12-2012

Motion Capture Study of Human Movement Recognition

Nicholas Kinerd
Clemson University, nkinerd@gmail.com

MOTION CAPTURE STUDY OF HUMAN MOVEMENT RECOGNITION

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Fine Arts
Digital Production Arts

---

by
Nicholas I. Kinerd
December 2012

Accepted by:
Dr. Brian A. Malloy, Committee Chair
Dr. Andrew T. Duchowski
Dr. Donald H. House

Abstract

In this thesis, we address questions about recognition of human movement through motion capture. We begin by capturing the martial arts movements of four actors as they each perform three different techniques. We then conduct a survey to determine if participants in the survey are able to distinguish the various martial arts techniques and differentiate the actors that portray said techniques. A secondary consideration of our survey is to determine if the participants can distinguish the sex of the actors as they perform the martial arts techniques. Our hypothesis is that people are able to distinguish actors and sex; however, determining sex may be more difficult with digital actor representations that are anatomically ambiguous. With this thesis we will attempt to provide insight into human perception and motion capture, and help validate the use of motion capture in video games and movies for realistic human animation and interactions and to help improve the immersion of the player/viewer.

Dedication

We fall so that we may learn to pick ourselves up.

Thomas Wayne

Acknowledgments

I would like to acknowledge the following people for their help with my thesis: my advisor, Dr. Brian Malloy for his patience and guidance; Dr. Andrew Duchowski for his assistance with making this study possible; Dr. Donald House for allowing us to pave the way for this different type of thesis; my roomate and best friend, Darius Jones for his support when I needed him most; my girlfriend, Abbie Tremblay for building up my confidence and supporting me when others did not; and lastly, my family for always believing in me. I love you all. Thank you.

Table of Contents

Table of Contents (Continued)

List of Figures

Page

Chapter 1

Introduction and Motivation

Video games and movies have been using motion capture, the process of recording the movements of objects or persons, to perform animation since the mid 1990s [12]. This type of capture can provide more realistic animation than standard animation techniques and also allows subtle human expression to be captured. Subtle expressions can be difficult to portray using regular keyframe animation. Currently, motion capture is used for other applications as well. For example, the medical field uses motion capture to provide biometric feedback to patients, and motion capture is used to build training videos for employees.

In the film industry, motion capture continues to be used for virtually all types of animation, as well as live streaming. Live streaming allows directors to see a scene rendered in real time and to direct the actors during the capture of the scene. Live streaming is also used in video games, and some game developers attempt to use movie stars and sports icons for motion capture. Using celebrities allows game players to place themselves in the shoes of someone famous. Recently, the Kinect has been developed for the Microsoft Xbox game console; the Kinect provides a type of motion capture that captures the players and uses their basic movements and gestures to control game characters. This use of motion capture is an advancement for video games, which may one day allow players to perform very complex moves with their motions and these moves can be incorporated into games and mimicked by game characters in real time. However, in order to fully achieve this technology, the science of motion capture must be further developed to provide better understanding of the effects of using motion capture, sports figures and actors on the immersion of the player into the video game.

The purpose of this thesis is to address questions relating to the use of motion capture in video games. Many sports games, and action games that use martial arts, use different actors to portray movements, such as swinging a bat or kicking a ball, that may look different due to each actor's unique style. However, it is an open question as to whether or not the use of different actors to portray these movements is necessary. For example, can someone playing a sports game discern the difference between two different actors performing the same action?

To address this question of actor discrimination, we begin by capturing martial arts movements performed by four actors, two males and two females, as they perform three different techniques. We then conduct a survey to determine if the participants in the survey are able to distinguish the various martial arts techniques. A secondary consideration of our survey is to determine if participants can distinguish the sex of the actors as they perform the moves. Our hypothesis is that people are able to distinguish actors and sex; however, determining sex may be more difficult with relatively ambiguous digital actor representations. It is our intention that this thesis will provide insight into human perception and motion capture, and validate the use of motion capture in video games and movies for realistic human animation and interactions and to improve the immersion of the player.

In the next chapter, we provide background information and definitions about the terms and concepts that we use in this thesis. In Chapter 3 we describe previous research about motion capture and identification, and in Chapter 4 we describe the process whereby we capture the movements of four actors as they perform three distinct martial arts techniques. In Chapter 5 we describe the process of building an interface and the practice of performing a survey to capture participant responses about the martial arts techniques. In Chapter 6 we describe the results of the study and in Chapter 7 we discuss the results and describe some problems with the study and summarize some approaches to solving these problems for future research. Finally, in Chapter 8 we draw some conclusions.

Chapter 2

Background

In this section, we review terms and concepts that we use in this thesis. We begin by describing motion capture and the use of markers, and we then describe the software that we use to process the data. In Section 2.1.1 we review the use of motion capture in video games, and we conclude the section in 2.1.2 by describing the particular form of martial art movements that we identify in our study.

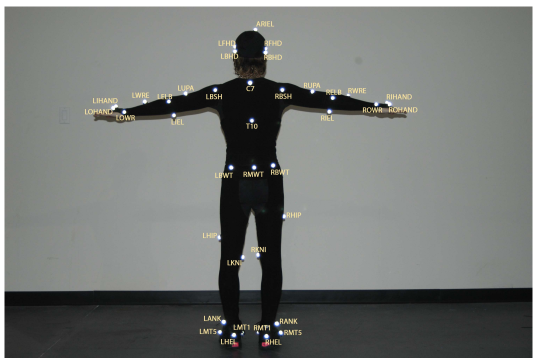


Figure 2.1: **Reflector Locations**. This figure illustrates the placement of reflectors on major joints and bone locations.

2.0.1 Capturing Form and Motion

Motion capture is a process that builds a three-dimensional representation of a live figure or performance. In most motion capture systems, receivers are set around or on an actor, and sensors provide data to the receivers and capture information about the movements of the actor. Information is retrieved from the receivers and transmitted to hardware that translates the data, and then applies it to a digital character. Different types of motion capture (MoCap) techniques have been developed, including *passive marker*, *active marker*, *markerless*, *inertial*, and *magnetic* [12; 7]. These techniques differ in the type of equipment that is used to track the motions that the actor exercises. In our study, we use *passive markers*, which are orbs coated in reflective material. For human participants, the reflective orbs are placed on major joints and bones of the body, illustrated in Figure 2.1. Cameras, placed around the area of capture, emit or strobe infrared light that bounces off the reflective orbs and back to the camera, where it is transmitted to a computer for processing.

