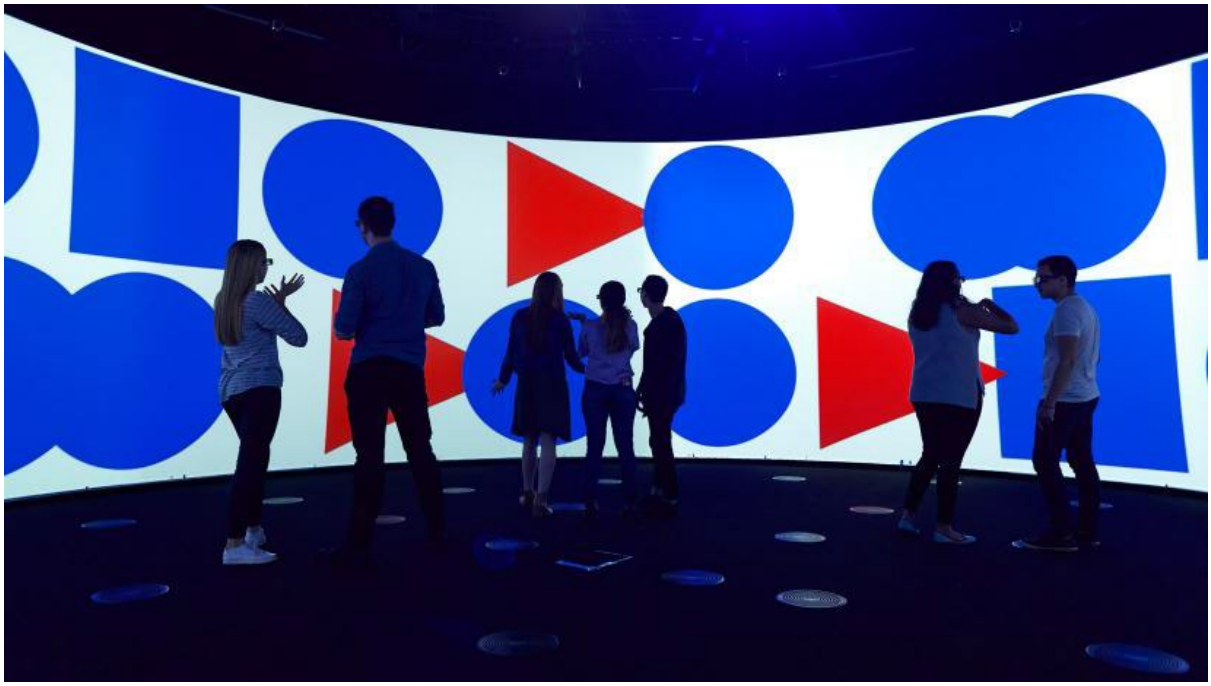

Fig. 7.7: The UTS Data Arena[53]

The immersion offered by the UTS Data Arena introduced a critical consideration: the positioning of Cloé in relation to the projected animations. While the venue provided impressive immersive audio-visual capabilities, careful planning was necessary to ensure a balanced presentation. It was determined that the Cloé's position would be directly opposite the entrance of the Data Arena, as that is where the audience would likely gather for the performance. To prevent any interference between Cloé's actions and the projected animations, a strategic decision was made to feature four animations – two on each side of Cloé. This arrangement ensured the animations would be visible to the audience without overshadowing the dancer's movements. The Cloé's proximity to the Data Arena's walls also required consideration due to the light projection. Performing too close to the walls could result in unwanted shadows and interference from the projector's light. To mitigate this, Cloé's was instructed to maintain a certain distance from the walls while performing, allowing the animations to be viewed clearly. Figure 7.8 shows a schematic diagram of the audience placement in the Data Arena in relation to Cloé.

[53] UTS Data Arena, https://www.uts.edu.au/discover/campus-future/data-arena

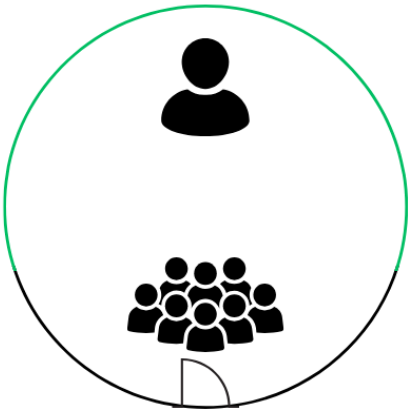

Fig. 7.8: A schematic diagram of the layout of the Data Arena. The audience was situated near the entrance opposite the performer (Cloé). The projection covered the entire room, but the animation covered the area in green.

The performer was given a cautionary instruction to avoid any dance moves that involved her body making contact with the floor, as a safety precaution. By only having her feet touch the ground, this would help prevent potential injuries from other body parts hitting the floor during the performance. This directive aimed to prevent potential injuries from accidental interactions with the air-conditioning vents. By excluding floor-bound movements from Cloé's choreography, the safety and well-being of the dancer was prioritised while ensuring the performance adhered to the constraints of the venue. Furthermore, all pertinent venue limitations were communicated transparently to Cloé to ensure a comprehensive understanding. This included the specifics of the floor vent placement and any other technical and spatial considerations that could influence the execution of the choreography. A collaborative approach was fostered by providing Cloé with a comprehensive overview of the venue's limitations, enabling the dancer to shape their performance to fit within the unique confines of the UTS Data Arena.

7.2.3 The virtual environment

The vision for the virtual environment revolved around creating an immersive space that could effectively showcase abstract animations as artistic pieces. Given the animation's likeness to sculptures within a still frame, the idea of an art gallery emerged as a potential

setting. By situating the animations within the virtual art gallery, each animation could be experienced as an exhibit, emphasising their visual and artistic qualities. An existing art gallery environment was selected from the Turbosquid[54] website to bring this vision to fruition. However, significant customisation was undertaken to align the chosen art gallery with the unique dimensions of the UTS Data Arena while extending the virtual environment beyond the Arena's physical boundaries. This extended dimension allowed the audience to perceive a continuation of the gallery outside the physical confines, enhancing the illusion of immersion. Figure 7.9 displays various perspectives of the virtual art gallery environment, arranged from top to bottom as north view, east view, west view, and south view.

[54] Turbosquid, https://www.turbosquid.com/

Fig. 7.9: The art gallery virtual environment.

The virtual art gallery was designed to evoke realism while accentuating its aesthetic allure. Wooden floors and high ceilings replicated the grandeur of an actual gallery. The paintings on the walls were the most distinctive elements in this virtual environment. The abstract paintings were generated using the DALL-E 2 (Ramesh et al., 2022) AI image creation tool.

The static gallery background was a deliberate aesthetic choice that served to frame the animations conceptually as moving sculptures. By using a minimal, gallery-like setting, the animations could be viewed as dynamic artworks in a curated space, rather than theatrical scenes. This understated backdrop ensured that viewers' attention remained focused on the relationship between the performer and the resulting animations, without competing visual elements in the environment that might distract from these core elements.

7.2.4 Choreography design

Once the animation styles had been established, the next step was to translate this vision to the performer/choreographer, Cloé Fournier. Since Cloé was the performer involved in prior experiments, her previous experience would have shaped her comprehension of the process and the potential capabilities of machine learning approaches. The selected animation styles were shared with Cloé, to help her conceive the overall performance aesthetic and she was invited to the UTS Data Arena, where both the virtual environment and temporary animations would be projected. This visual preview offered her a tangible representation of the forthcoming performance, enabling her to envisage the choreographic interaction between her movements and the projected animations. The musical element of the performance was also important in shaping the choreography. The chosen piano piece (Lightfoot, 2017) was selected to mirror the ambience of the art gallery setting, provided further context for the dancers’ movements.

The spatial constraints of the Data Arena significantly influenced the choreographic development. With a performance area of approximately two square meters, Cloé needed to craft movement that would be impactful within this confined space while remaining aware of the surrounding projected animations. The choreography had to balance between utilizing the available space effectively and maintaining appropriate distance from the projection surfaces to avoid casting shadows.

To facilitate the rehearsal process, Cloé was provided with an online video link featuring the animations synchronised with the chosen musical composition. The animations provided are abstract animations applied to motion capture data from Mixamo, solely for the performer to visualise how the animation appears with motion. This arrangement allowed the dancer to

experiment and rehearse within the comfort of their own space, enabling them to explore the choreography’s dynamics and keeping the space of the Data Arena in mind. The music's timing to the performance's length provided a structured framework for the dancer to refine their movements, ensuring temporal coherence. Safety considerations also shaped the choreographic choices. Specifically, floor work was limited to foot contact only, avoiding movements that would bring other body parts in contact with the ground due to the presence of air conditioning vents in the Data Arena floor. This constraint led to a focus on upright movement vocabulary that emphasized upper body articulation and standing positions.

The final choreography was designed to complement – rather than compete with – the projected animations. Movements were crafted to both initiate and respond to the various animation styles - flowing motion trails, geometric forms, floating strands, and orbiting spheres. This careful integration helped create the fusion of live performance and digital elements.

After a few weeks of rehearsal and preparation, the performer was poised to undertake the motion capture recording. During the rehearsal, the performer practiced the performance with the provided music track. No motion capture was conducted at this stage. The choreographic journey, characterised by artistic alignment, experimentation, and rehearsal, converged when it was time to commence the motion capture recording. The carefully designed choreography, intricately woven with the musical accompaniment and animations in mind, laid the foundation for the forthcoming collaborative step in creating the performance.

7.2.5 Recording the motion capture

The dimensions of the performance space within the UTS Data Arena were carefully noted to ensure the accurate translation of the dancers' movements. The boundaries were marked on the floor, enabling the performer to stay within these confines during the motion capture process. This spatial precision was integral to maintaining consistency between the dancer’s live movements and the subsequent animations. The performance creation sequence involved recording the performer's motion capture data, applying abstract motion to this data, and then projecting the abstract animation in the Data Arena while the performer executes the same choreography as captured in the motion capture.

Seven cameras played a pivotal role in capturing the performer's movements. An eighth camera with a fish-eye lens was also trialled as an experiment. However, this camera introduced undesirable distortions and blurriness along its edges. To maintain the accuracy and clarity of the data, the experimental camera was replaced with a GoPro[55] camera equipped with a wide-angle lens. This pragmatic adjustment ensured the integrity of the captured motion data, safeguarding against distortions and maintaining visual fidelity. Figure 7.10 displays the impact of a fish-eye lens on one of the devices, resulting in noticeable blurring and significant distortions. Figure 7.11 illustrates the arrangement of cameras to capture the motion for the performance. A variety of capture devices are utilised, spanning from an iPhone 4 to an iPhone 13, as well as a GoPro Hero 8 action camera and iPads. This setup serves as a valuable test to determine the feasibility of achieving high-quality motion capture for abstract animation using affordable equipment.

Fig. 7.10: Fish-eye lens.

[55] Go Pro, https://gopro.com/en/au/

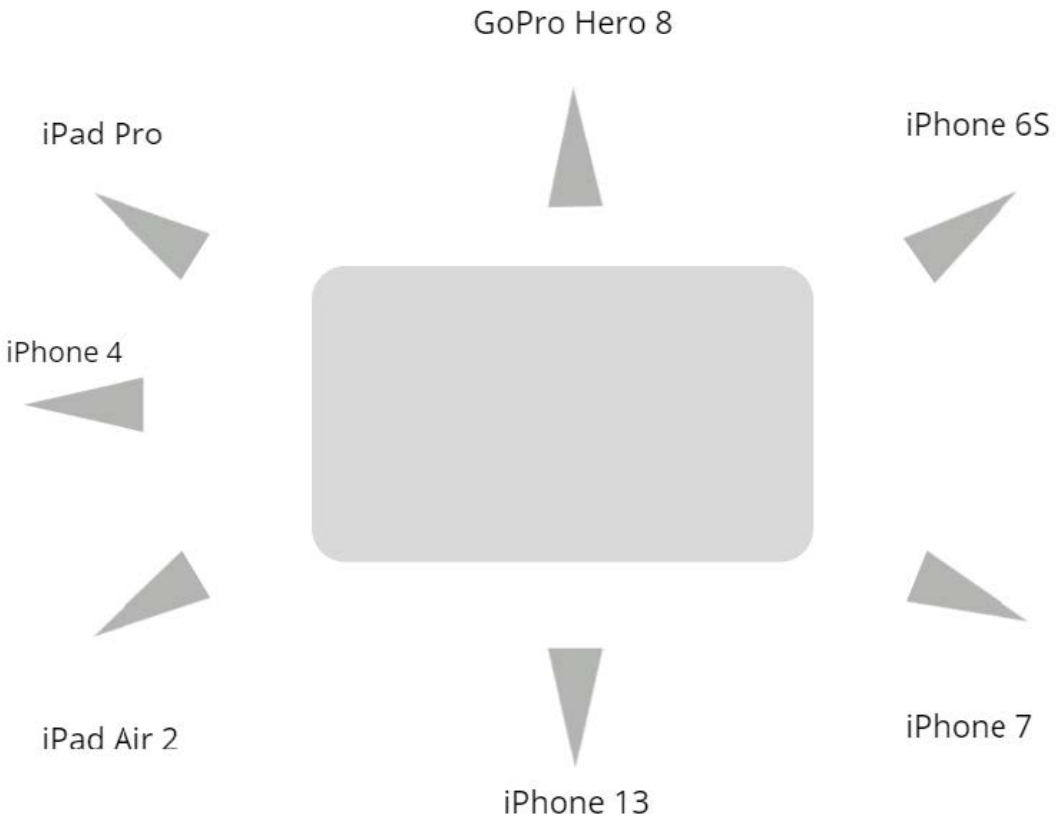

Fig. 7.11: Camera layout for performance.

7.2.6 Running the model

Thanks to prior testing and optimisation, the application of the MPP2SOS model proceeded smoothly. The model, designed to convert the performer's movements into a skeletal structure, provided a foundational framework instrumental in driving the subsequent animation process. While the MPP2SOS model offered a skeletal representation of the performer's movements, the animation required a mesh as input. This necessitated the manual application of a mesh to the skeletal structure. A refinement phase followed the initial application of the skeleton and mesh. The skeleton needed further adjustments to ensure temporal coherence. Manual edits were applied to areas where arm movements exhibited slight inaccuracies, fine-tuning the animation to better align with the original performance. There was, however, a realisation that the animation need not be flawlessly aligned with the original performance due to its abstract nature. In Figure 7.12, an overlay of the refined mesh on the original video utilised for motion capture is presented. Figure 7.13 depicts different frames of the motion capture overlaying the performer. The grey mesh on the left side of the performer represents the raw pose detection, while the areas highlighted in red

indicate the regions that required refinement to align with the performance. On the right side of the performer is the refined mesh.

Fig. 7.12: Mesh overlayed on input video.

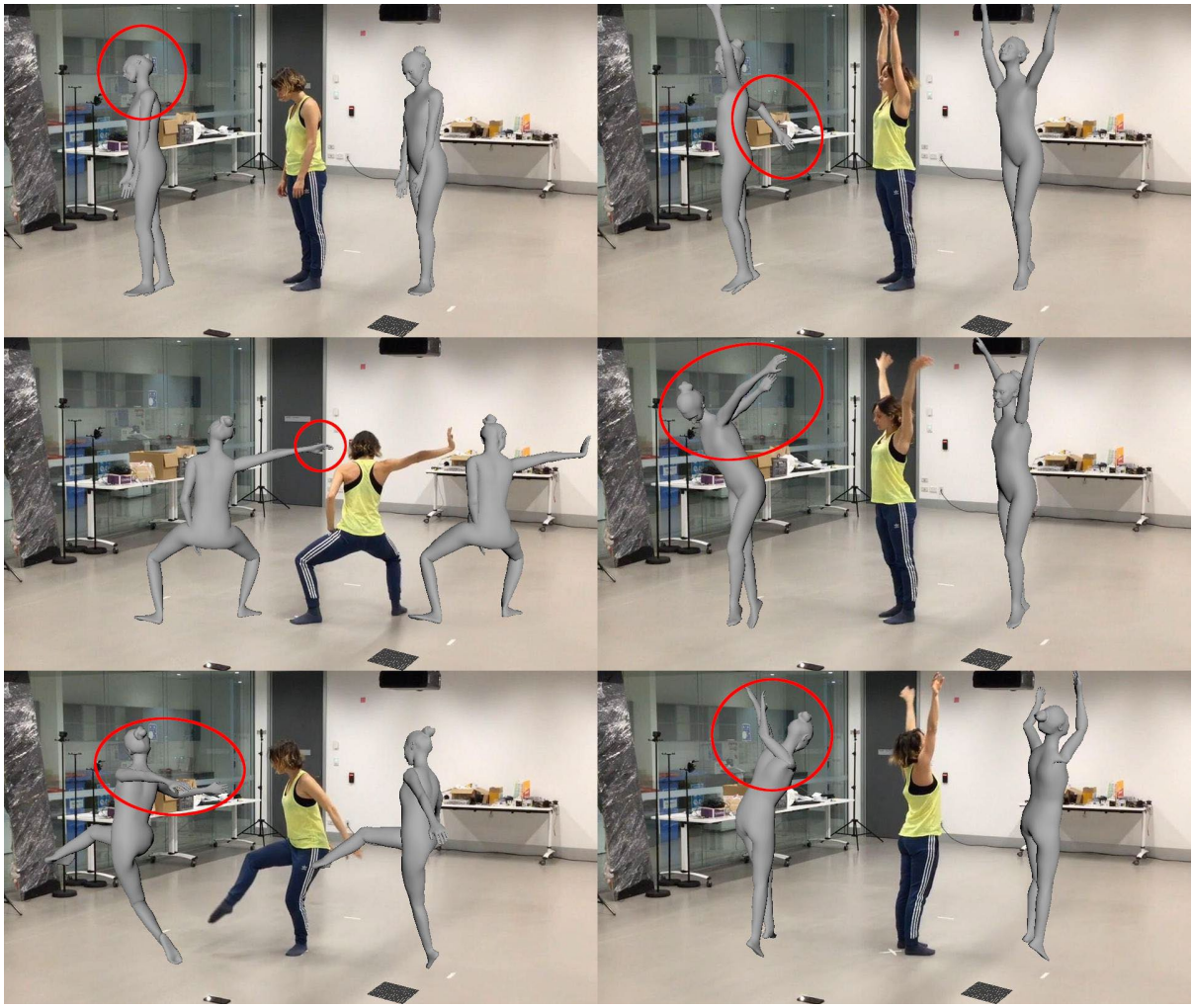

Fig. 7.13: Raw detection and refined mesh comparison. Each frame depicts the raw detection (left of the performer) and the refined mesh pose (right of the performer). Highlighted in red are the areas that needed to be refined.

7.2.7 Testing in the Data Arena

Projection in the Data Arena required a resolution of 10576 x 2400 pixels per eye. With the performance comprising 6132 frames (3 minutes and 24 seconds at 30 fps), the rendering resulted in a file exceeding 50 gigabytes in size. The files were downloaded and subsequently compressed by using the FFmpeg[56] software to allow smooth playback. The tests illuminated subtleties that necessitated adjustments, most notably the impact of the stereoscopic glasses on image quality. The glasses introduced a darker and less saturated appearance than intended, prompting colour grading and lighting adjustments to achieve the intended visual aesthetic.

[56] FFmpeg, https://www.ffmpeg.org/

The computational power required to render the animations and the virtual environment in a manageable time frame was considerable. The render farm at the UTS Animal Logic Academy [57], housing an array of 62 computers, became instrumental in this endeavour. To optimise the rendering further, a pragmatic approach was employed where only one frame of the static art gallery environment was rendered and composited beneath the animations, significantly reducing the rendering time. Despite these optimisations, the rendering process spanned approximately 12 hours. The table below lists the specifications for the computers on the render farm.

Product line	Dell PowerEdge C6525
CPU	2 x AMD EPYC 7552 2.20GHz
Memory (RAM)	512 GB
Storage	800 GB SSD

Table 7.1: UTS Animal Logic render render farm. machine specifications

7.2.8 Rehearsal

The rehearsal phase enabled the performer to refine their choreography and ensure that her timing was in sync with the animation and music. The dynamic interplay between the performer's physicality and the virtual narrative was fine-tuned, emphasising the artistic synergy between the performer and the projections. An integral aspect of the rehearsal was finalising lighting arrangements within the Data Arena. Three programmable spotlights were strategically positioned to illuminate the performer from various angles, creating a focused and dramatic effect. The lighting scheme emphasised the performer's presence within the immersive environment, accentuating their movements and distinguishing them from the animated backdrop. A video example of the performer rehearsing with the animation and music is available here [vidStream_Q]. Figure 7.14 depicts the performer rehearsing alongside the animation.

[57] UTS Animal Logic Academy, https://www.uts.edu.au/about/faculty-engineering-and-information-technology/animal-logic-academy

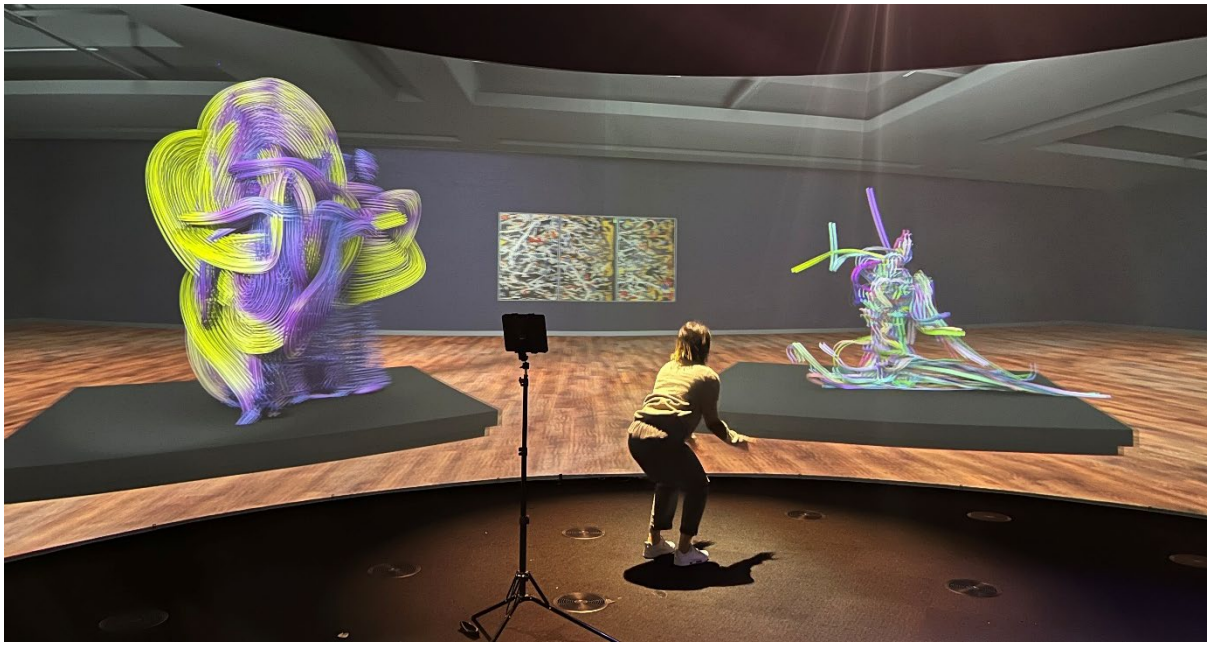

Fig. 7.14: Performer rehearsing.

An iPad displayed a reference to the performer’s motion-captured choreography to enable the alignment of their movements with the intended narrative. The reference acted as a guide in rehearsals, ensuring a cohesive fusion between the live performance and the animations within the immersive environment.

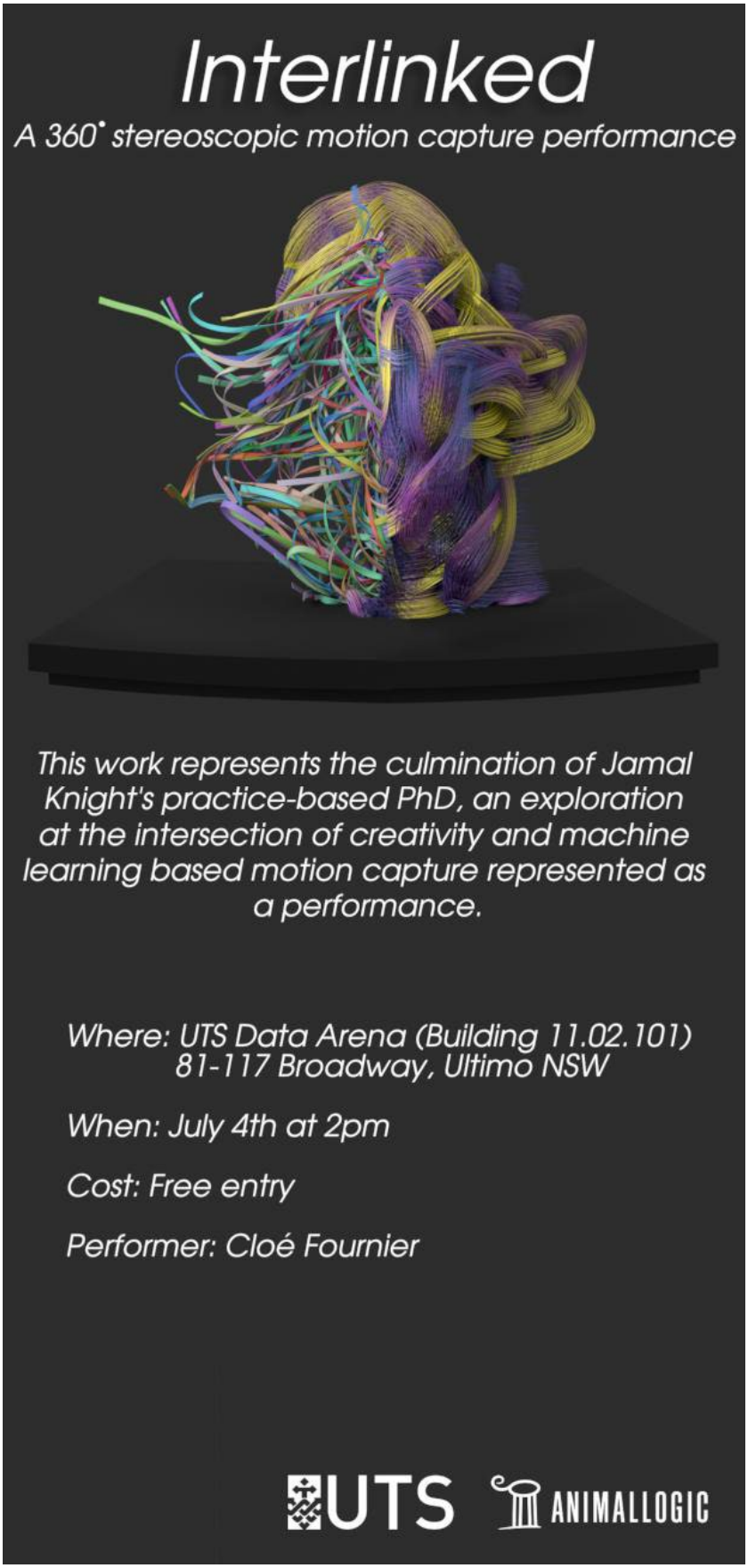

Fig. 7.15: *Interlinked* flyer.

7.2.9 Performance

Attended by an audience of 17 individuals, the performance was screened twice to ensure every audience member could absorb the experience. The audience was invited to participate in a short interview about their experience after the screenings. The performance was synchronized with prerecorded music, which served as Cloé's cue to begin. Prior to the live performance, Cloé had rehearsed with a video that combined the projected animations and music track. During the actual performance in the Data Arena, once the audience was seated, I operated the playback system from the back, ensuring the animations and music were synchronised from the start.

Being deeply involved in both the technical and creative aspects of this project provided a unique perspective on the performance's execution. During the performance, I observed subtle incongruencies between the animation and the performer's movements, particularly in timing and spatial relationships. These minor discrepancies, while perhaps not immediately apparent to the audience, highlighted the importance of extensive rehearsal time in achieving seamless integration between pre-recorded motion capture animations and live performance. A longer rehearsal period would have allowed the performer to better synchronize with both the musical score and the animated elements, potentially resolving these timing inconsistencies. This insight suggests that future projects utilizing ML motion capture for pre-recorded performances should allocate substantial rehearsal time for performers to familiarize themselves with the virtual elements and develop a more intuitive understanding of the spatial and temporal relationships between their movements and the animated responses.

A 360-degree video example of the performance is available here [vidStream_R]. A short clip of the performance is available here [vidStream_S]. Figure 7.16 captures the performance within the Data Arena.

Fig. 7.16 Data Arena performance.

7.3 Audience interviews

A series of post-screening interviews were conducted to gather feedback on audience engagement during the performance. Three fundamental questions guided the interviews, each designed to elicit a particular facet of the audience experience.

The figure below shows how research question two was addressed by conducting interviews with audience members who attended the performance.

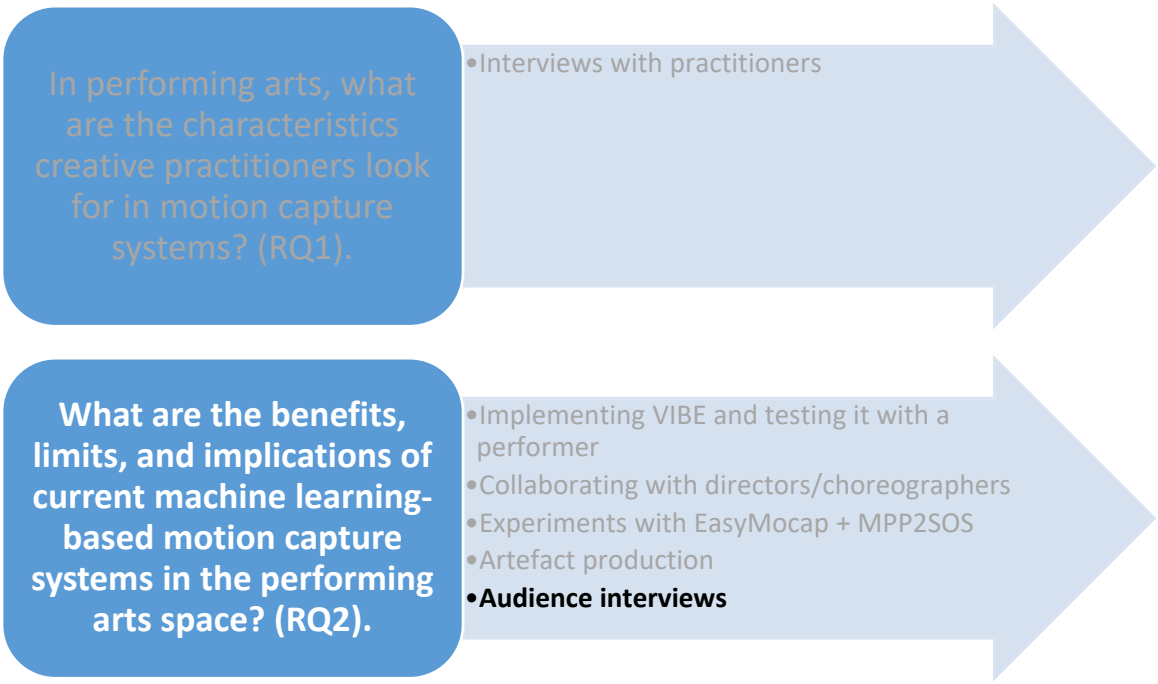

Fig. 7.17 Audience interviews to address research question two.

The first audience question invited attendees to share their immediate thoughts about the performance, shedding light on its emotional impact and visual resonance. The main purpose of this question was to gain insight into the initial impressions into audience engagement. Attendees' immediate thoughts allowed me to gain a better understanding of the first impressions of the performance and how effectively it captured the audience attention from the outset.

Audience question one:

- How would you describe the overall experience of the performance?
    - Were there any specific moments in the performance that stood out to you and why?