2.0.2 Blade Software

For our study, we use the passive marker system by Vicon, and we process the data using the corresponding *Blade* software [2]. The Blade software records and processes the captured data and is used during all aspects of the motion capture performances utilized in our study, including set-up, calibration, recording, post-processing, and creating and exporting the captured videos for each performance.

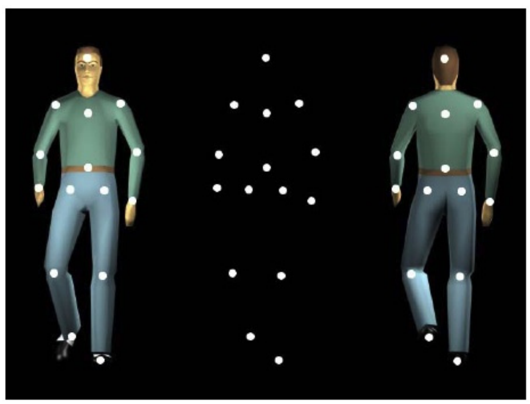


Figure 2.2: **Point Light Display**. This figure illustrates a point light display and how the dots line up on a 3D character [1].

2.1 Point Light Displays

Reference to point light displays(PLDs) may be noted in the *Related Works* chapter. These point light displays are considered the precursor to certain types of motion capture. Figure 2.2 references a point light display. A PLD is essentially a set of (usually) white dots on a contrasting (usually black) background. When the dots are arranged in certain patterns relating to major joints on a person/animal's body the dots represent said person/animal. Examples are provided in the *Related Works* chapter. This arrangement of PLDs can also be referred to as biological motion or the visual perception of biological motion. This technique was first demonstrated by Gunnar Johansson in 1973[6].

Figure 2.3: **Baseball Mocap**. This image shows Minnesota Twin's catcher, Joe Mauer, participating in motion capture for the PS3 game, MLB 2010 The Show[14].

2.1.1 Motion Capture in Video Games

One of the video games that first exploited motion capture was *Soul Edge*, developed by Namco in 1995, Since then, video games have been capturing motions such as fight sequences, tackles, pitches, and punches, to engender a more realistic experience for the player [4]. Video games have more potential than movies to completely immerse the player into their world. This immersian is enhanced for many action and sports games through the use of configurable characters and motion captured animations of actual sports figures and players. For example, a 2004 video game, *Jet Li: Rise to Honor*, features captured motions by the martial arts legend, Jet Li, providing an enhanced immersion into the game [5; 8]. Many current action and sports video games use motions captured from athletes to enhance the game experience for the player. An example of an athlete in motion capture for a video game is illustrated in Figure 2.3.

2.1.2 Martial Arts

The style of martial art that we use in our study is *Cuong Nhu*, pronounced "Kung New", which is actually a mix of seven other styles of martial arts. These styles include: Western

Boxing, Shotokan Karate, Judo, Wing Chun, Vovinam, Aikido, and Tai Chi Chuan. Each of these styles brings a balance to Cuong Nhu between hard and soft martial arts. The harder styles incorporate strengthening one's self, using striking contact, and power in order to overcome one's opponent. Examples of hard style influence include, Western Boxing, Shotokan Karate, and some parts of Wing Chun and Vovinam. Softer styles incorporate using the opponent's body or energy against them. Blocks guide and deflect instead of using raw power to stop the attack. Examples of Cuong Nhu's soft style influences include, Aikido, Judo, Tai Chi Chuan, and some parts of Wing Chun and Vovinam. In Vietnamese "Cuong Nhu" means "hard soft" [3].

Chapter 3

Related Work

McLendon et al. investigated creature motion recognition using point light displays (PLDs) [11]. Their goals were to find recognizable traits in animals that may be communicated through motion using aspects of a point light display, such as size and attitude, and to use this information to facilitate management of animation. These goals are similar to ours so that both studies attempt to recognize traits of the actor; in the case of McLendon et al. the goal is to recognize types of animals and in our case the goal is to recognize martial arts movements. In their study, participants viewed 20 videos of 10 animals. In the first viewing, the animals were represented as PLDs, and in the second viewing the animals were shown in full video. When viewing the full videos, the participants were able to identify the animals in all cases. When viewing the PLDs, the participants were able to recognize the animals in 25 % of the cases, and this recognition varied by a large margin for some of the animals. For example, participants were able to recognize the kangaroo 50 % of the time; however, no participants were able to recognize the bear in PLD form. Comparatively, In our related study, participants viewed 12 videos of two actors side by side and were asked to determine if the same movement was performed in both videos. Our results concluded that people could easily tell if the actors were doing the same move or not.

Loula, et al. studied the recognition of people from their movement [9]. This study is similar to ours in that both studies are attempting recognition from human actors doing certain types of movements. While the Loula et al. study based their experiment on several different types of human movements, our study focuses on martial arts movements. In their study the actors performed the experiments involving different motions and came back 3 months later to view PLDs of their movements. The actors were asked to distinguish between themselves, their friends, and strangers for each clip. The results of this study showed that 69 % of actors were able to identify themselves doing actions based on the

PLDs. In our study we did not use the actors in the survey part of the experiment so all of our participants and actors were strangers. This had seemingly no effect on our results of whether strangers could identify differentmovements but may have had an efffect on the results of how participants were able to distinguish the different actors performing the same move.

McDonnell et al. researched the effects that form and motion had on sex perception [10]. This study relates to ours a lot, especially since they used motion capture to capture their movements. However, the main difference is that the McDonnell et al. study focused on distinguishing sex, but for our study the main focus is on distinguishing actors regardless of sex. However, even though it is not the main purpose, our study also includes distinguishing sex as a secondary goal. The experiment done by McDonnel et al. yielded mostly positive results. When their motion captured walks were placed on an androgynous model, most responses were accurate for male and female walks. In our study, we did not fair as positively with our results for sex. Reasons for our outcome will be discussed in the *Results* and *Discussion* sections.

Pyles et al. studied visual perception and neural correlation as it related to biological motion[13]. In this set of studies the authors looked at whether or not the motions from created apparatus can be animated to look alive. This paper relates to our work due to the fact that they were trying to recognize motion from their characters. Some of the characters were human, some were created with software and modeled in different apparatus shapes. Participants were then asked if the PLDs and modeled characters were alive. Their results yielded that half of the creatures were thought to be alive, and almost all of the human motions were percieved as being alive. Although we did not do any work with real or fake creatures, both studies use basic human movement recognition as a control. Our results compared with theirs, as almost all of the participants were able to distinguish an actor performing different moves.

Chapter 4

Motion Capture Methodology

In this chapter we explain our methodology in designing the experiment and conducting a study to evaluate our motion capture technique. The process requires several steps and we look into each part separately as it relates to the work being done for the study. Our use of motion capture drives the study that we conduct. The type of motion capture that we use employs a passive marker system. At Clemson University, the Digital Production Arts (DPA) lab provides motion capture equipment and space to perform the capture. Also, we chose Cuong Nhu's martial arts techniques because of the complexity of the movements. The martial arts techniques permit more complex movements than those provided by movements such as walking or jumping. The martial arts techniques also provide a link to video games, because many fighting games use martial arts, and the ability to use actual martial arts techniques has the potential to increase immersion into the game and to increase the realism of playing the game.

We use several martial artists (herby referred to as *actors*) to capture the same technique performed by different actors. Part of our study is to determine if these variations obfuscate identification of the technique. We record videos of each actor's techniques and then videos of two actors performing either the same technique or different techniques are shown side-by-side. Participants in our study then view the videos and are asked questions that relate to our hypothesis. These questions are posed to the participants through an automated interface that factors our personal characteristics of the questionnaire administrator.

4.1 Acquiring Actors

To perform the martial arts techniques, we chose actors who were high ranking students in Cuong Nhu. These students possess knowledge of all the techniques used for the performance and have experience doing these techniques. To make the study more consistent, certain actors were not chosen due to exaggerated height or weight differences. The four actors that we chose are three upper rank students and one Sensei (instructor). There are two male actors and two female actors. The first set of techniques that we capture include a basic low block, followed by a punch, and then a kick. The next two sets of techniques that we use are progressively more complex and stylistic while still being easy to perform by the actors. Figure 4.1 shows an example of a technique being performed. Prior to the actual capture, each technique is practiced with all four actors in a group so that each of the techniques are performed within roughly the same time frame, in an effort to prevent the actors from doing the same technique at radically different speeds, which may confuse participants in the survey.

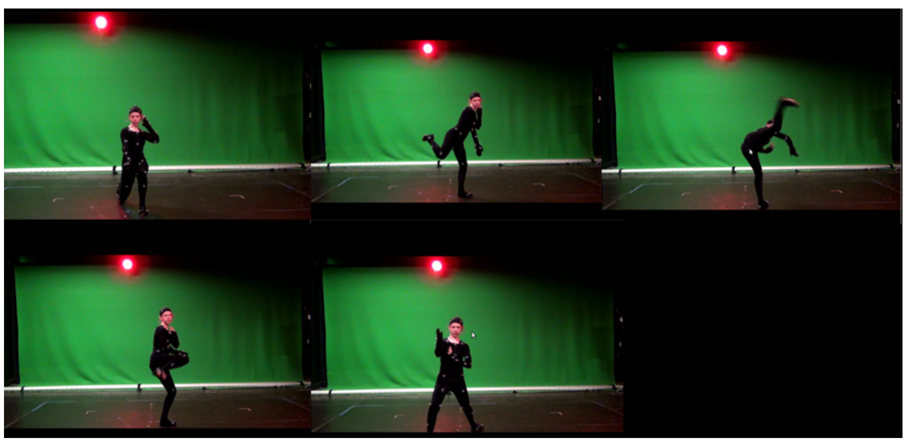


Figure 4.1: **Acting out**. This series of captures show the progression of one type of martial art techniques.

4.2 Calibration

4.2.1 Camera Calibration

After technique rehearsal, each actor is suited and filmed in the motion capture studio. In order to start the motion capture sequence the equipment must be calibrated. After equipment calibration, each actor is also calibrated, a multi-step process that proceeds as follows. Connecting through a computer, the Blade software can connect each of the eight infrared cameras. In order to increase the accuracy of this calibration, all reflective materials not in use must be put away and out of sight from the cameras. This can be tested by turning on the live calculation settings in the Blade perspective view. From this perspective, anything being captured by the infrared cameras will show up as a white dot as shown in Figure 4.2. During our experience with the use of this motion capture system the overhead lights became a large source of reflectivity and were shut off in order to not interfere with other mocap data.

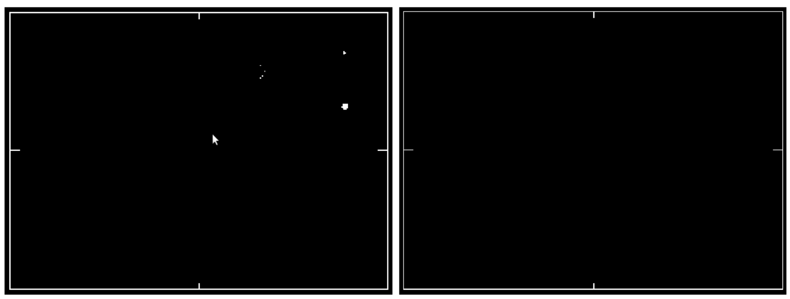


Figure 4.2: **Clearing the Background**. This figure illustrates the camera space before and after it is cleared. This can be done manually by clearing reflective objects or masking them in Blade.

Once the infrared cameras are turned on and connected to the system and the area is clear of reflective debris, the cameras need to be calibrated (see Figure 4.3) by a series of wand waves with a 240mm wand. The calibration wand contains three small reflective sensors which test the reflectivity and accuracy of the cameras. In order to perform the calibration a person needs to walk around the perimeter of the motion capture scene with the wand in order to test all areas of the space. After starting the calibration sequence in Blade, the calibrator holds the wand out in front of their body and motioning with forward figure-eight arm movements walks the perimeter of the motion capture area, allowing the wand sensors to reach high and low areas within the capture space. Once the Blade software captures approximately 3-5,000 frames (about a minute's worth of walking) the person calibrating stops the process. Blade then calculates each camera's accuracy and informs the person calibrating if any of the cameras need to be adjusted. If a camera is shown as anything less than *Good*, a camera re-calibration should be performed or the cameras need to be adjusted. For our experiment, calibration before each of our motion capture sessions is performed. During our calibrations, each of the eight cameras calibrated at *Good, Excellent*, or *Awesome* as shown in Figure 4.4.