The second question aimed to explore the relationship between the live performance and the projected animations. The question was a two-part question asked in a single inquiry. The intention was to ascertain whether the audience discerned whether the performance resulted from live motion capture or a pre-recorded animation. The responses to this question provided insight into the success or otherwise of blending technology and live artistry

seamlessly. Attendees' descriptions provided feedback on whether the two elements complemented each other and contributed to a cohesive artistic experience.

Audience question two:

- Can you describe how you perceive the relationship between the dancer and the animated projection?

The final question invited attendees to contextualise the performance within their broader experience of dance presentations. By prompting a comparison to other dance performances, this question probed the qualities that set this collaborative piece apart from other performances.

Audience question three:

- How would you compare this performance to other dance performances you have seen?

The interview questions were deliberately designed as semi-structured and open-ended to allow audience members to express their perceptions without excessive guidance. Particularly, the second question, "Can you describe how you perceive the relationship between the dancer and the animated projection?" was specifically crafted to probe deeper into the audience's understanding of the performer-technology alignment. This approach allowed participants to articulate their observations in their own terms, rather than responding to more prescriptive questions that might have influenced their responses.

The questions maintained a careful balance between structure and openness, enabling audience members to naturally focus on what resonated most strongly with them. This semi-structured format also provided the flexibility to explore interesting themes that emerged during the interviews through follow-up questions. By avoiding overly specific or leading questions, we sought to capture authentic audience responses that reflected their genuine engagement with the performance, whether positive or negative.

Results

The responses to the first audience question, seeking their thoughts about the performance, were mostly positive and elicited emotional responses. One audience member described their reaction as "I found it quite meditative… it was, like, peaceful and reflective, but still,

playful.” The term “meditative” emerged repeatedly, indicating that the performance invoked a tranquil and reflective quality. Another individual expressed, “The music was sad, but it was just kind of mellowing. It was meditative.” This recurring reference to a meditative quality underscored the performance’s ability to create a contemplative atmosphere that resonated with the audience. The meditative quality that audience members noted may have emerged from the intentional interplay between multiple elements of the performance. While the music did contribute to this atmosphere, as one audience member observed, I concluded that the overall meditative experience arose from the combination of Cloé's choreography and the musical accompaniment. The slow, deliberate nature of Cloé's movements, when paired with the contemplative musical score, likely created this meditative effect.

The overall feedback was positive, with one participant summarising their experience as “It was fantastic... Pretty amazing in terms of the project”. The auditory component also received positive feedback for the most part, with comments like “quite beautiful… like for the piano, it was gorgeous.” However, it is important to note that one participant found the audio to be somewhat distracting and perceived the quality as being lower than expected. This could be due to one of the speakers near this participant being slightly distorted.

The second question, designed to gauge the audience’s perception of the relationship between the performer and the projected animation, yielded a spectrum of insights. Some participants believed that the animations resulted from live interaction, assuming a direct connection between the performer and the projected animation. Comments such as: “I just liked how the figure mimics the dancer… (It felt) cohesive, like moving as one” and “(the animation was) obviously tracking like her body” underscored the successful integration that fostered the illusion of a cohesive, synchronised performance. While some attendees believed that the animations were live, others were intrigued by the technicalities and unsure. Comments like: “I was quite curious as to how that was done… there was a bit of a lag in terms of what was happening, but if it were to be real-time, then I would say both dancer and the animation is like one” and “(I was) trying to figure out how the performance was done behind the scenes. So, seeing your performance in real-time is quite impressive” illustrated curiosity and the desire to unravel the technical intricacies that blurred the line between live and pre-recorded elements. One participant noted a distinction in the perception of different animations and thought that some animations were pre-recorded and some were real-time. They observed: “… the other two were real-time, it felt like real-time except for the long wavy one (animation).” Interestingly, one participant astutely recognised

that the animations were pre-recorded. Their insight: “My assumption was that it was a pre-choreographed dance that was being performed again, but had previously been used to, like recorded and used to make the avatars. So, I didn’t interpret it as being live or real-time,” provided a distinct perspective that acknowledged the pre-recorded nature of the animation, highlighting the blend of perceptions within the audience.

The second research question, ‘What are the benefits, limits, and implications of current machine learning-based motion capture systems in the performing arts space?’ (RQ2), is partially addressed by the audience's response to the second question, as some attendees noted their unawareness that the projected animation was pre-rendered and not directly controlled by the performer during the performance. This also suggests that real-time processing speed may not be a crucial factor for this type of performance.

There is an intriguing aspect regarding how audience members appeared to be curious about the technical aspects of the performance, wondering about how it was executed. This curiosity might create a certain level of detachment from the performance itself, which could be considered sub-optimal as it potentially distracts from the intended artistic experience. However, this curiosity also presents an opportunity for exploration or exploitation in future performances. By incorporating elements that address or play with audience curiosity, performers could enhance engagement and create a more immersive and interactive experience. Thus, while the distance generated by curiosity may pose challenges, it also opens doors for innovative approaches to audience interaction and participation in the performing arts. Additionally, it could imply that such performances might prompt audience members to contemplate the role, function, and behaviour of technology.

The third and final question of the audience interviews prompted a comparison between the witnessed performance and their past encounters with dance presentations. This enquiry aimed to contextualise the immersive experience within the landscape of their artistic encounters, drawing insights into the performance’s novelty and impact. One participant noted the influence of the setting on their perception, stating, “The setting in the art gallery, I think, made me assume that this was art rather than performance”. This observation highlighted the ability of the performance to challenge traditional categorisations. A positive sentiment emerged regarding the visual elements of the performance. A participant remarked, “[The visuals] just elevates the dance performance because it’s happening at the

same time, so you have lots to look at.” This acknowledgement of the co-incidence between the dance and the animations highlights how the integration of technology could potentially enrich the viewer’s experience. This performance marked one attendee's first exposure to this blend of dance and technology. As they shared, “This is probably the first one (performance) I’ve seen the dancer dance with other animated figures… So, it was really, really cool to see them work together.” This comment highlighted the performance’s role in pushing artistic boundaries and introducing novel experiences to at least some audience members. The responses to this question revealed comparisons and reflections, providing insights into how the performance resonated within the framework of the audience’s artistic encounters.

The various animation styles were specifically designed to highlight different aspects of this technological relationship. Some animations responded in real-time to the performer's movements, while others, like the motion trails, had an intentional delay. This variety of response times and visualization styles provided multiple entry points for audiences to understand and question the technology's operation.

Other design elements that supported audience engagement included the choice of the Data Arena as a venue, the use of stereoscopic projections, and the integration of music with movement. These technical and artistic choices were made with awareness that audience members attending technology-enhanced performances could be interested in the underlying technical processes.

This curiosity about the technical aspects was further explored through post-performance interviews, where questions like "what is the link between the performer and the animations, and how is that link established?" allowed audiences to articulate their understanding of these connections. This approach acknowledged that understanding the technical "how" is often a key part of audience engagement with technological performances, while still maintaining focus on the artistic experience.

7.4 Performer interview

Designed to amplify the understanding of motion capture's role in performing arts, this section delves into the performer's (Cloé) past experiences in motion capture performances. The interview process uncovers their involvement in various motion capture systems, delves into the methodologies of capturing motion data, and concludes with comparing traditional methods to the novel machine learning techniques deployed in the recent performance. It is important to note that thematic analysis was not used for this interview. The aim of conducting an interview with Cloé after the performance was to gather her insights on the process and to compare her experience with more traditional methods of incorporating motion capture in performances, without the use of machine learning. The interview took place via a Zoom[58] video call a few weeks following the performance.

Previous motion capture systems

Cloé, who boasts a wealth of experience in performing with technology, spoke about her involvement with diverse motion capture systems. She highlighted her collaboration with the dance production company Box of Birds, where they interacted with large-scale motion capture systems. The systems that she was involved with included LiDAR scanning and the use of Intel RealSense[59] cameras which mapped the environment and performer in three dimensions. Cloé has been involved in projects using various real-time motion capture systems, each of which offered distinctive challenges and opportunities. A noteworthy instance entailed a performance that engaged augmented reality and required the audience to wear augmented reality glasses. The live performance featured two performers: Cloé, who was visible to the audience, while the other remained concealed behind a wall of mirrors. The audience, through their augmented reality glasses, enabled them to see Cloé and a digital projection of the concealed performer. Although the projected performer appeared distinctly digital, they interacted with Cloé in the performance space. This interaction was facilitated by a 3D camera capturing the concealed performer's image, which was then projected onto the audience's glasses.

Cloé recounted: "This one was super tricky just because you have to split your brain in so many different ways. And whatever you know about choreography, you have to reconsider completely because you have the real space and you have the, what I'm going to call, the

[58] Zoom, https://zoom.us/
[59] Intel RealSense, https://www.intelrealsense.com/lidar-camera-l515/

virtual space, but then you have somehow an in-between space, I don't really know how to call it so it's a hard one." This description vividly portrays the complexities and cognitive demands of executing a real-time performance within an intricate technological landscape.

Additionally, Cloé's international collaboration with a dancer in The Netherlands showcased their adaptability and proficiency in real-time motion capture collaborations. As they recounted, the intercontinental synergy involved dancing in Australia while simultaneously interacting with a distant counterpart overseas. The international collaboration brought its own set of challenges. Dancing remotely with a partner in Holland required recalibrating the choreography to accommodate the dancer's smaller performance space constraints. The performers account, "Then there was a live performance on the Netherlands side… you have to hit that particular point, that you have to hit otherwise you get out of the frame, and the image can't be projected properly to the Australian side," underscored the meticulous spatial awareness and choreographic adaptability often required in performances with a strong reliance on technology integration.

The process of capturing motion relating to traditional motion capture systems

Cloé mentioned that in her past experience with motion capture systems, the process of motion capture often involved a substantial element of trial and error. She shared, "There was a lot of trial and error, I guess. Because it was something we didn't know." This sentiment highlights the iterative nature of working with motion capture systems. In their pursuit of creating a seamless fusion of movement and technology, experimentation was intrinsic to refining the choreography and optimising the use of the technology.

Interestingly, Cloé shared instances where pre-recorded motion capture was employed. This strategy allowed them to address technical issues that posed challenges in real-time performances. Her reflection, "There was some areas where we use pre-records, because we couldn't do it in real-time", illuminates the pragmatic decisions taken to overcome technical obstacles and ensure a fluid artistic outcome when deploying motion capture for performance.

Comparison of traditional methods of motion capture to machine learning methods

During the interview, Cloé revealed the distinctions between the traditional methodologies of motion capture and the machine learning methods used for the collaborative performance *Interlinked*. A marked departure from the conventional practices was evident in *Interlinked*,

as she attested, “what we did, which was more about filming the body, so it was about having the choreography that stays and then capturing it then you went to do your amazing work with the graphics. And then for me to basically do what I had prepared and do it in real-time with the sculptures, which was probably the easiest.”

During the motion capture phase of performance preparation, Cloé's movements were captured and translated into motion capture data. In the actual performance, Cloé replicated the choreography captured during motion capture, but the projected animation was not directly linked to her movements as it was pre-recorded. This approach has both advantages and disadvantages. On the positive side, Cloé could accurately replicate the rehearsed movements, ensuring synchronisation with the music and animation. However, a potential downside is that any discrepancies between Cloé's movements and the pre-recorded animation would be noticeable to the audience. Additionally, using pre-recorded animations allows for high-quality render settings, which may not be feasible in real-time performances due to the processing requirements and resolution limitations of the Data Arena projection.

A noteworthy point was the perceptual shift in awareness and synchronisation between Cloé and the animations. In non-machine learning-based methods of motion capture, the performers' movements are captured and emulated by technology, requiring careful choreographic planning and post-performance refinement. However, in *Interlinked*, Cloé's awareness was oriented towards the animations, ensuring synchronisation between their movements and the graphics. As Cloé phrased it, “The (digital) sculptures didn’t have an awareness of me. I had an awareness of them, so that’s where the difference was. So, if I were late, the sculpture would not catch up with me. It would just continue on.”

The duality of this dynamic is captured in the Cloé's commentary, “So I was in a way I had the upper hand in terms of, like, the timing, but I was very much also at the mercy of it. So, depending on which angle you look at it.” This observation underscores the reciprocal interplay where awareness influences synchronisation and vice versa, culminating in a complex interaction that explores the performer’s relationship with technology.

7.5 Discussion

This chapter followed the journey of conceiving and realising a new performance piece, *Interlinked*, by integrating machine learning-based motion capture. The performance was examined by gathering audience feedback and in an interview with the performer. The

interviews with the audience provided positive feedback, reflecting an appreciation for the work's meditative and reflective qualities, and indicating a successful fusion of technology and art. Such feedback demonstrates that using machine learning methods in motion capture can result in captivating and resonant performances that engage the audience in unique ways.

While the Data Arena contributed to the performance's production, it's essential to acknowledge that not everyone can access such facilities. However, it should be noted that the venue for Interlinked can be any location capable of projecting animation. Furthermore, it's important to underline that Interlinked was created using accessible tools – a few tripods, consumer devices and open-source software, demonstrating that innovation in performance does not necessitate extravagant resources.

Additionally, the artefact's association with the specialised venue of the Data Arena does not constrain its adaptability. The demonstrated process could be translated to diverse venues, whether big or small, providing a versatile framework for practitioners in the performing arts.

It is also important to note that while machine learning methods bring innovation and opportunities, they are not devoid of limitations. As evidenced by Cloé's insights, synchronisation and awareness remain subjects of consideration, at least for non-real-time based approaches. It is worth noting that in principle, both the motion capture and the generated animations could run in real-time if using a single camera for capture, although this would compromise the quality of the animation.

The scope of this thesis intentionally encompassed a broad range of implications for motion capture technologies in dance environments, with choreography being one of many important aspects explored rather than the central focus. While choreographic considerations were addressed, they were examined alongside numerous other technical, practical, and artistic elements that contribute to the integration of these technologies in performance contexts.

The interview approach with Cloé used open-ended questions that allowed her to focus on what she found most significant in her experience with the system. Her responses naturally gravitated toward technical and logistical aspects, and this emphasis reflects her concerns and interests during the process. Had she chosen to elaborate more on choreographic elements or how the technology influenced her creative process, these insights would have been included in the analysis.

This broader approach was deliberate, as narrowing the focus exclusively to choreographic considerations would have limited our ability to examine other crucial aspects of implementing these technologies in dance performance. The intention was to maintain an expansive view that could accommodate multiple perspectives and concerns, rather than privileging any single aspect of the performance-technology relationship.

7.6 Conclusion

This chapter examined the creative and technical process involved in conceiving and executing the machine learning-powered performance piece *Interlinked*. The chapter traced the journey from conceptualisation to implementation, highlighting key milestones such as venue selection, animation design, motion capture, and integration. Audience feedback and the performer's insights highlighted the artefact's artistic impact and the distinctive nature of the machine learning-based motion capture approach. The audience responses affirmed the performance's ability to evoke meditative and reflective emotions through dance, music and animation synergy. The performer's interview offered valuable perspectives into the evolution of motion capture practices and the streamlined artistic process facilitated by machine learning methods in this project.

Ultimately, the artefact showed that using machine learning-based motion capture can deliver compelling artistic experiences, and can do so at low cost, with consumer-grade hardware, and without necessitating the use of markers, specialised clothing or highly controlled capture environments. While not devoid of challenges, this collaborative performance points to an exciting future where this technology enables the creation of many kinds of immersive experiences that resonate with audiences. The chapter thus makes a case for greater exploration at the intersection of creative practice and machine learning-based motion capture within the performing arts domain.

Cloé's contemporary dance training enabled her to deliberately explore and test the system's capabilities through controlled, precise movements. Meanwhile, Harrison and Sam's background in physical theatre and experimental performance informed their approach of intentionally challenging the system's limits by using atypical movements and body poses (discussed in Chapter 5). Each performer's specialized training allowed them to interact with the technology in distinct ways - Cloé through refined technical execution, and Harrison and

Sam through purposeful experimentation that pushed against conventional movement patterns. Their respective movement vocabularies and performance experiences directly influenced not only how they chose to move, but also how they conceptualized and explored the potential of the motion capture system.