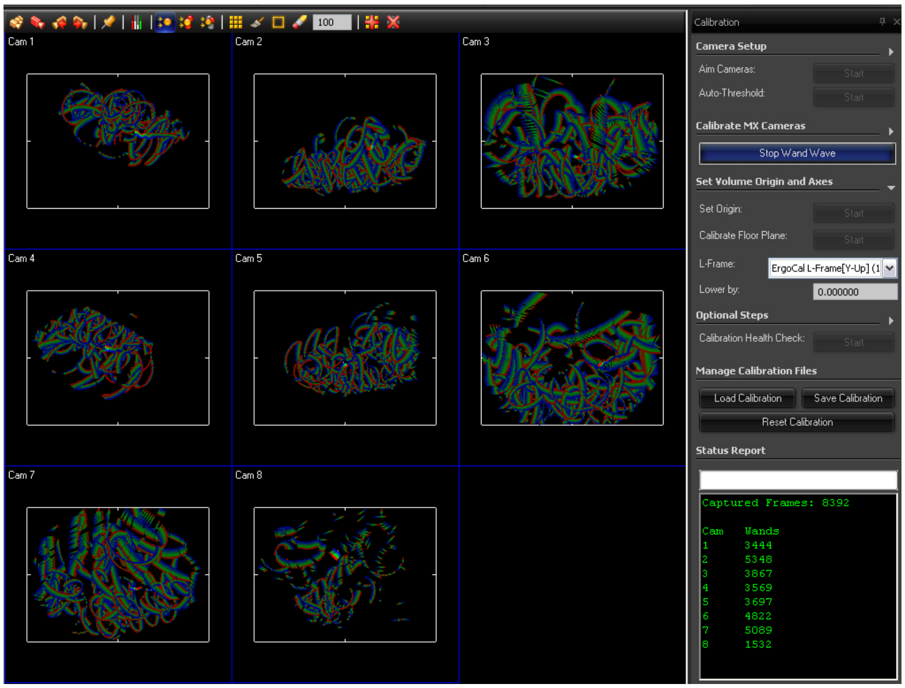


Figure 4.3: **Wand Waving Calibration**. This figure illustrates an example of the calibration wand wave process. Each camera captures what it can "see" from the wand as it moves around.

After the wand wave calibration an EngoCal L-frame is used for camera placement calibration in order to establish the center of the mocap area. The triangle shaped calibration tool rests on the floor and contains four small reflective sensors as shown in Figure 4.5. Assuming the Y-axis will be perpendicular to the ground, the side with three aligned sensors are placed along the X-axis of the floor in relation to the viewing plane in Blade. Once this is complete Blade now knows the position of the cameras and can position them in the perspective viewing plane.

One final camera calibration tool is the floor plane calculation. This is an optional calibration that can adjust the floor plane of viewing area if the actual floor of the mocap space is not level. This is done by scattering 4 or more reflective markers on the floor and

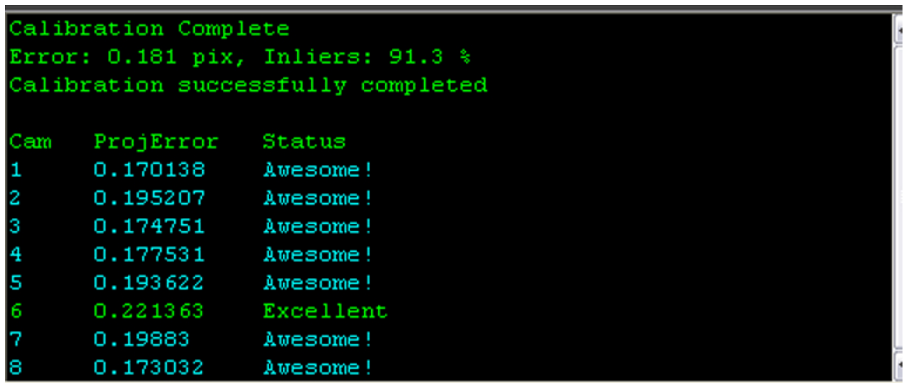


Figure 4.4: **Wand Wave complete**. This figure shows the completed analysis of the wand wave.

their relative heights are calculated and Blade adjusts the floor plane to match the marker data. Camera calibration is performed at the beginning of each shooting day.

4.2.2 Actor Calibration

After camera calibration, each actor must also be calibrated to the motion capture system, which we explain in this section. The first step is to equip an actor with a suit and place reflective markers on the suit in 53 specific joint and bone locations (for reference see Figure 2.1 in the *Background* chapter). We choose the 53 marker setup, which does not include separate markers for all the fingers or toes, because none of the techniques require the level of detail for individual toe and finger recognition. Thus, the 53 marker setup is sufficient for our martial arts technique recognition.

All parts of the suit must be as tight and secure on the actor as possible in order to keep the markers from moving during the capture. Any loose part of the suit must be tightened with Velcro strips. After the actor is suited up, the next step is to perform his or her calibration. The actor calibration is performed in Blade and utilizes a Range of Motion capture (ROM) which goes over most of the basic body motions that might be used in a scene. It is also a good way for the actor to warm up their body in the suit and get used to moving in it. Examples of ROM motions include: neck rolls, arm rotations, body twists, and simple stretches. Once the ROM is captured the data is then processed in Blade and

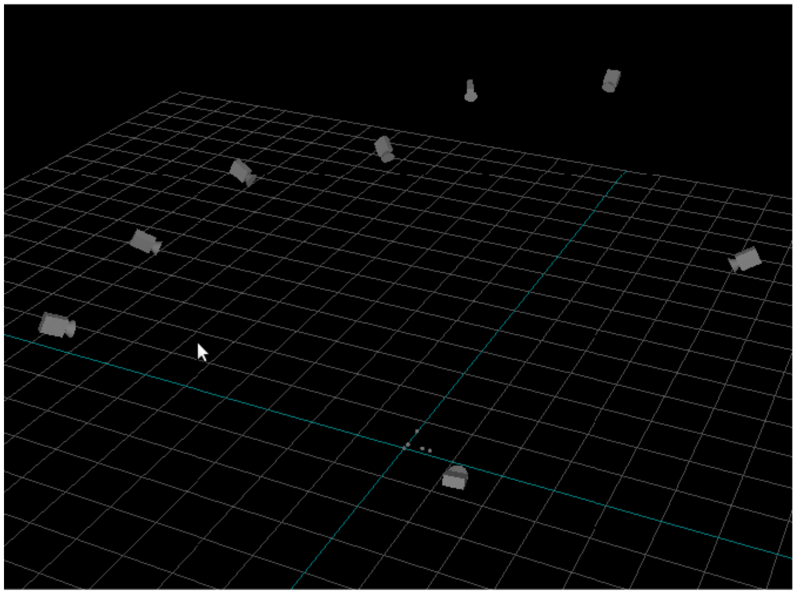


Figure 4.5: **Camera Setup**. This figure illustrates an example of the perspective view of the camera setup. The cameras are located based on the 4 marker L-frame shown on the ground plane.