8 Future Work

8.1 Introduction

In the context of technological advancement and artistic expression, the domain of creating animation for performing arts has undergone a notable transformation, driven by the integration of machine learning-based motion capture methodologies. This thesis has examined the intersection of machine learning-based motion capture and performance, specifically focusing on generating motion capture data to integrate animation into live performance. However, even as we conclude this investigation, it is evident that we are at the early stages of exploring a vast field of potential. Integrating machine learning-based motion capture and performing arts is an emerging concept poised for further exploration and exploitation. The preceding chapters have systematically explored the application of machine learning techniques to create sophisticated animations capable of engaging audiences with their emotional resonance. Despite this work and the work of others, the landscape remains relatively novel, leaving ample room for advancement.

Looking ahead, the landscape is dotted with the promise of novel innovations, models and techniques that can be discovered, refined, and integrated into performing arts. The rapid evolution of machine learning suggests a continuing influx of new ideas that can be harnessed to enhance the essence of performing arts. The convergence of technology and creativity is far from peaking. Furthermore, the potential for interdisciplinary collaboration is apparent, inviting artists, computer scientists, engineers, and researchers to converge and enrich the symbiotic relationship between machine learning and performing arts. The interaction of diverse perspectives can trigger innovative approaches, spark unconventional ideas, and catalyse breakthroughs that push the boundaries of artistic and technical possibilities.

8.2 Cross-disciplinary collaboration

As we look ahead to the future of machine learning methods for creating animation for performing arts, it is evident that the potential for growth and innovation lies in the convergence of different disciplines. The scope of this research has primarily focused on individual or small-scale collaborations that draw on open-source machine learning models. However, the true depth and richness of the field can only be fully realised through the active collaboration of artists, researchers, and technologists from diverse backgrounds. There is

the potential for a significant impact that cross-disciplinary collaboration can have on advancing the field. The fusion of art, research, and technology holds immense promise for pushing the boundaries of machine learning methods for creating animation for performing arts. Artists possess a unique understanding of the emotional and creative dimensions essential to captivating an audience. Researchers contribute with their analytical skills and scientific methodologies to enhance animations' technical accuracy and quality. Engineers bring expertise in implementing cutting-edge tools and platforms to bring ideas to life. The synergy among these disciplines can lead to novel solutions that would not be achieved in isolation.

Using machine learning methods to create animation for performing arts is still relatively new, leaving unexplored territories to investigate. Collaborative efforts have the potential to uncover undiscovered paths and possibilities. From experimenting with new algorithms that capture the nuances of human motion more accurately to exploring interactive elements that engage the audience in unprecedented ways, the collaborative approach may lead to new outcomes that reshape the landscape of performing arts. Encouraging collaboration necessitates creating platforms and opportunities for artists, researchers, and technologists to interact and exchange ideas. Workshops, conferences, and online communities can serve as avenues for interdisciplinary knowledge sharing. Additionally, dedicated projects that bring together experts from diverse fields could yield outcomes that drive the field forward.

8.3 Integration of new machine learning models

Machine learning is characterised by its rapid evolution. Developing novel architectures, training methodologies, and optimisation techniques is ongoing. The integration of newer models could provide improvements and breakthroughs. As new models emerge, there is potential for improved accuracy, greater efficiency and improved ease-of-use. Exploring these models in the context of performing arts opens doors to expanding the scope of artistic expression, creating performances that accurately mimic human motion and open up new areas for creative exploration.

8.4 The enhancement of current models with added functionality

The modular design of the multi-camera MPP2SOS model inherently offers a level of flexibility conducive to evolution. Its modular components can be replaced or updated with more effective alternatives as they become available. For instance, the pattern detection

module's reliability could be improved by substituting it with a more robust pattern recognition technique. By embracing this modularity, we open the door to a model that can adapt to emerging challenges. An intriguing prospect for the future of the MPP2SOS model involves the integration of the SMPL (Skinned Multi-Person Linear) model. This integration could enable the model to output meshes, streamlining the animation creation process by eliminating the need for manual skinning. By leveraging the capabilities of SMPL, the efficiency of the MPP2SOS model could be significantly enhanced, reducing the manual effort required for skinning.

An unexplored avenue that holds potential is the performance space of the MPP2SOS model in large capture spaces. While the model has been validated under controlled conditions, the resilience of its performance in vast capture spaces, where some cameras might not have direct visibility of the performer, remains to be tested. Investigating the model’s adaptability in scenarios where a performer is visible through some cameras but less visible through others could significantly extend the model’s application scope, making it suitable for expansive performance settings.

Finally, incorporating an automated noise reduction mechanism could substantially enhance the user-friendliness of the monocular VIBE model. An integrated noise reduction module could significantly enhance pose detection’s temporal coherence and mitigate jittering. Further improvements could allow users to adjust the degree of jitter reduction, offering a more personalised and versatile animation generation process.

8.5 Multiple performers

An area for investigation is the application of models to scenarios involving multiple performers. While the current research predominantly focuses on individual performers, expanding this paradigm to accommodate multiple performers could reveal novel insights into the capabilities and limitations of the models. The interaction between performers can give rise to complex occlusions, overlaps, and interactions that may challenge the model’s ability to capture and represent movement, particularly in monocular models. Evaluating the models’ performance in such settings can provide valuable insights into their adaptability and effectiveness in diverse performing arts contexts.

8.6 User-friendly enhancements

One of the pivotal aims in advancing the field of creating animation for performing arts is to make the technology more accessible and user-friendly for artists and practitioners. The Blender MPP2SOS add-on demonstrated a more simplistic approach to creating an animated mesh without the need to deal with code. By integrating the complex processes within a widely used and open-source 3D software like Blender, the technical barriers for artists and practitioners were substantially lowered. This integration not only sidestepped the need for coding expertise but also enabled artists to harness the power of machine-learning models directly within their creative environment. The success of the MPP2SOS Blender add-on prompts exploration into replicating this integration with other popular 3D software commonly used in performing arts and animation. The integration of machine learning-based motion capture models into software platforms like Maya[60], Houdini, or Unreal Engine[61] could provide artists with a broader spectrum of tools to choose from, each tailored to their preferred creative workflow.

Another avenue to enhance accessibility is the provision of pre-made animations that artists can easily apply to their generated meshes. These pre-made animations can serve as inspiration, allowing artists to visualise the animation potential of their models quickly. These pre-made animations can be a starting point, nurturing creative experimentation and enabling practitioners to discover new artistic expressions. Even if these pre-made animations are created in Houdini, a non-commercial version is free. It does have limitations, such as a watermark in the corner of the image, but it can still be a source of inspiration. Such tools could allow artists to assess the impact of animations on their creations and make informed decisions before moving forward with the entire animation process.

8.7 Audience interaction

Embedding audience interaction within a performance introduces a dynamic and collaborative dimension that can revolutionise how performances are experienced. Audience members become active participants, shaping and influencing the progression of the performance. This level of engagement can bridge the gap between the performer and the audience, creating a shared experience that blurs the lines between creator and spectator. Integrating real-time tracking systems that detect audience members' movements, gestures, or proximity opens avenues for dynamic interactivity. Imagine animations responding to the

[60] Autodesk Maya, https://www.autodesk.com/products/maya
[61] Unreal Engine, https://www.unrealengine.com/en-US

audience’s presence, allowing them to alter the narrative or appearance of animations based on their actions. Performance becomes a collaborative effort, co-created by both performers and spectators in real-time.

By inviting the audience to participate actively, the potential for unexpected outcomes and spontaneous creative expressions multiplies. This unpredictability element injects surprise and innovation into the performance, enhancing its artistic depth and uniqueness. While audience interaction is enticing, it comes with technical challenges and ethical considerations. Reliable tracking systems must be implemented to capture the audience's movements accurately. Additionally, maintaining a balance between audience interaction and the performers’ intended narratives requires careful design and coordination.

8.8 Performers insights

The invaluable perspective of experienced performers can shed light on unchartered possibilities within the realm of creating animation for performing arts. These insights offer novel directions for future exploration, emphasising the interplay between the performer and the animation. One of the performer’s suggestions centres on redefining the interaction between the performer and the animation. Instead of the performer simply mimicking the animation, the animation could influence the performer’s actions, creating a dynamic dialogue between the two. This concept introduces a captivating layer of responsiveness, where the animation becomes an active collaborator, influencing the performer’s movements in real-time.

8.9 Data Availability and accessibility

The availability of diverse and comprehensive motion capture data plays a pivotal role in shaping the capabilities of machine learning models for use in performing arts. Leveraging open datasets and collaborating with motion capture studios are crucial avenues to consider. Motion capture studios possess a treasure trove of motion capture data accumulated over years of capturing various performances. While some data might be subject to copyright and intellectual property concerns and perhaps ethical considerations, a substantial portion could be made accessible for training machine learning models. These diverse datasets, spanning a broad spectrum of motions and interactions, could significantly enhance the robustness and versatility of animation-generating models.

The experiences with the VIBE and MPP2SOS models shed light on the potential benefits of data enrichment for model training. A more extensive and diverse dataset could facilitate more accurate predictions in cases where models struggled with specific poses, such as performers suspended in the air on aerial slings. Incorporating motion capture data of performers on aerial slings or engaging in unconventional movements could aid in refining the models' pose estimation capabilities. The availability of open datasets curated for performing arts can accelerate progress in the field. Creating open datasets encompassing a variety of performance styles, genres, and contexts could provide researchers, artists, and technologists with a foundation for model training and experimentation. Such datasets would promote research collaboration and democratise access to motion capture data for artists exploring this field.

An important consideration is the need for training data that encompasses diverse body shapes and movement styles. While new motion capture data from professional studios would be the gold standard, an interim solution could be to collect and annotate videos of existing artistic performances. Particularly valuable would be performances involving atypical movement patterns, unique body positioning, the use of specialised equipment like aerial slings, ropes, or props, and performers with diverse body shapes. Curating and annotating this type of diverse performance data could significantly improve the available training sets for machine learning models aimed at performance arts applications. This diverse data could then be leveraged to fine-tune existing models, enabling them to better generalise to the full range of artistic expression across different body types, movement vocabularies, and performance styles. Prioritising this diversity in training data could be crucial for realising the potential of machine learning-based approaches in creative performance contexts.

While the potential benefits of expanded motion capture data availability are substantial, ethical considerations must be navigated carefully. Protecting performers' rights and ensuring proper anonymisation of data is paramount. Collaborative efforts should prioritise transparency and consent, fostering an environment where data sharing aligns with ethical principles.

8.10 Conclusion

The integration of machine learning-based motion capture and performing arts has only begun to reveal its potential. While the current exploration has yielded promising results, it represents a fraction of the possibilities. As machine learning continues to evolves and

performing arts seeks new frontiers for exploration, their convergence will unlock novel techniques, innovations, and experiences. However, realising this potential necessitates a commitment to ongoing research, collaboration and ethical data practices. Cross-disciplinary partnerships among artists, researchers, and technologists can catalyse breakthroughs, combining creativity, analytical rigour, and technical expertise. Tapping into emerging machine learning models and enhancing current approaches will ensure progress aligned with cutting-edge advancements. Democratising access through intuitive tools and open datasets is vital to empowering broader participation. Throughout, upholding ethical standards regarding data practices and intellectual property will foster an ecosystem of trust and transparency.

The shared vision to transform performing arts through technology is at the heart of this effort. While the path ahead holds challenges, its horizons promise to expand artistic capabilities and captivate audiences. With perseverance, imagination and synergy, the future offers endless opportunities.

9 Conclusion

As this thesis draws to a close, this chapter will reflect on the journey to address the two fundamental research questions. We began by uncovering the characteristics sought by creative practitioners in motion capture systems, encapsulated in the first research question: ‘In performing arts, what are the characteristics creative practitioners look for in motion capture systems?’ (RQ1). The second research question sought to uncover not only the potential advantages offered by machine learning technologies but also the inherent challenges and implications associated with their integration into performing arts: ‘What are the benefits, limits, and implications of current machine learning-based motion capture systems in the performing arts space?’ (RQ2).

9.1 Key findings for RQ1

9.1.1 Literature review

The literature review highlights the evolution of motion capture techniques, from the early labour-intensive methods of hand-drawing images to the more automated digital approaches that significantly expedited the process. While precise optical systems like Vicon offer highly accurate tracking, their lack of portability and budgetary constraints often render them unsuitable for creative practitioners. Alternative motion capture methods, such as active marker systems, although less prevalent than their passive counterparts, are not well-suited for the performing arts due to their requirement for line-of-sight visibility to cameras and the need for performers to wear restrictive suits (Latulipe et al., 2011; Jobson, 2015; Meador et al., 2004).

There have been alternatives that eliminate the need for markers or specialised suits, such as performing with gyroscopes (Latulipe et al., 2011), the Microsoft Kinect (Jobson, 2015), depth cameras (Meador et al., 2004), or sensor-based motion capture systems like Rokoko, Xsens, or Perception Neuron. The growing preference among practitioners for motion capture systems that do not require suits or markers suggests a desire for solutions that do not hinder performers' movements or attire during performances. Machine learning-based motion capture presents a new opportunity for potentially satisfying this preference for markerless capture approaches.

9.1.2 Interviews with practitioners using motion capture for performance

During interviews with practitioners, it was established that their choice of technology is largely dependent on the specific project and the skills of the individuals within the group, as they may opt for technologies they are already familiar with. Practitioners expressed that the level of accuracy in a motion capture system varies, but it is not necessarily a deciding factor in their choice of system.

Portability emerged as an essential consideration, as practitioners require the ability to change rehearsal spaces or travel to performance venues with ease. They mentioned that some motion capture methods necessitate maintenance and may sometimes not function as intended, such as experiencing magnetic interference. Practitioners expressed a need for hardware to work more seamlessly, as hardware malfunctions can temporarily halt shows, which should be avoided.

Due to budgetary constraints, cost played a significant role in the choice of motion capture systems. However, practitioners worked with the resources available and were willing to accept some downsides of their chosen system, such as imperfect accuracy. They mentioned that pre-recorded motion capture was sometimes used in performances, enhancing the overall production.

Interestingly, practitioners found errors or anomalies in the captured motion data intriguing and intentionally utilised these glitches to create interesting visual effects within their productions. This openness to embracing imperfections or unexpected outcomes highlights a unique perspective among practitioners in performing arts.

Overall, the interviews revealed that practitioners' needs encompass a range of factors, including project-specific requirements, team skillsets, portability, cost-effectiveness, reliability, and a willingness to explore creative avenues through the unintended artifacts or glitches generated by motion capture systems.

9.1.3 Collaboration with professional choreographers and directors

Working with experienced choreographers and directors in the performing arts space revealed their openness to incorporating technology into their performances, even if it involved emerging technologies that had not been extensively tested. Their approach was not to dictate the outcome of the motion capture process but rather to observe how the

technology would interpret and represent the movement, leaving the output open-ended and experimental.

The practitioners embraced the ability to work remotely, not only because of geographical distances (as they were based in a different state) but also because it allowed them the freedom to collaborate while simultaneously working on other projects worldwide. This flexibility was a welcome advantage.

When working with professional choreographers and directors, it was refreshing to discover that they were not fixated on achieving perfect accuracy in motion capture, unlike practitioners in the film, animation, and gaming industries, where mimicking the exact movements of the performer verbatim is often a priority. Instead, these practitioners were interested in anomalies, such as glitches and unexpected outcomes. They deliberately provided video footage that would naturally be very challenging for the motion capture system to generate coherent data, as if the more incoherent the data, the more fascinating it was for them. The practitioners were curious to discover the interpretations of the emerging motion capture system when presented with footage designed to hinder its detection capabilities.

Through these collaborations, it became evident that the practitioners worked not only with live motion capture but also incorporated pre-recorded motion capture and animation into their performances. This multifaceted approach demonstrated their versatility and willingness to explore various techniques and technologies within their creative process.

9.2 Key findings for RQ2

9.2.1 Interviews from practitioners using motion capture for performance

Based on the responses from practitioners involved in motion capture for performance, it is evident that machine learning methods for motion capture have the potential to address their diverse needs and requirements, as highlighted by the following points:

The choice of motion capture system is largely context-dependent, but ease of use emerges as a critical factor in ensuring accessibility. With their relatively straightforward implementation, machine learning approaches align well with this requirement.