cleaned. A .vsk is then created by the software. This .vsk is a skeletal structure like a character rig that is created from the movements of the ROM and will be used for that actor in any future scenes that day. As long as the actor does not take off the suit, the same .vsk can be used. For this reason, we give each actor a day and shoot all of their techniques in that day.

4.3 Capturing Data

Once the actor is calibrated, the scene is ready to be shot. As mentioned in the *Acquiring Actors* section, each actor is going to perform three series of techniques. To obtain continuity, each set of techniques is performed multiple times until at least three good takes are captured. Each technique is practiced a few times before any capturing commenced in case the actor has any questions about timing or technique specifics like stances or hand/foot positions.

The filming of each take consists of five parts. In the beginning of each take the actor starts in a T-pose with arms held straight out to the side and legs shoulder width apart. This is held for about one second. Next the actor brings their arms down in front of them in a natural ready stance that is often used in Cuong Nhu before techniques are performed. After a brief pause in the ready stance the actor begins the technique set. Once finished with the set the actor returns to the ready stance and after a brief pause returns to the T-pose. After the T-pose is held for an aditional second the capture take ends. From start to finish the capturing process takes around 8-11 seconds depending on the technique being done. This process is repeated for each set of takes for each technique. The three best takes for each technique are chosen. Thus, a minimum of nine takes are needed to be captured, with most actors having more than nine takes.

4.4 Cleaning Data

The passive marker system of motion capture allows for high accuracy marker detection; however, occlusion can occur if a marker interferes with another marker or if not enough cameras can see the marker. Thus, varied steps of post processing must occur after the take including some cleaning of the data. The sections below follow the steps of cleaning the captured data.

Once the performance is captured the data must be processed with the Blade software. A .x2d file is created after each capture. The .x2d file is the raw data from the cameras and must be compiled and translated into 3d data in order to clean it. Figure 4.6 shows the data management screen where each take is logged by day and scene. Double clicking on the .x2d icon brings each camera's data into view as shown in Figure 4.7. At this point it is possible to view the data collected from each camera individually, but until it is processed a 3d representation cannot be viewed in the perspective viewscreen. There are several ways to process the data at this point. The easiest way is to use the provided pipelines for post processing data. The pipelines in Blade are preset lists of commands that enable a user to follow step by step in order to complete certain functions. In most cases the

Basic Post Processing pipeline is sufficient. Because of the camera set up in the DPA lab, a more in depth post processing is needed so we use a mixture of Basic and Advanced post-processing.

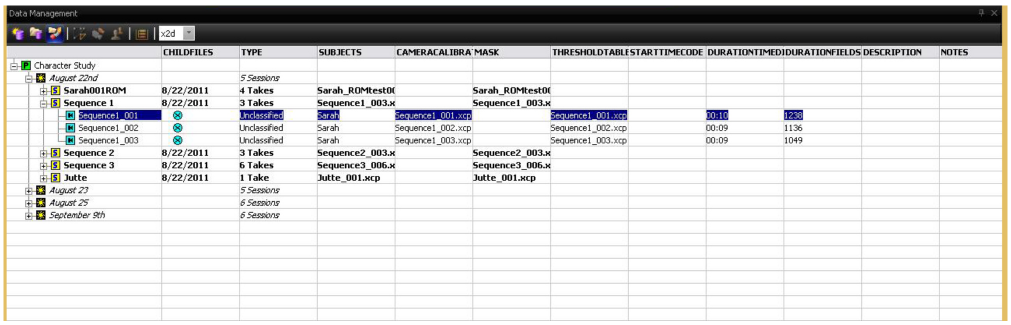


Figure 4.6: **Data Management**. This figure shows the location of all captured data and information about each capture session

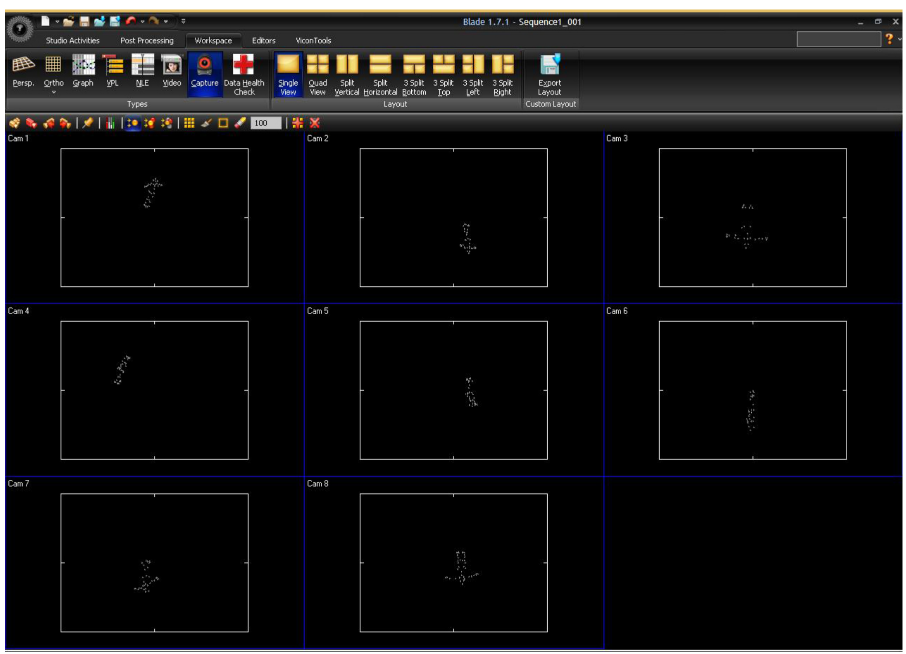
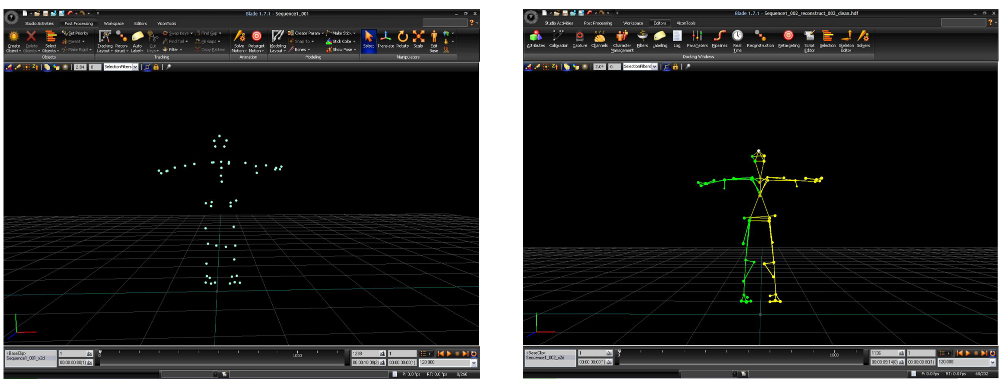


Figure 4.7: **Data Loaded**. This figure illustrates the loaded data from the .x2d file. Each camera's data can be viewed individually at this stage of reconstruction.

If the post processing is done immediately after the calibration the actor's .vsk will be present. If this is not the case and the actor's ROM capture and scene capture occur without calibration then the actor's .vsk will not be present and will have to be imported once calibration is complete. Although this is the method used in our experiment due to time constraints for the actors, it is not suggested and the preferred method is to calibrate the actors immediately after the ROM and then capture the scene data. Once the .vsk is imported post processing can commence.

The first step in the pipeline is to reconstruct the data. Single clicking on the *Reconstruct* button brings up a menu with variables to change and fine tune the reconstruction. For our purposes we want to enable *Grayscale Circle Fitting* and set the *Minimum Number of Cameras* to 2. This maximizes the reconstruction of data and prevents loss of markers; however, it does increase the chance of ghosting. Double clicking on the *Reconstruct* button will assimilate the data into a 3d view as seen in Figure 4.8(a).



(a) Reconstructed (b) Labeled

Figure 4.8: These figures illustrate the process of loading the data and labeling the joint locations.

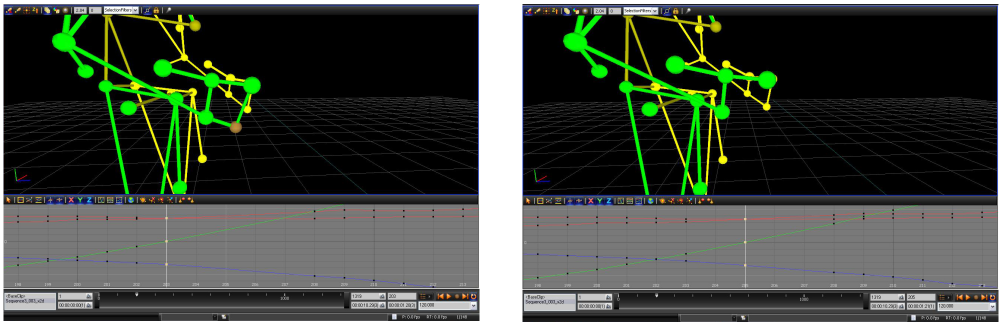
The next step in the pipeline is to label the markers. At this time in the cleaning process, the software does not know what each marker is called in relation to the actual locations on the actor. There are a few options that can be chosen manually, but the one we use most often is the T-pose label. This is the default labeling for human actors and is the basis for how each of our actor's markers are attached. Thus, clicking on the *T-pose Label* button will attempt to label all 53 markers on the participant. This is one of the main reasons for an actor to begin and end each take in a good T-pose. The labeling process will continue throughout the take until any occlusion occurs. If any occlusion does occur the marker will disappear and Blade will not know what the marker is to be labeled if it reappears.

The *Auto Label* button will try to guess based on the approximate location of the marker in relation to where it was before it disappeared. This process is not 100 % accurate so it is best to not depend solely on the auto label process. Marker swapping can occur during this process when the software cannot tell one marker from another. Thus, some manual labeling is sometimes needed in our cleaning sessions. Figure 4.8(b) shows an example of a fully labeled character. Once everything is labeled correctly the main post processing component can commence: cleaning.

Cleaning the data is, in essence, filling any gaps that occur due to occlusion or marker swapping. In our capture sessions, the rate of occlusion is high due to the positions of the motion capture cameras. The cameras are positioned from the ceiling at varying heights. However, the lowest positioned camera is still over eight feet high, which is not the optimal capturing height for a small size studio. This means that all of the cameras are too high, and can contribute to more chances of occlusion, which require more post processing.

Thus, in order to maximize data efficiency and accuracy in future cleanings, it is strongly recommended to move all of the cameras to a lower position and make the cameras more level in relation to each other. As mentioned previously, filling gaps is the primary goal of cleaning the data. A gap is measured in frames and occurs when a marker disappears due to occlusion. The length of the gap varies depending on how long the marker is occluded. Most minor occlusions can be filled with a script that searches and finds any gaps less than 10 frames. These are filled automatically with interpolation algorithms that guess the path the marker is taking between its disappearance and reappearance. Because the cameras are capturing data at at least 120 frames per second A 10 frame gap can be interpolated without causing any noticeable loss of accuracy. Interpolation beyond 10 frames can increase the chances of errors occurring with the gap filling.

For larger gaps a rigid fill algorithm is used. Rigid fills base their calculations on surrounding markers. By using a rigid fill calculation, a gap will be filled based on the trajectory of clean markers that are neighboring the occluded one. For this to be done correctly, at least two markers must be chosen along with the marker intended for rigid filling. Also,



(a) Gap Found (b) Marker Lost

Figure 4.9: These figures illustrate the process of loading the data and labeling the joint locations.