While accuracy is relative to the context of the motion capture output, practitioners acknowledge that machine learning models for motion capture are continually improving and that some find the inherent inaccuracies of these models appealing for creative exploration. This is relevant to machine learning-based motion capture systems because inaccuracies and glitches can somewhat be controlled by manipulating the input data.

Portability is a high priority for performers, and modern machine learning methods for motion capture excel in this regard, as they have been trained on and perform well with relatively low-quality video data that can be captured on compact, portable, consumer-grade video capture devices (such as mobile phones).

Practitioners recognised the hardware limitations of traditional motion capture systems and expressed appreciation for the relatively low-maintenance nature of machine learning approaches, which could offer a more streamlined solution.

Cost considerations are a significant emphasis, and the potential cost advantages of machine learning methods, which have been shown to work well with less specialised equipment in this work, make them an attractive option for practitioners operating under budgetary constraints.

Incorporating both pre-recorded animation and real-time motion capture is desirable, and machine learning-based systems could facilitate a seamless integration of these workflows.

Practitioners expressed openness to the creative possibilities offered by glitch artifacts and deliberate data manipulation, areas where machine learning models could provide avenues for experimentation through techniques like camera calibration adjustments.

Notably, the feedback from performing arts practitioners regarding the potential of machine learning for motion capture was positive, with all interviewees expressing receptiveness to the technology, particularly if it could lead to time savings, cost reductions, and increased accessibility.

The use of real-time motion capture creation for performing arts faces limitations when employing machine learning-based methods. These limitations arise from the substantial processing power required and the challenges associated with outputting high-quality animation in real-time.

However, it is evident that practitioners in the performing arts field often incorporate pre-recorded motion capture and animation into their performances. This approach can involve either utilising pre-recorded motion capture data as a standalone element or integrating it with live motion capture.

The pre-recorded motion capture data and animations can be processed and refined in advance, allowing practitioners to overcome the real-time processing constraints imposed by machine learning methods. By combining these pre-processed elements with live motion capture, practitioners can create a blend of pre-rendered and real-time components, enabling them to leverage the benefits of both approaches.

This hybrid technique addresses the limitations of real-time processing and offers practitioners greater flexibility and control over the visual outcomes. Pre-recorded motion capture and animations can be crafted and polished, while the live motion capture component introduces a dynamic and interactive element to the performance.

Overall, while real-time motion capture creation for performing arts faces challenges with machine learning methods, practitioners have adapted by incorporating pre-recorded motion capture and animation, either as standalone elements or in conjunction with live capture, to create captivating and visually rich performances.

In summary, the responses from practitioners involved in motion capture for performance suggest that machine learning methods have the potential to address their diverse needs, offering advantages in areas such as ease of use, portability, cost-effectiveness, creative exploration, and seamless integration into existing workflows.

9.2.2 My experience using single and multi-camera models

During the reflective practice of investigating single and multi-camera machine learning models for their use in the performing arts, the initial challenge was finding a suitable machine learning model that could effectively output an animation. Once an appropriate model was identified, the process involved running and testing the open-source models. Relating to the research question ‘What are the benefits, limits, and implications of current machine learning-based motion capture systems in the performing arts space?’, my experience as a practitioner meant I was aware of the potential difficulties in adopting this technology. It can be challenging to locate the desired model and navigate the installation

process, as these models are often not presented as user-friendly software packages with straightforward installation and execution.

However, once the model is up and running, the setup process for single or multiple cameras is reasonably quick. While processing the video may take some time, depending on the resolution, the resulting animatable skeleton output is valuable for making further adjustments after processing.

For my purposes, the level of accuracy achieved by the output skeleton did not need to match that of high-end systems like Vicon, as the sentiment from interviews with practitioners in the field also suggests that accuracy is not always a defining factor. The ability to set up the system quickly and easily change locations became evident, particularly with single-camera models, where a single handheld device was effectively used in an uncontrolled performance area.

The reflective practice highlighted the initial hurdles in locating and setting up suitable machine learning models for motion capture, but once operational, these models offer advantages in terms of cost-effectiveness, ease of use, and portability – characteristics that align with the needs of creative practitioners in the performing arts. The level of accuracy, while not on par with high-end solutions, was deemed sufficient for my purposes and aligned with the perspectives of practitioners in the field.

9.2.3 The artefact design and production process

Using machine learning methods to create an artefact also provided valuable insights that helped answer the first research question regarding the characteristics that creative practitioners look for in motion capture systems. The ability to set up rapidly and change capture locations streamlined the process, aligning with one of the key feedback points from practitioner interviews: performers dislike waiting for extended periods while equipment is being set up, as they need to be ready once they are warmed up.

Regarding the hardware costs, the most expensive components were the iPhones, iPads, and a GoPro action camera used for capturing the motion capture footage. No additional hardware modifications were required. The devices used for capturing multiple angles were borrowed and of diverse form factors and video capabilities, which could be a common scenario for those without readily available equipment.

The processing of the motion capture data was performed on a four-to-five-year-old home PC. The rendering was done on a render farm at UTS to expedite the process, as the high resolution required for projection in the Data Arena necessitated a larger video size.

Overall, the artefact design and production phase provided valuable insights into the characteristics that creative practitioners seek in motion capture systems, such as rapid setup, portability, and the ability to integrate smoothly into existing workflows while minimizing downtime and maximizing efficiency.

9.2.4 Collaboration with professional choreographers and directors

Collaborating with experienced practitioners who use technology and motion capture in their performances provided valuable insights into their openness and receptiveness towards these technologies, even when they yield unpredictable results. It became evident that performers are open to and encourage interesting glitch behaviours as outputs of captured motion data. Given that some practitioners were based in Melbourne while I was in Sydney, the collaboration benefited from the convenience of working remotely without needing physical co-location. This flexibility was facilitated by using a machine learning model for generating motion capture, eliminating the requirement for the performer to be present in the same room, as video recordings of the performances could be shared over the internet. This setup enabled a more flexible working situation for choreographers and performers.

Based on interviews with practitioners, accessibility, affordability, and portability were identified as critical issues, and through this collaborative experience, these areas were found to be favourable when using machine learning-based motion capture systems. Accessibility was enhanced by the ability to utilise consumer devices, such as mobile phones, for recording motion capture data and the flexibility to use any space that can accommodate the performer. The equipment required for this approach is relatively affordable compared to traditional motion capture methods. Additionally, the setup was highly portable, as a single camera was sufficient, and the system tolerated suboptimal lighting conditions, further contributing to its accessibility and ease of use.

9.2.5 The interview with the performer

During the interview, Cloé described her process of performing for motion capture as a trial-and-error process with the technology, mainly when working in a real-time setting. While she acknowledged the benefits of real-time feedback, she mentioned that using pre-recorded motion capture is a common practice for her.

In contrast to the traditional approach of experimenting with motion capture as the performer is dancing, the method employed in this project involved using previously captured motion capture data. This approach allowed experimentation with the captured data, effectively bypassing the need for the performer to be directly involved in the experimentation stage. Consequently, this method saved valuable time for the performer, eliminating the need for extensive trial-and-error during the performance itself.

The project streamlined the process by leveraging pre-recorded motion capture data, enabling efficient experimentation and refinement without requiring the performer's constant presence or involvement in the iterative stages. This approach aligned with Cloé's familiarity with working with pre-recorded motion capture in her practice while simultaneously optimising the use of the performer's time and minimising potential disruptions to their workflow.

9.2.6 Audience feedback

During interviews with the audience after the production of *Interlinked*, they generally expressed pleasure with the overall performance. However, there were mixed views regarding whether the performance drove the animation in a live manner. Interestingly, except for one participant, the audience thought the animation was real-time. Some participants believed that some of the animation was pre-recorded, while others thought it was entirely live.

This divergence of perspectives suggests that real-time processing speed may not be a crucial factor for this type of performance, challenging the notion that real-time processing is the sole determinant of success. Considering the audience's feedback made it evident that the performance was engaging, reiterating that machine learning methods for motion capture in performance art are a feasible option.

Currently, the limitations of machine learning-based motion capture systems are that they do not allow for driving animations at the high resolution required for the Data Arena or achieving the quality that was created in real-time. However, the audience's positive reception indicates that these limitations did not significantly detract from the overall experience, highlighting the potential of machine learning approaches in this domain.
While the findings connect with academic research, they are particularly relevant to dancers, choreographers, directors, and other creative professionals who are looking to understand the practical opportunities and limits of contemporary machine-learning-based motion capture technology in the performing arts space. This study provides these practitioners with insights into the potential benefits and challenges of adopting such technologies, offering a foundation for informed decision-making and creative exploration. This work examines the technical performance and artistic implications of current machine learning-powered motion capture systems in a range of practical artistic environments and through engagement with a diversity of practitioners. It highlights how, why and when performing arts may benefit from the off-the-shelf machine learning motion capture systems available to creative professionals today. The accessibility of the methods described, coupled with the real-world examples provided, makes this work a valuable resource for practitioners seeking to innovate and expand their artistic horizons through technology.

The preceding discussion of relevance to practitioners should be understood in the context of knowledge contribution rather than technical accessibility. By documenting our findings about the interaction between technology and artistic practice, we provide insights that inform decision-making and creative exploration. This differs from creating technical documentation or tools - instead, we offer a broader understanding of how these systems operate within performance contexts.

The contribution lies in understanding how the technical strengths and limitations of current off-the-shelf machine learning-powered motion capture systems inform their potential adoption and application in performance art contexts. This insight helps practitioners make informed decisions about where, how, and why to integrate these systems into their creative practice, even as the underlying technology continues to evolve.

This approach aligns with our research questions, which focus on understanding the relationship between technology and artistic practice. The core contributions centre on this understanding, providing insights into the practical and artistic implications of using these systems in performance contexts.

As machine learning-based motion capture continues to evolve, its integration into the performing arts will open up new creative possibilities for artists. While still an emerging area, it may allow performers and choreographers to explore novel ways of expression through movement. The combination of machine learning-based motion capture's technical capabilities and the artistic vision of the performing arts community has the potential to lead to innovative performances that push boundaries. The effect is potentially transformative, unlocking low-cost and flexible motion capture that is of a sufficient quality that it can be used to complement and enhance live performance. With an open-minded approach to experimentation and a willingness to embrace the inherent imperfections of these technologies, I believe that performing artists will find unique avenues to engage audiences in compelling new experiences.

10 References

4K support on Smartphones - Playback, Recording & Display. (n.d.). Retrieved January 29, 2022, from https://7labs.io/mobile/4k-supported-phone.html

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., & Dean, J. (2016). TensorFlow: A System for Large-Scale Machine Learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16).*, 786.

Aerial Silks vs. Aerial Sling: What's the Difference and Which One is Right for You? (2023, August 11). https://www.verticalwise.com/aerial-silks-vs-aerial-sling/

Akhter, I., & Black, M. J. (2015). Pose-conditioned joint angle limits for 3D human pose reconstruction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 1446–1455. https://doi.org/10.1109/CVPR.2015.7298751

Alaoui, S. F., Bevilacqua, F., Pascual, B. B., & Jacquemin, C. (2013). Dance interaction with physical model visuals based on movement qualities. *International Journal of Arts and Technology*, *6*(4), 357–387. https://doi.org/10.1504/IJART.2013.058284

Alaoui, S. F., Carlson, K., & Schiphorst, T. (2014). *Choreography as Mediated through Compositional Tools for Movement: Constructing A Historical Perspective*. www.motionbank.org

Alaoui, S. F., Schiphorst, T., Cuykendall, S., Carlson, K., Studd, K., & Bradley, K. (2015). *Strategies for Embodied Design: The Value and Challenges of Observing Movement*. https://doi.org/10.1145/2757226.2757238ï

Alemi, O., Li, W., & Philippe Pasquier. (2015). *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on : date 21-24 Sept. 2015*.

Anderson, C., & Kirkpatrick, S. (2016). Narrative interviewing. *International Journal of Clinical Pharmacy*, *38*(3), 631–634. https://doi.org/10.1007/s11096-015-0222-0

Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2D human pose estimation: New benchmark and state of the art analysis. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3686–3693. https://doi.org/10.1109/CVPR.2014.471

Anwar, A. (2022). What are Intrinsic and Extrinsic Camera Parameters in Computer Vision? In *Towards Data Science*.

Art, G., Fry, C. J., & Hons, B. A. (2021). *Encounters with Errors : How the error shapes relationships with digital media practice*. 399–408.

Azure Kinect body tracking joints. (2019, June 26). https://learn.microsoft.com/en-gb/previous-versions/azure/kinect-dk/body-joints

Baker, R. (2007). The history of gait analysis before the advent of modern computers. *Gait and Posture*, *26*(3), 331–342. https://doi.org/10.1016/j.gaitpost.2006.10.014

Barbacci, S. (2002). Labanotation: a universal movement notation language. *Journal of Science Communication*, *01*(01), A01. https://doi.org/10.22323/2.01010201

Barreto, C. (n.d.). *MPP2SOS*. Gumroad.Com. Retrieved August 14, 2023, from https://carlosedubarreto.gumroad.com/l/mocap_mpp2sos?layout=profile

Bastien Girschig. (2019a). *Living Archive by Wayne McGregor*. https://experiments.withgoogle.com/living-archive-wayne-mcgregor

Bastien Girschig. (2019b). *Living Archive by Wayne McGregor*.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). *BlazePose: On-device Real-time Body Pose tracking*.

Bell, P. (2000). Dialogic Media Productions and Inter-Media Exchange. *Journal of Dramatic Theory and Criticism*, *14*(2), 41,Äì56.

Bénard, P., Hertzmann, A., & Kass, M. (2014). Computing smooth surface contours with accurate topology. *ACM Transactions on Graphics*, *33*(2). https://doi.org/10.1145/2558307

Betancourt, M. (2016). Glitch Art in Theory and Practice. In *Glitch Art in Theory and Practice*. https://doi.org/10.4324/9781315414812

Bird, C. M. (2005). How I stopped dreading and learned to love transcription. *Qualitative Inquiry*, *11*(2), 226–248. https://doi.org/10.1177/1077800404273413

Bisig, D. (2021a). *Granular Dance*. *July*.

Bisig, D. (2021b). *Raw Music from Free Movements : Early Experiments in Using Machine Learning to Create Raw Audio from Dance Movements*. 1–11.

Bisig, D., & Wegner, E. (2021). Puppeteering an AI - Interactive Control of a Machine-Learning based Artificial Dancer. *XXIV Generative Art Conference - GA2021*, 315–332.

Blacking, J., & Keali'inohomoku, J. W. (1980). The Performing Arts: Music and Dance. In *Ethnomusicology* (Vol. 27, Issue 2, p. 359). Walter de Gruyter Gmbh US SR. https://doi.org/10.2307/851081

Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., & Black, M. J. (2016). Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9909 LNCS*, 561–578. https://doi.org/10.1007/978-3-319-46454-1_34

Booth, W. C., Colomb, G. G., & Williams, J. M. (2003). *The Craft of Research*.

Bowers, J., Norman, S. J., Staff, H., Schwabe, D., Wallen, L., Fleischmann, M., Sundblad, Y., & Bowers, A. J. (1998). Extended Performances: Evaluation and Comparison. *Royal Institute of Technology*, *May*, 1–43.

Boyatzis, R. E. (1998). *Transforming Qualitative Information Thematic Analysis and Code Development* (p. 200).

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology Using thematic analysis in psychology. *Psychiatric Quarterly*, *0887*(1), 37–41.

Brenton, H., Kleinsmith, A., & Gillies, M. (2014). Embodied design of dance visualisations. *ACM International Conference Proceeding Series*, 124–129. https://doi.org/10.1145/2617995.2618017

Bryman, A. (2001). Social research methods. In *OUP Oxford*.

Candy, L., & Edmonds, E. (2011). Interacting: Art, Research and the Creative Practitioner. In *Interacting. Libri, Oxfordshire* (Issue April 2011).

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2018). Openpose. *ArXiv*, *2017-Janua*(Xxx), 1302–1310. https://doi.org/10.1109/CVPR.2017.143

Carlson, K., Pasquier, P., Tsang, H. H., Phillips, J., Schiphorst, T., & Calvert, T. (2014). *Cochoreo: A Generative Feature in idanceForms for Creating Novel Keyframe Animation for Choreography*.

Carlson, K., Schiphorst, T., Cochrane, K., Phillips, J., Tsang, H. H., & Calvert, T. (2015). Moment by moment: Creating movement sketches with camera stillframes. *C and C 2015 - Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, 131–140. https://doi.org/10.1145/2757226.2757237

Carlson, K., Tsang, H. H., Phillips, J., Schiphorst, T., & Calvert, T. (2015). Sketching movement: Designing creativity tools for in-situ, whole-body authorship. *ACM International Conference Proceeding Series*, *14-15-August-2015*, 68–75. https://doi.org/10.1145/2790994.2791007

Casa Paganini InfoMus. (2019). *Casapagnini.Org*. http://www.casapaganini.org/assets/CasaPaganini-InfoMus-Nov2019.pdf

Cascone, K. (2001). *The Aesthetics of Failure: "Post Digital" Tendencies in Contemporary Computer Music*.

Chan, C., Ginosar, S., Zhou, T., & Efros, A. (2019). Everybody dance now. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 5932–5941. https://doi.org/10.1109/ICCV.2019.00603

Chang, V. (2019). Catching the ghost: the digital gaze of motion capture. *Journal of Visual Culture*, *18*(3), 305–326. https://doi.org/10.1177/1470412919841022

Chen, Y., Shen, C., Wei, X. S., Liu, L., & Yang, J. (2017). Adversarial PoseNet: A Structure-Aware Convolutional Network for Human Pose Estimation. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-October*, 1221–1230. https://doi.org/10.1109/ICCV.2017.137

Choi, J. H., Lee, J. J., & Nasridinov, A. (2021). Dance self-learning application and its dance pose evaluations. *Proceedings of the ACM Symposium on Applied Computing*, 1037–1045. https://doi.org/10.1145/3412841.3441980

Colyer, S. L., Evans, M., Cosker, D. P., & Salo, A. I. T. (2018). A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sports Medicine - Open*, *4*(1). https://doi.org/10.1186/s40798-018-0139-y

Crnkovic-friis, L., & Crnkovic-friis, L. (2016). Generative Choreography using Deep Learning Long Short-Term Memory. *7th International Conference on Computational Creativity, ICCC2016*, 1–6.

Cubitt, S. (2017). Glitch. *Cultural Politics*, *13*(1), 19–33. https://doi.org/10.1215/17432197-3755156

Dabral, R., Gundavarapu, N. B., Mitra, R., Sharma, A., Ramakrishnan, G., & Jain, A. (2019). Multi-Person 3D Human Pose Estimation from Monocular Images. *Proceedings - 2019 International Conference on 3D Vision, 3DV 2019*, 405–414. https://doi.org/10.1109/3DV.2019.00052

Dalmazzo, D., & Ramírez, R. (2019). Bowing gestures classification in violin performance: A machine learning approach. *Frontiers in Psychology*, *10*(MAR). https://doi.org/10.3389/fpsyg.2019.00344

David, M. (2005). Real-time motion-e capture makes dance a digital art. *Electronic Design*, *53*(10), 19.

Dent, S. (2014). What you need to know about 3D motion capture. In *Engadget* (pp. 1–5).

Dixon, S. (2007). *Digital Performance : A History of New Media in Theater, Dance, Performance Art, and Installation*.

Donahue, C., Lipton, Z. C., & Mcauley, J. (2017). Dance Dance Convolution. *Proceedings of the 34th International Conference on Machine Learning*.

Dong, J., Fang, Q., Jiang, W., Yang, Y., Huang, Q., Bao, H., & Zhou, X. (2021). Fast and Robust Multi-Person 3D Pose Estimation and Tracking from Multiple Views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(8). https://doi.org/10.1109/TPAMI.2021.3098052

Downie, M., & Kaiser, P. (2018). *Marc Downie / Paul Kaiser OpenEndedGroup selected artworks 1998 – 2018*.

Dudley, J. J., & Kristensson, P. O. (2018). A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems*, *8*(2). https://doi.org/10.1145/3185517

Dulari Bhatt. (2021). *A comprehensive guide for Camera calibration in computer vision*.

Durkin, J., Jackson, D., & Usher, K. (2020). Qualitative research interviewing: reflections on power, silence and assumptions. *Nurse Researcher*, *28*(4), 31–35. https://doi.org/10.7748/nr.2020.e1725

Egan, K. F. W. (2020). 'Tones from Out of Nowhere' and Other Non-sensedness: Re-membering the Synthetic Sound Films of Oskar Fischinger and Làszló Moholy-Nagy. *Animation*, *15*(2), 160–178. https://doi.org/10.1177/1746847720938230

Failes, I. (2019, September 10). *"Computer pajamas": the history of ILM's IMocap - befores & afters*. https://beforesandafters.com/2019/09/10/computer-pajamas-the-history-of-ilms-imocap/

Fdili Alaoui, S., Françoise, J., Schiphorst, T., Studd, K., Bevilacqua, F., Francoise, J., & Bevilacqua, F. (2017). *Seeing, Sensing and Recognizing Laban Movement Qualities*. https://doi.org/10.1145/3025453.3025530ï

Felciano, R. (1999). Dance a little dream of Merce. *Dance Magazine*, *73 (7)*, 72–72.

Foster, J. J., & Parker, I. (1995). Carrying out investigations in psychology: Methods and Statistics. In *BPS Books (British Psychological Society)*.

Fraleigh, S. H. (1999). *Researching Dance. Evolving Modes of Inquiry.*

Frayling, C. (1993). *Research in Art and Design*.

Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., & Rochester, L. (2014). Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait and Posture*, *39*(4), 1062–1068. https://doi.org/10.1016/j.gaitpost.2014.01.008

Goldberg, R. (1988). Performance Art: From Futurism to the Present. In *Leonardo* (Vol. 21, Issue 4, p. 464). https://doi.org/10.2307/1578724

Goldman, C. (1985). Abel, Ketchum Create A Sexy Robot for Can Council. *Back Stage (Archive: 1960-2000); New York*, *26*(4), 1, 10, 38.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *3*(January), 2672–2680.

Greco, E., & Scholten, P. C. (1997). *Double Skin/Double Mind*. Ickamsterdam.Com.

Grigonis, H. (2023). *How Lens Distortion Works in Photography (And How to Fix It)*. Expert Photography. https://expertphotography.com/what-is-lens-distortion/#:~:text=Barrel distortion is created by,more extreme the optical distortion.

Gu, L., Istook, C., Ruan, Y., Gert, G., & Liu, X. (2019). Customized 3D digital human model rebuilding by orthographic images-based modelling method through open-source software. *Journal of the Textile Institute*, *110*(5), 740–755. https://doi.org/10.1080/00405000.2018.1548079

Guest, A. H., & Ryman, R. (1998). Choreo-Graphics: A Comparison of Dance Notation Systems from the Fifteenth Century to the Present. In *Dance Research Journal* (Vol. 23, Issue 1, p. 35). Psychology Press. https://doi.org/10.2307/1478698

Gunerli, J. H. C., Deshpande, M., & Magerko, B. (2024, May 30). Video Segmentation Pipeline For Co-Creative AI Dance Application. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3658852.3659085

Gupta, A., Mittal, A., & Davis, L. S. (2006). Constraint integration for multiview pose estimation of humans with self-occlusions. *Proceedings - Third International Symposium on 3D Data Processing, Visualization, and Transmission, 3DPVT 2006*, *30*(3), 900–907. https://doi.org/10.1109/3DPVT.2006.45

Haag, J. (2008). *Inertial motion capture and live performance (with a focus on dance)*. https://ausdance.org.au/articles/details/inertial-motion-capture-and-live-performance

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 2980–2988. https://doi.org/10.1109/ICCV.2017.322

Hobby, K., Carlson, K., & Wheaton, D. (2017, July). *Rehabilitating Experience: Designing an Aesthetic and Movement-based Game for Physical Therapy*. https://doi.org/10.14236/ewic/eva2017.38

Höysniemi, J. (2006). International survey on the dance dance revolution game. *Computers in Entertainment*, *4*(2), 1–30. https://doi.org/10.1145/1129006.1129019

Intersecting geometry. (2014). Autodesk Maya Forum. https://forums.autodesk.com/t5/maya-animation-and-rigging/problems-with-geometry-overlapping/td-p/4771109

Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C. (2014). Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(7), 1325–1339. https://doi.org/10.1109/TPAMI.2013.248

Iskakov, K., Burkov, E., Lempitsky, V., & Malkov, Y. (2019). Learnable triangulation of human pose. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 7717–7726. https://doi.org/10.1109/ICCV.2019.00781

Joao Alves, A. (2021). *Overview Of Depth Cameras*. Aivero.Com. https://aivero.com/topic/overview-of-depth-cameras/

Jobson, C. (2015, March 9). *Asphyxia: A Striking Fusion of Dance and Motion Capture Technology — Colossal*. Colossal. https://www.thisiscolossal.com/2015/03/asphyxia-a-striking-fusion-of-dance-and-motion-capture-technology/

Johnson, S., & Everingham, M. (2010). Clustered pose and nonlinear appearance models for human pose estimation. *British Machine Vision Conference, BMVC 2010 - Proceedings*, *i*, 1–11. https://doi.org/10.5244/C.24.12

Johnston, A. (2014). Keeping Research in Tune with Practice. *Interactive Experience in the Digital Age*, 49–62.

Johnston, A. (2015a). Conceptualising interaction in live performance: Reflections on “Encoded.” *ACM International Conference Proceeding Series*, *14-15-Augu*(2006), 60–67. https://doi.org/10.1145/2790994.2791003

Johnston, A. (2015b). Conceptualising interaction in live performance: Reflections on “Encoded.” *ACM International Conference Proceeding Series*, *14-15-Augu*(2006), 60–67. https://doi.org/10.1145/2790994.2791003

Joo, H., Simon, T., Li, X., Liu, H., Tan, L., Gui, L., Banerjee, S., Godisart, T., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S., & Sheikh, Y. (2016). Panoptic Studio: A Massively Multiview System for Social Interaction Capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(1), 190–204. https://doi.org/10.1109/TPAMI.2017.2782743

Jordan, J. (2017). Evaluating a machine learning model. In *Data Science* (pp. 1–19).

Joslin, C. (n.d.). *Optical Motion Capture*. Retrieved July 24, 2023, from https://mocap.csit.carleton.ca/index.php?Section=System&Item=Markers&Page=Default

Jung, D., Jensen, M. H., Laing, S., & Mayall, J. (2012). Cyclic.: An interactive performance combining dance, graphics, music and kinect-technology. *ACM International Conference Proceeding Series*, 36–43. https://doi.org/10.1145/2379256.2379263

Kalmakurki, M. (2018). Snow White and the Seven Dwarfs, Cinderella and Sleeping Beauty: The Components of Costume Design in Disney’s Early Hand-Drawn Animated Feature Films. *Animation*, *13*(1), 7–19. https://doi.org/10.1177/1746847718754758

Kanazawa, A., Black, M. J., Jacobs, D. W., & Malik, J. (2018). End-to-End Recovery of Human Shape and Pose. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 7122–7131. https://doi.org/10.1109/CVPR.2018.00744

Kane, C. L. (2019). High-Tech Trash: Glitch, Noise, and Aesthetic Failure. In *High-Tech Trash: Glitch, Noise, and Aesthetic Failure*. https://doi.org/10.1525/luminos.83

Kennedy, D. (2003). The Oxford Encyclopedia of Theatre and Performance. In *The Oxford Encyclopedia of Theatre and Performance*. https://doi.org/10.1093/acref/9780198601746.001.0001

Kitagawa, M., & Windsor, B. (2008). MoCap for Artists : Workflow and Techniques for Motion Capture. In *Actual Problems of Economics* (Vol. 142, Issue 4). https://doi.org/10.33178/alpha.3.06

Knight, J., Johnston, A., & Berry, A. (2023). *Artistic control over the glitch in AI-generated motion capture*.

Kocabas, M., Athanasiou, N., & Black, M. J. (2020). *VIBE: Video Inference for Human Body Pose and Shape Estimation*.

Kocabas, M., Karagoz, S., & Akbas, E. (2019). Self-supervised learning of 3D human pose using multi-view geometry. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 1077–1086. https://doi.org/10.1109/CVPR.2019.00117

Kong, Q., Siauw, T., & Bayen, A. (2020). *Python numerical methods*.

Lachat, E., Macher, H., Mittet, M. A., Landes, T., & Grussenmeyer, P. (2015). First experiences with kinect V2 sensor for close range 3D modelling. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *40*(5W4), 93–100. https://doi.org/10.5194/isprsarchives-XL-5-W4-93-2015

Lapadat, J. C., & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry*, *5*(1), 64–86. https://doi.org/10.1177/107780049900500104

Lassner, C., Romero, J., Kiefel, M., Bogo, F., Black, M. J., & Gehler, P. V. (2017). Unite the people: Closing the loop between 3D and 2D human representations. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 4704–4713. https://doi.org/10.1109/CVPR.2017.500

Latulipe, C., Wilson, D., Gonzalez, B., Huskey, S., & Word, M. (2011). Temporal Integration of interactive technology in dance: Creative process impacts. *C and C 2011 - Proceedings of the 8th ACM Conference on Creativity and Cognition*, *November*, 107–116. https://doi.org/10.1145/2069618.2069639

Leavy, P. (2020). *Method Meets Art. Arts-Based Research Practice.*

Levin, G. (2006). Computer vision for artists and designers: Pedagogic tools and techniques for novice programmers. *AI and Society*, *20*(4), 462–482. https://doi.org/10.1007/s00146-006-0049-2

Liu, L., Long, D., Gujrania, S., & Magerko, B. (2019, October 10). Learning movement through human-computer co-creative improvisation. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3347122.3347127

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics*, *34*(6). https://doi.org/10.1145/2816795.2818013

Mahmood, N., Ghorbani, N., Troje, N. F., Pons-Moll, G., & Black, M. (2019). AMASS: Archive of motion capture as surface shapes. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 5441–5450. https://doi.org/10.1109/ICCV.2019.00554

Maranan, D. S., Alaoui, S. F., Schiphorst, T., Pasquier, P., Subyen, P., & Bartram, L. (2014). Designing for movement: Evaluating computational models using LMA effort qualities. *Conference on Human Factors in Computing Systems - Proceedings*, 991–1000. https://doi.org/10.1145/2556288.2557251

Marker Placement Guide. (2007). *Database, C M U Graphics Lab Motion Capture*.

Marois, B., & Syssau, P. (2008). LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1–3), 157–173.

Martinez, J., Hossain, R., Romero, J., & Little, J. J. (2017). *A simple yet effective baseline for 3d human pose estimation*.

Mccarthy, K. F., Brooks, A., Lowell, J., & Zakaras, L. (2001). *Performing Arts in a New Era*.

Mccormick, J., Vincs, K., Nahavandi, S., & Creighton, D. (2013). Learning to dance with a human. *Proceedings of the 19th International Symposium of Electronic Art, ISEA2013,*.

McDonald, K. (2019, September 10). *SO-FAR - Issue Article - Dance and Machine Learning: First Steps*. So-Far. https://so-far.xyz/issue/dance-and-machine-learning-first-steps

Mclaren, N. (2017). *Norman McLaren: Between the Frames*.