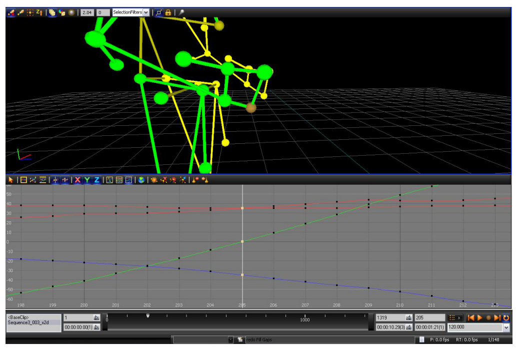


Figure 4.10: **Filled**. The gap has been filled and the marker is fixed.

each marker's relative distance must remain constant. For example, if a hip marker has a gap, it is best to choose the other hip markers for the rigid fill, because they do not move in relation to each other. If an elbow marker is chosen to make the rigid fill, however, an error would occur since the arms can move independently of the hips. An example of filling gaps is shown in Figures 4.9 and 4.10.

Once the data is labeled and all gaps are filled the cleaning process can be completed by solving the .vsk to the marker data. The .vsk is created in the calibration process and is imported when the actor's scene data is loaded. Clicking on the *Solve Motion* button makes the skeletal structure connect to the marker data and completes the post processing. An example of a solved character is shown in Figure 4.11. A solved character can then be exported into a 3d program such as Maya or Motion Builder in order to use as an animated rig for a 3d character. For our purposes we decided to stop here because we felt a 3d model would ultimately interfere with the sex distinction portion of our study and even the base models provided in Motion Builder were inherently male looking, which could cause sex bias in our study.

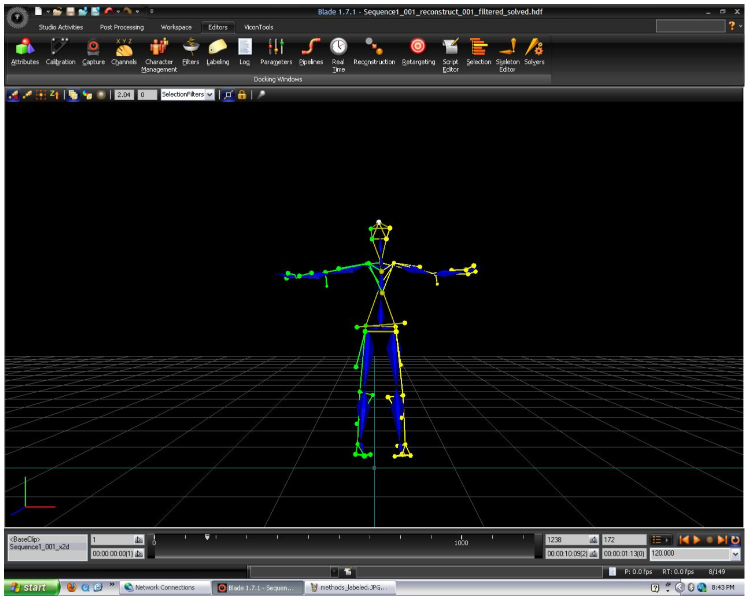


Figure 4.11: **Solved Motion**. This shows the .vsk being attached to the actor and the motion being solved together.

After everything is cleaned and solved, we export the file as a video, providing us with high quality capture data that can be brought into a video editing software program such as Final Cut, in order to composite these videos for viewing in the survey. This will be covered more in the next section: Video Editing.

4.5 Video Editing

For our experiment, we use four actors with each actor performing three different techniques, for a total of twelve distinct videos. We composite each set of actor/technique combinations together, providing us with 144 composited videos of every combination of actor pairs. This pairing allows participants to view the videos chosen randomly and answer questions about any differences they see between each pair of videos. In order to composite the videos together a simple process is completed and repeated in Final Cut. Each video is imported into Final Cut, and each combination is achieved by pasting the videos side by side and then exporting them as one composited video. Thus allowing participants to see two videos at the same time and make comparisons in real time without having to remember what they saw previously. Having the videos paired should also help to minimize the participant's need to guess an answer for any questions asked.

Chapter 5

Empirical Evaluation

5.1 Interface

To view the videos and enable the participants to answer questions, we developed an interface that plays the videos and displays the questions directly beneath the videos. The interface is built in Python, where we exploit its rapid prototyping, and we utilize the Tkinter plugin to handle the questions, and a GStreamer plugin to handle the video playing. The videos did not inherently play on all machines with the GStreamer plugin so the videos were encoded with an H.264 format to enable viewing with Gstreamer. This allows the videos to play on almost any Linux machine since H.264 file formats are more popular and are compatible with most computers. The H.264 file format is also meant for high definition video, and since we want our participants to be able to view the highest video quality possible, this file format is a good choice.

To enable the participants to make a fair comparison of the techniques, we juxtapose the two videos that will be compared on the same canvas, with the questions posed beneath the videos. The videos loop until all questions about the video have been answered; this approach permits the participants to answer each question while the video is playing. To further facilitate a fair comparison, we also provide a pause button so that the participants might periodically pause the videos for a frame by frame comparison. However, it should be noted that although some of the participants took advantage of the pause option, none of them attempted to view the videos on a frame-by-frame basis. Figure 5.1 shows a screenshot of the interface when a participant is taking the survey.