Meador, W. S., Rogers, T. J., O'Neal, K., Kurt, E., & Cunningham, C. (2004). Mixing dance realities. *Computers in Entertainment*, *2*(2), 12–12. https://doi.org/10.1145/1008213.1008233

MediaPipe | Google Developers. (n.d.-a). Retrieved April 18, 2023, from https://developers.google.com/mediapipe

MediaPipe | Google Developers. (n.d.-b). Retrieved April 18, 2023, from https://developers.google.com/mediapipe

Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W., & Theobalt, C. (2018). Monocular 3D human pose estimation in the wild using improved CNN supervision. *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, 506–516. https://doi.org/10.1109/3DV.2017.00064

Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Elgharib, M., Fua, P., Seidel, H.-P., Rhodin, H., Pons-Moll, G., & Theobalt, C. (2020). *XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera*. *39*(4), 1–24. https://doi.org/10.1145/3386569.3392410

Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Sridhar, S., Pons-Moll, G., & Theobalt, C. (2018). Single-shot multi-person 3D pose estimation from monocular RGB. *Proceedings - 2018 International Conference on 3D Vision, 3DV 2018*, 120–130. https://doi.org/10.1109/3DV.2018.00024

Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, H., Seidel, H., Xu, W., Casas, D. A. N., Theobalt, C., Rey, U., & Carlos, J. (2017). *VNect : Real-time 3D Human Pose Estimation with a Single RGB Camera*. *36*(4).

Menache, A. (2011). Understanding Motion Capture for Computer Animation. In *Understanding Motion Capture for Computer Animation* (pp. 1–46).

Merriaux, P., Dupuis, Y., Boutteau, R., Vasseur, P., & Savatier, X. (2017). A study of vicon system positioning performance. *Sensors (Switzerland)*, *17*(7). https://doi.org/10.3390/s17071591

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. *Journal of Environmental Psychology*, *14*(1), 336–337. https://doi.org/10.1016/0149-7189(96)88232-2

Miller, B., Fisher, K., Mateo, J., & David, I. (2006). Bebe Miller. *Dance Magazine*.

Mills, M. J. L., Sale, K. L., Simmons, B. A., & Popelier, P. L. A. (2017). Rhorix: An interface between quantum chemical topology and the 3D graphics program blender. *Journal of Computational Chemistry*, *38*(29), 2538–2552. https://doi.org/10.1002/jcc.25054

Mitchell, Th., Hyde, J., Tew, P., & Glowacki, D. (2018). *Art-Science Interrogations of Localization in Neuroscience*. *51*(2), 111–117. https://doi.org/10.1162/LEON

Miyoshi, E. (2021). *Artificial intimacy | SHIBUI collective*. Shibu Collective. https://www.shibuicollective.com/artificialintimacy

Moen, K. (2019). Expressive Motion in the Early Films of Mary Ellen Bute. *Animation*, *14*(2), 102–116. https://doi.org/10.1177/1746847719859194

Moura, J. M., Barros, N., & Ferreira-Lopes, P. (2019). *From real to virtual embodied performance-a case study between dance and technology*.

movementcomputing.org. (n.d.). Retrieved January 19, 2025, from movementcomputing.org

Mullis, E. (2013). Dance, interactive technology, and the device paradigm. *Dance Research Journal*, *45*(3), 111–123. https://doi.org/10.1017/S0149767712000290

Muybridge, E. (1882). *The Attitudes of Animals in Motion, illustrated with the Zoopraxiscope* (pp. 1–28).

Muybridge, E. (1899). *Muybridge_Animals in motion.pdf*.

Muybridge, E. (1901a). The Human Figure in Motion , by Eadweard Muybridge. In *The American Biology Teacher* (Vol. 25, Issue 8, pp. 638–638). https://doi.org/10.2307/4440492

Muybridge, E. (1901b). *The Human Figures in motion muybridge1901*.

Neagle, R. J., Ng, K., & Ruddle, R. A. (2004). Developing a Virtual Ballet Dancer to Visualise Choreography. *{Proceedings of the AISB 2004 Symposium on Language, Speech and Gesture for Expressive Characters (AISB'04})*, *January 2004*, 86–97.

Newell, A., Yang, K., & Deng, J. (2016). *Stacked Hourglass Networks for Human Pose Estimation 26 Jul 2016*.

Nie, Q., Liu, Z., & Liu, Y. (2021). Lifting 2D Human Pose to 3D with Domain Adapted 3D Body Concept. *International Journal of Computer Vision*, *131*(5), 1250–1268. https://doi.org/10.1007/s11263-023-01749-2

Nogueira, M. R., Menezes, P., & Maçãs de Carvalho, J. (2024). Exploring the impact of machine learning on dance performance: a systematic review. *International Journal of Performance Arts and Digital Media*, *20*(1), 60–109. https://doi.org/10.1080/14794713.2024.2338927

Okun, J. A., & Zwerman, S. (2021). *The VES Handbook of Visual Effects | ScienceDirect*. Focal Press.

Pagnon, D., Domalain, M., & Reveret, L. (2022a). *Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 2: Accuracy*.

Pagnon, D., Domalain, M., & Reveret, L. (2022b). Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software*, *7*(77), 4362. https://doi.org/10.21105/joss.04362

Pavlakos, G., Zhou, X., Derpanis, K. G., & Daniilidis, K. (2017). *Harvesting Multiple Views for Marker-less 3D Human Pose Annotations*. 1–10.

Pavlakos, G., Zhu, L., Zhou, X., & Daniilidis, K. (2018). Learning to Estimate 3D Human Pose and Shape from a Single Color Image. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 459–468. https://doi.org/10.1109/CVPR.2018.00055

Payne, W. C., Bergner, Y., West, M. E., Charp, C., Shapiro, R. B., Szafr, D. A., Taylor, E. V., & DesPortes, K. (2021). Danceon: Culturally responsive creative computing. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445149

Pettee, M., Shimmin, C., Duhaime, D., & Vidrin, I. (2019). Beyond imitation: Generative and variational choreography via machine learning. *Proceedings of the 10th International Conference on Computational Creativity, ICCC 2019*, 196–203.

Phillips, I. (2016). *The actor behind the new droid in "Rogue One" acted on stilts for the entire movie*. Insider. https://www.insider.com/star-wars-rogue-one-alan-tydyk-k-2so-stilts-2016-12

Piloto, C. (2022, December 26). *Artificial Intelligence vs Machine Learning: What's the difference?* https://professionalprograms.mit.edu/blog/technology/machine-learning-vs-artificial-intelligence/

Protopapadakis, E., Grammatikopoulou, A., Doulamis, A., & Grammalidis, N. (2017). Folk dance pattern recognition over depth images acquired via kinect sensor. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2W3), 587–593. https://doi.org/10.5194/isprs-archives-XLII-2-W3-587-2017

Qianwen, L. (2024). Application of motion capture technology based on wearable motion sensor devices in dance body motion recognition. *Measurement: Sensors*, *32*, 101055. https://doi.org/10.1016/j.measen.2024.101055

Raj, T., Hashim, F. H., Huddin, A. B., Ibrahim, M. F., & Hussain, A. (2020). A survey on LiDAR scanning mechanisms. In *Electronics (Switzerland)* (Vol. 9, Issue 5). MDPI AG. https://doi.org/10.3390/electronics9050741

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). *Hierarchical Text-Conditional Image Generation with CLIP Latents*. *Figure 3*.

Reddy, N. D., Guigues, L., Pishchulin, L., Eledath, J., & Narasimhan, S. G. (2021). TesseTrack: End-to-end learnable multi-person articulated 3D pose tracking. *Proceedings of the IEEE Computer*

Society Conference on Computer Vision and Pattern Recognition, 15185–15195. https://doi.org/10.1109/CVPR46437.2021.01494

Reicher, S., & Taylor, Stephanie. (2005). Similarities and differences between traditions. *Psychologist*, *18*(9), 547–549.

Rindler, R., Eshkar, S., & Kaiser, P. (1999). *Ghostcatching: A Virtual Dance Installation*. Riverbed Media.

Rizzo, A., El Raheb, K., Whatley, S., Cisneros, R. M., Zanoni, M., Camurri, A., Viro, V., Matos, J.-M., Piana, S., Buccoli, M., Markatzi, A., Palacio, P., Zohar, O. E., Sarti, A., Ioannidis, Y., & Fletcher, E.-M. (2018). WhoLoDancE: Whole-body Interaction Learning for Dance Education. *Proceedings of the Workshop on Cultural Informatics*. https://ceur-ws.org/Vol-2235/paper5.pdf

Rogez, G., Weinzaepfel, P., & Schmid, C. (2019). LCR-Net++: Multi-Person 2D and 3D Pose Detection in Natural Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(5), 1146–1161. https://doi.org/10.1109/TPAMI.2019.2892985

Rose, A. (2021). *Forgery – Australasian Dance Collective Dance Against The Machine*. Scenestr.Com.Au. https://scenestr.com.au/arts/forgery-australasian-dance-collective-dance-against-the-machine-20210819

Sadekar, K., & Mallick, S. (2020, February 25). *Camera Calibration using OpenCV | Learn OpenCV*. https://www.learnopencv.com/camera-calibration-using-opencv/

Sahni, M., Sahni, R., & M Merigo, J. (2022). Neural Networks, Machine Learning, and Image Processing. In *Neural Networks, Machine Learning, and Image Processing*. https://doi.org/10.1201/9781003303053

Saltz, D. Z. (2001). Live Media: Interactive Technology and Theatre. *Theatre Topics*, *11*(2), 107–130. https://doi.org/10.1353/tt.2001.0017

Sarafianos, N., Boteanu, B., Ionescu, B., & Kakadiaris, I. A. (2016). 3D Human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding*, *152*(September), 1–20. https://doi.org/10.1016/j.cviu.2016.09.002

Schiphorst, T., & Pasquier, P. (2015). *movingstories*. Interactions.Acm.Org. https://interactions.acm.org/enter/view/movingstories-simon-fraser-university

Schön, D. A. (1992). *The Reflective Practitioner : How Professionals Think in Action*. *1992*.

Scrivener, S. (2000). Reflection in and on action and practice in creative-production doctoral projects in art and design. *Working Papers in Art and Design*, *1*(January 2000), 1–13.

Segal, L. (2005). *Dance; REVIEW; "Landing/Place" is left dancing in dark: [HOME EDITION] - ProQuest*. Los Angeles Times. https://www-proquest-com.ezproxy.lib.uts.edu.au/docview/422069511?parentSessionId=pmX904HQj%2FpfEcClnRGpF2F0UNUBTqRdnfVbHaxR47Y%3D&pq-origsite=primo&accountid=17095

Serry, T., & Liamputtong, P. (2013). The in-depth interviewing method in health. In *Research methods in health: foundations for evidence-based practice*.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2011). *Real-Time Human Pose Recognition in Parts from Single Depth Images*.

Shuai, Q. (2021). *EasyMocap*. Github. https://github.com/zju3dv/EasyMocap

Sigal, L., Balan, A. O., & Black, M. J. (2010). HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, *87*(1–2), 4–27. https://doi.org/10.1007/s11263-009-0273-6

SimTK: OpenSim: Project Home. (n.d.). Retrieved April 18, 2023, from https://simtk.org/projects/opensim

Singh, A., Prakash, S., Kumar, A., & Kumar, D. (2022). A proficient approach for face detection and recognition using machine learning and high-performance computing. *Concurrency and Computation: Practice and Experience*, *34*(3), 1–12. https://doi.org/10.1002/cpe.6582

Sipe, D. (2020). Aesthetics and the methods of visual enquiry in the photography ofétienne-jules marey. *French Studies*, *74*(4), 554–571. https://doi.org/10.1093/FS/KNAA170

Smys, S., Tavares, J. M. R. S., Bestak, R., & Shi, F. (Eds.). (2021). *Computational Vision and Bio-Inspired Computing* (Vol. 1318). Springer Singapore. https://doi.org/10.1007/978-981-33-6862-0

Sommer, S. R. (1980a). Loie fuller's art of music and light. *Dance Chronicle*, *4*(4), 389–401. https://doi.org/10.1080/01472528008568817

Sommer, S. R. (1980b). Loie fuller's art of music and light. *Dance Chronicle*, *4*(4), 389–401. https://doi.org/10.1080/01472528008568817

Su, C., Lyu, M., Mähönen, A. P., Helariutta, Y., De Rybel, B., & Muranen, S. (2023). Cella: 3D data visualization for plant single-cell transcriptomics in Blender. *Physiologia Plantarum*, *175*(6), 1–7. https://doi.org/10.1111/ppl.14068

Synchronous Objects Archive Site. (n.d.). Retrieved October 19, 2024, from https://synchronousobjects.osu.edu/

Tadic, V., Odry, A., Kecskes, I., Burkus, E., Kiraly, Z., & Odry, P. (2019). *Application of Intel RealSense Cameras for Depth Image Generation in Robotics*.

Tokui, N. (2020, June 2). *A renowned dancer performed with an AI model — Can AI stimulate the dancer's creativity? | by Nao Tokui | Qosmo Lab | Medium*. https://medium.com/qosmo-lab/ai-and-a-renowned-dancer-performed-together-can-ai-stimulate-the-dancers-creativity-2e8715fd32d6

Tomlinson, L. (2014). The dance of the live and the animated: Performance animation by Kathy Rose, Miwa Matreyek and Eva Hall. *Animation Practice, Process & Production*, *3*(1), 17–55. https://doi.org/10.1386/ap3.3.1-2.17_1

Trajkova, M., Long, D., Desphande, M., Knowlton, A., & Magerko, B. (2024, May 11). Exploring Collaborative Movement Improvisation Towards the Design of LuminAI - a Co-Creative AI Dance Partner. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3613904.3642677

Tsuchida, S., Fukayama, S., Hamasaki, M., & Goto, M. (2019). Aist Dance Video Database : for Dance Information Processing. *Ismir 2019*, 501–510.

Tuckett, A. G. (2005). Applying thematic analysis theory to practice: a researcher's experience. *Contemporary Nurse : A Journal for the Australian Nursing Profession*, *19*(1–2), 75–87. https://doi.org/10.5172/conu.19.1-2.75

Ulaby, N. (2010). *Muybridge The man who made pictures move*. NPR. https://www.npr.org/2010/04/13/125899013/muybridge-the-man-who-made-pictures-move

University of Georgia. (2000, March 1). *Innovative Production of Shakespeare's "The Tempest."* Newswise. https://www.newswise.com/articles/innovative-production-of-shakespeares-the-tempest

Varol, G. (2018). *Learning from Synthetic Humans*. 1–10.

Vondrak, M., Sigaly, L., Hodgins, J., & Jenkinsx, O. (2012). Video-based 3D motion capture through biped control. *ACM Transactions on Graphics*, *31*(4). https://doi.org/10.1145/2185520.2185523

Walkerden, G. (2009). Researching and developing practice traditions using reflective practice experiments. *Quality and Quantity*, *43*(2), 249–263. https://doi.org/10.1007/s11135-007-9103-5

Wallace, B., & Martin, C. P. (2022). *Embodying the Glitch: Perspectives on Generative AI in Dance Practice*. *1*(1), 1–5.

Wang, Q., Kurillo, G., Ofli, F., & Bajcsy, R. (2015). Evaluation of pose tracking accuracy in the first and second generations of microsoft Kinect. *Proceedings - 2015 IEEE International Conference on Healthcare Informatics, ICHI 2015*, 380–389. https://doi.org/10.1109/ICHI.2015.54

Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8798–8807. https://doi.org/10.1109/CVPR.2018.00917

Zhang, Z. (2008). *A Flexible New Technique for Camera Calibration*.

Zhang, Z. (2021). Camera Extrinsic Parameters. In *Computer Vision* (pp. 77–77). Springer, Cham. https://doi.org/10.1007/978-0-387-31439-6_154

Zhang, Z., Wang, C., Qiu, W., Qin, W., & Zeng, W. (2021). AdaFuse: Adaptive Multiview Fusion for Accurate Human Pose Estimation in the Wild. *International Journal of Computer Vision*, *129*(3), 703–718. https://doi.org/10.1007/s11263-020-01398-9

11 Appendices

11.1 Ethics consent forms

Below is the ethics application and consent form sent to practitioners who use motion capture for performances:

Creativity & Cognition Studios

PARTICIPANT INFORMATION SHEET
Generating Abstract Animation for Performance Art
UTS HREC ETH19-3452

CCS Project Reference Number: 2022-4

WHO IS DOING THE RESEARCH?
My name is *Jamal Knight* and I am a student at UTS. My supervisor is *Prof. Andrew Johnston (Andrew.Johnston@uts.edu.au)*

WHAT IS THIS RESEARCH ABOUT?
This research is to find out more about the experiences of practitioners who work in the field of motion capture and/or performance art.

FUNDING
Funding for this project has been received from the Australian Government Research Training Program.

WHY HAVE I BEEN ASKED?
You have been invited to participate in this study because you have experience in the field of motion capture and/or performance art.

IF I SAY YES, WHAT WILL IT INVOLVE?
If you decide to participate, I will invite you to attend a zoom video call to answer some questions relating to your experiences with motion capture and/or performance art. The video and audio of the call will be recorded for transcribing purposes. Recording is voluntary and you may choose to not be recorded (video, audio or both). The interview will take 30 minutes to an hour to complete.

ARE THERE ANY RISKS/INCONVENIENCE?
Yes, there are some risks/inconvenience. You will be asked to give insights to your experience on your previous projects involving motion capture and/or performances. There may be questions asked about your present or past methods and techniques. There may be a risk that you may reveal information that may reduce your competitive advantage in the marketplace. However, if this does occur we will remove any of this information at your request.

DO I HAVE TO SAY YES?
Participation in this study is voluntary. It is completely up to you whether or not you decide to take part.

WHAT WILL HAPPEN IF I SAY NO?
If you decide not to participate, it will not affect your relationship with the researchers or the University of Technology Sydney. If you wish to withdraw from the study once it has started, you can do so at any time without having to give a reason, by contacting *Jamal Knight (jamal.knight@student.uts.edu.au)* or Prof. Andrew Johnston *(Andrew.Johnston@uts.edu.au)*.

If you withdraw from the study, any recordings and transcripts and participation will be erased.

CONFIDENTIALITY
By signing the consent form you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. Your name will be removed from any transcripts and your personal details will be anonymised. All identifying information will be erased. Only generalised information relating to your role will be known. Files will be stored in Aarnet's CloudStor online secure storage. Your information will only be used for the purpose of this research project and it will only be disclosed with your permission, except as required by law.

We plan to publish the results in academic publications including Phd thesis, journal and conference papers and/or books. In any publication, information will be provided in such a way that you cannot be identified.

WHAT IF I HAVE CONCERNS OR A COMPLAINT?
If you have concerns about the research that you think I or my supervisor can help you with, please feel free to contact me on jamal.knight@student.uts.edu.au or Prof. Andrew Johnston *(Andrew.Johnston@uts.edu.au)*.

You will be given a copy of this form to keep.

NOTE:
This study has been approved in line with the University of Technology Sydney Human Research Ethics Committee [UTS HREC] guidelines. If you have any concerns or complaints about any aspect of the conduct of this research, please contact the Ethics Secretariat on ph.: +61 2 9514 2478 or email: Research.Ethics@uts.edu.au], and quote the UTS HREC reference number: ETH19-3452. Any matter raised will be treated confidentially, investigated and you will be informed of the outcome.

Creativity & Cognition Studios

CONSENT FORM
Abstract Movement
UTS HREC ETH19-3452

CCS Project Reference Number: 2002-4

I _____ agree to participate in the research project 'Generating Abstract Animation for Performance Art' being conducted by Jamal Knight (jamal.knight@student.uts.edu.au) Tel: [redacted]. I understand that funding for this research has been provided by the Australian Government Research Training Program.

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research as described in the Participant Information Sheet.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time without affecting my relationship with the researchers or the University of Technology Sydney.

I understand that I will be given a signed copy of this document to keep.

I agree to be:

☐ Audio recorded
☐ Video recorded

I agree that the research data gathered from this project may be published in a form that:

☐ Does not identify me in any way
☐ May be used for future research purposes

I am aware that I can contact Jamal Knight if I have any concerns about the research.

_____ ____/____/____
Name and Signature [participant] Date

_____ ____/____/____
Name and Signature [researcher or delegate] Date

Witness to the consent process
If the participant, or if their legally acceptable representative, is not able to read this document, this form must be witnessed by an independent person over the age of 18. In the event that an interpreter is used, the interpreter may not act as a witness to the consent process. By signing the consent form, the witness attests that the information in the consent form and any other written information was accurately explained to, and apparently understood by, the participant (or representative) and that informed consent was freely given by the participant (or representative)

_____ ____/____/____
Name and Signature [witness*] Date

Below is the ethics application and consent form sent to the performer before the development of Interlinked.

Creativity & Cognition Studios

PARTICIPANT INFORMATION SHEET
MoCap performance
UTS HREC ETH19-3452

CCS Project Reference Number: 2023-3

WHO IS DOING THE RESEARCH?

My name is *Jamal Knight* and I am a student at UTS. My supervisor is *Prof. Andrew Johnston (Andrew.Johnston@uts.edu.au)*

WHAT IS THIS RESEARCH ABOUT?

The aim of the research is to assess the pipeline of creating a performance which uses motion capture generated by machine learning methods.
It will also document performers' experience in relation to other performances that they have participated in as a comparison.

FUNDING

Funding for this project has been received from the Australian Government Research Training Program.

WHY HAVE I BEEN ASKED?

You have been invited to participate in this study because you are an experienced dancer/performer and have experience with motion capture technology.

IF I SAY YES, WHAT WILL IT INVOLVE?

If you decide to participate, we will begin to plan the performance. The performance will be around two minutes in duration. We will start by viewing the space in the Data Arena to get an idea of what kind of motion is possible. I will send you the music for the performance, where you can begin to create the choreography. I will also provide examples of animation that will be driven by the motion capture you create from your movement, to get an idea of what the final look will be like. Once the choreography is decided, we will then capture your movement using RGB cameras. This data will be processed and cleaned, before applying to the animation. The animation will be displayed in the Data Arena simultaneously as you are performing the choreography.

The performance will be video recorded and the recording will be included in a PhD thesis and made publicly available.

Preparation for the performance may be conducted on floor 2, building 12 on the UTS campus upon confirmation.

After the performance, a zoom video call will be set up to ask you some questions relating to your experiences during the process.

The video and audio of the call will be recorded for transcribing purposes. The interview will take 30 minutes to an hour to complete.

ARE THERE ANY RISKS/INCONVENIENCE?

Yes, there are some risks/inconvenience. The floor in the Data Arena may not be ideal for any type of dancing style, as there are a few air-conditioning vents installed. The performance will need to be adapted to take this into consideration to avoid injury or damage to the vents. Some inconvenience might involve the amount of time coming up with the choreography, discussing the planning of the performance and the actual performance itself.

DO I HAVE TO SAY YES?

Participation in this study is voluntary. It is completely up to you whether or not you decide to take part.

WHAT WILL HAPPEN IF I SAY NO?

If you decide not to participate, it will not affect your relationship with the researchers or the University of Technology Sydney. If you wish to withdraw from the study once it has started, you can do so at any time without having to give a reason, by contacting *Jamal Knight (jamal.knight@student.uts.edu.au)* or Prof. Andrew Johnston *(andrew.johnston@uts.edu.au)*.

If you withdraw from the study, any recordings and transcripts and participation will be erased.

CONFIDENTIALITY

By signing the consent form you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. Files will be stored in Aarnet's CloudStor online secure storage. Basic information like your name and performance resume will be recorded for the purposes of documenting the performance. The performance will be video recorded. Your information will only be used for the purpose of this research project and it will only be disclosed with your permission, except as required by law.

We plan to publish the results in academic publications including Phd thesis, journal and conference papers and/or books.

WHAT IF I HAVE CONCERNS OR A COMPLAINT?

If you have concerns about the research that you think I or my supervisor can help you with, please feel free to contact me on jamal.knight@student.uts.edu.au or Prof. Andrew Johnston *(andrew.johnston@uts.edu.au)*.

You will be given a copy of this form to keep.

NOTE:

This study has been approved in line with the University of Technology Sydney Human Research Ethics Committee [UTS HREC] guidelines. If you have any concerns or complaints about any aspect of the conduct of this research, please contact the Ethics Secretariat on ph.: +61 2 9514 2478 or email: Research.Ethics@uts.edu.au], and quote the UTS HREC reference number: ETH19-3452. Any matter raised will be treated confidentially, investigated and you will be informed of the outcome.

Creativity & Cognition Studios

CONSENT FORM
MoCap performance
UTS HREC ETH19-3452

CCS Project Reference Number: 2023-3

I _____ agree to participate in the research project 'MoCap performance' being conducted by Jamal Knight (jamal.knight@student.uts.edu.au) Tel: [redacted]. I understand that funding for this research has been provided by the Australian Government Research Training Program.

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research as described in the Participant Information Sheet.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time without affecting my relationship with the researchers or the University of Technology Sydney.

I understand that I will be given a signed copy of this document to keep.

I agree to be video/audio recorded. I give permission for the video of the performance to be publicly released.

I agree that the research data gathered from this project may be published in a form that identifies me.

☐ I agree the data may be used for future research purposes.

I am aware that I can contact Jamal Knight if I have any concerns about the research.

_____ ____/____/____
Name and Signature [participant] Date

_____ ____/____/____
Name and Signature [researcher or delegate] Date

Below is the ethics application and consent form sent to the participants who attended the performance Interlinked, who volunteered their time to be interviewed.

Creativity & Cognition Studios

PARTICIPANT INFORMATION SHEET
MoCap/performance
UTS HREC ETH19-3452

CCS Project Reference Number: 2023-3

WHO IS DOING THE RESEARCH?

My name is *Jamal Knight* and I am a student at UTS. My supervisor is *Prof. Andrew Johnston (Andrew.Johnston@uts.edu.au)*

WHAT IS THIS RESEARCH ABOUT?

The aim of the research is to assess the pipeline of creating a performance which uses motion capture generated by machine learning methods.
It will also document performers' experience in relation to other performances that they have participated in as a comparison.

FUNDING

Funding for this project has been received from the Australian Government Research Training Program.

WHY HAVE I BEEN ASKED?

You have been invited to participate in this study because you are an experienced dancer/performer and have experience with motion capture technology.

IF I SAY YES, WHAT WILL IT INVOLVE?

If you decide to participate, we will begin to plan the performance. The performance will be around two minutes in duration. We will start by viewing the space in the Data Arena to get an idea of what kind of motion is possible. I will send you the music for the performance, where you can begin to create the choreography. I will also provide examples of animation that will be driven by the motion capture you create from your movement, to get an idea of what the final look will be like. Once the choreography is decided, we will then capture your movement using RGB cameras. This data will be processed and cleaned, before applying to the animation. The animation will be displayed in the Data Arena simultaneously as you are performing the choreography.

The performance will be video recorded and the recording will be included in a PhD thesis and made publicly available.

Preparation for the performance may be conducted on floor 2, building 12 on the UTS campus upon confirmation.

After the performance, a zoom video call will be set up to ask you some questions relating to your experiences during the process.

The video and audio of the call will be recorded for transcribing purposes. The interview will take 30 minutes to an hour to complete.

ARE THERE ANY RISKS/INCONVENIENCE?

Yes, there are some risks/inconvenience. The floor in the Data Arena may not be ideal for any type of dancing style, as there are a few air-conditioning vents installed. The performance will need to be adapted to take this into consideration to avoid injury or damage to the vents. Some inconvenience might involve the amount of time coming up with the choreography, discussing the planning of the performance and the actual performance itself.

DO I HAVE TO SAY YES?

Participation in this study is voluntary. It is completely up to you whether or not you decide to take part.

WHAT WILL HAPPEN IF I SAY NO?

If you decide not to participate, it will not affect your relationship with the researchers or the University of Technology Sydney. If you wish to withdraw from the study once it has started, you can do so at any time without having to give a reason, by contacting *Jamal Knight (jamal.knight@student.uts.edu.au)* or Prof. Andrew Johnston *(andrew.johnston@uts.edu.au)*.

If you withdraw from the study, any recordings and transcripts and participation will be erased.

CONFIDENTIALITY

By signing the consent form you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. Files will be stored in Aarnet's CloudStor online secure storage. Basic information like your name and performance resume will be recorded for the purposes of documenting the performance. The performance will be video recorded. Your information will only be used for the purpose of this research project and it will only be disclosed with your permission, except as required by law.

We plan to publish the results in academic publications including Phd thesis, journal and conference papers and/or books.

WHAT IF I HAVE CONCERNS OR A COMPLAINT?

If you have concerns about the research that you think I or my supervisor can help you with, please feel free to contact me on jamal.knight@student.uts.edu.au or Prof. Andrew Johnston *(andrew.johnston@uts.edu.au)*.

You will be given a copy of this form to keep.

NOTE:

This study has been approved in line with the University of Technology Sydney Human Research Ethics Committee [UTS HREC] guidelines. If you have any concerns or complaints about any aspect of the conduct of this research, please contact the Ethics Secretariat on ph.: +61 2 9514 2478 or email: Research.Ethics@uts.edu.au], and quote the UTS HREC reference number: ETH19-3452. Any matter raised will be treated confidentially, investigated and you will be informed of the outcome.

Creativity & Cognition Studios

CONSENT FORM
MoCap performance
UTS HREC ETH19-3452

CCS Project Reference Number: 2023-3

I _____ agree to participate in the research project 'MoCap performance' being conducted by Jamal Knight (jamal.knight@student.uts.edu.au) Tel: [redacted]. I understand that funding for this research has been provided by the Australian Government Research Training Program.

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research as described in the Participant Information Sheet.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time without affecting my relationship with the researchers or the University of Technology Sydney.

I understand that I will be given a signed copy of this document to keep.

I agree to be video/audio recorded. I give permission for the video of the performance to be publicly released.

I agree that the research data gathered from this project may be published in a form that identifies me.

☐ I agree the data may be used for future research purposes.

I am aware that I can contact Jamal Knight if I have any concerns about the research.

_____ ____/____/____
Name and Signature [participant] Date

_____ ____/____/____
Name and Signature [researcher or delegate] Date

11.2 Interview questions

The interview questions for practitioners who utilise motion capture are provided below (Section 3.3.2):

1) What have you used motion capture for?
2) Which method/s of motion capture have been used in your research/performances?
3) Why was/were those method/s selected?
4) In your experience, what are the benefits of those method/s?
5) In your experience, what are the drawbacks of those method/s?

6) How do you incorporate motion capture into the development of performances? i.e. How does it fit into the creative process?
7) What kind of feedback, good or bad, did you get from the performers about the use of motion capture?
8) What do you look for in performers when creating works that utilise motion capture?
9) What is your process of reviewing motion capture data? i.e. Reviewing after capture, watching the performance and re-taking any actions, or watching a live feed of the animation in real-time?
10) What was your best experience with motion capture?
11) Have you had any particular bad experiences with motion capture? The biggest motion capture failure you have had to deal with?
12) What do you feel most inhibits your use of motion capture? i.e. what are the reasons why you would choose not to use it?
13) For you, how important is accuracy when using motion capture?
14) On a scale of 1-10, for you, how important is: accuracy, cost, portability, setup time, capture area, adaptability to different lighting conditions when setting up a motion capture system?
15) If technology allowed you to improve the setup time, cost, portability and size of capture area for a motion capture system, but those improvements came at the expense of accuracy, would the trade-off be worth it and would it unlock any new opportunities for you and your work?
16) Is there anything else do you want to say/any additional comments?

The interview questions for the audience after the performance are provided below:

1) How would you describe the overall experience of the performance?
    a. Were there any specific moments in the performance that stood out to you? Why?
2) Can you describe how you perceive the relationship between the dancer and the animated projection?
3) How would you compare this performance to other dancer performances you have seen?

The interview questions for the performer after the Interlinked performance are provided below:

1) What is your prior experience when performing with motion capture systems?
2) Please talk briefly about the process of the motion capture process to performing with that system.
3) How does your experience in preparing for and performing differ compared to ‘live’ motion capture vs ‘offline/pre-recorded/ motion capture?
4) How did you design the choreography for the data arena based on what you know of the system?
5) What other opportunities do you see in this space with performance?

11.3 Publications

During this PhD, I collaborated on the following publications, some of which explore works or concepts related to this thesis:

Knight, J., Johnston, A., & Berry, A. (2022). Machine Art: Exploring Abstract Human Animation Through Machine Learning Methods. *ACM International Conference on Movement and Computing (MOCO)*

Knight, J., Johnston, A., & Berry, A. (2023). *Artistic control over the glitch in AI-generated motion capture*. *Creativity and Cognition International Conference*