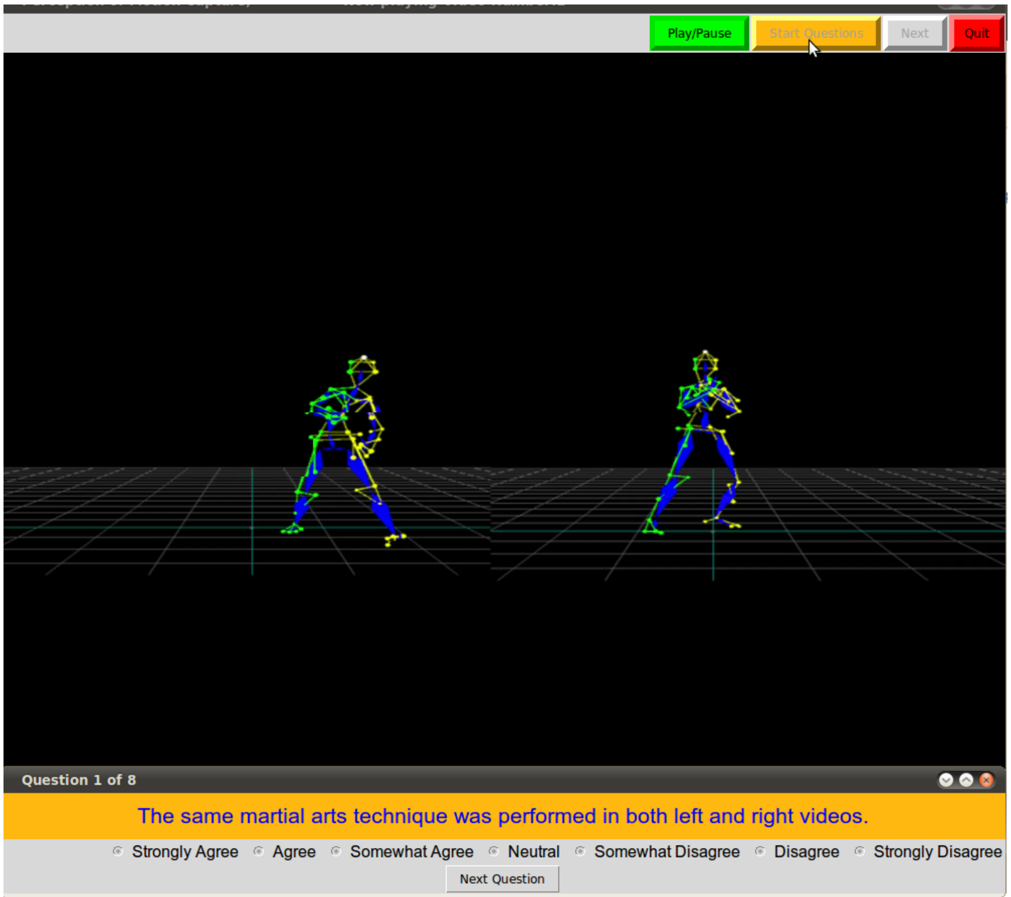


Figure 5.1: **Questions**. Clicking on the Yellow button starts the questions. The same eight questions were asked for each of the twelve videos.

5.2 Survey

Using the interface that we created to loop videos and allow the participants to answer questions about the videos, we set our survey into motion. In order to make this experiment valid we follow guidelines set up by the Institutional Review Board (IRB). These guidelines include taking a training quiz on using human participants for research to show that we are be able to identify all the risks involved in our process. By following these guidelines we are able to guarantee that our participants will be protected and treated fairly.

In order to gather as many participants as possible, we offer an incentive: pizza and sodas. On the day of the experiment we acquire a room in which to allow up to 30 participants take the survey at once. The room we use has over twenty Linux computers and so we are able to upload the survey onto each one.

Each participant is given twelve videos to watch, which are chosen randomly in groups created from the 144 composited video sets (see the *Video Editing* section in Chapter 4.5). The videos are followed by eight questions in which the participant can answer: *Strongly Agree*, *Agree*, *Somewhat Agree*, *Neutral*, *Somewhat Disagree*, *Disagree*, or *Strongly Disagree*. The eight questions are listed in Figure 5.2.

Figure 5.2: **List of Questions**. The Eight questions are shown one after another.

The approach that we use when constructing questions about the actors' actions involves the term *technique*. This term is a martial arts reference for any specific actions made by a performer of the martial art. However, this term *techniqe* proved to be confusing for participants in the study who are not martial artists. A better approach, for future surveys, would be to use the term *movement*, since this term is more familiar to participants who are unfamiliar with martial arts terminology.

Having the participants distinguish the sex of the actors is the secondary goal of this study. We attempt to accomplish this goal by using a Likert scale to determine the sex of the left and right actors in the video. We chose the Likert scale in an attempt to help prevent participants from being forced to guess between two choices of *male* and *female*. In the programming of the interface we ask the participants if the actor is male and then we reduntantly ask if the same actor is female. This line of questioning allows us to program the interface with the same answer possibilities for each question.

A problem arose during the survey in relation to the questions about sex distinction. Using the Likert scale from *Strongly Agree* to *Strongly Disagree*, we asked participants if they thought the actor was male, and then we asked again if the actor was female. We asked both questions in regards to the left actor and also again for the actor on the right side of the video. However, this proved to be tedious for the partici pants to have to answer a question twice. Some participants were noted as being prone to error by attempting to answer the questions too fast. Thus, there are a few answers that have the same response for an actor being male and female (i.e. a participant answered *Agree* for the left actor being male, and *Agree* for the left actor being female). One solution to this redundancy problem would be for the program to only ask if the actor is male/female once and base the answers off of that one question. If a person is asked if the actor is male and the participant answers *Disagree* the program can infer that the participant thinks the actor is female.

The next two chapters describe the results of the survey and address problems that arose during the survey, together with their effects on the outcome.

Chapter 6

Results

In this section we provide results from our study. The survey for each subject is saved in an anonymous XML data sheet that is unique for each participant. There were 48 subjects in the study, so that 48 unique XML data sheets were created. The XML data provides all information necessary to describe the subjects and their responses. The data shows that twenty five percent of the participants were female, and seventy five percent of the participants were male. The data also shows that the majority of participants were in the age range of 22-28 years.

We wrote an XML parser to extract the responses for each question. We analyzed the data using several programs that evaluate and categorize the results. Since the purpose of this experiment relies on distinguishing between two videos, four distinct video groups were created based on the relationship between the video pairs. These groups are: *Same Same* (denoted as *SS*), *Same Different* (denoted as *SD*), *Different Same* (denoted as *DS*), and *Different Different* (denoted as *DD*). These four groups refer to the actors and movement relationships. The *Same Same* group has the same actor performing the same move in both videos. The *Same Different* group has the same actor performing different moves in each video. The *Different Same* group has different actors performing the same moves in each, and the *Different Different* group has different actors performing different moves in each video. These four groups allow us to split up the 144 composited videos and randomize the videos while still allowing each participant to view an even distribution of the different types of paired videos. In this section we will show the results for each question and categorize the data into different groups based on the participant's responses.

6.1 Results About Technique Recognition

In *Question 1* participants are asked if the same move is performed in the left and right videos. The results for this question were as expected. Being independent of the actor the participants were able to agree or disagree accordingly. The graph in Figure 6.1 shows the output from all the participants and is categorized based on the four groups of different videos.

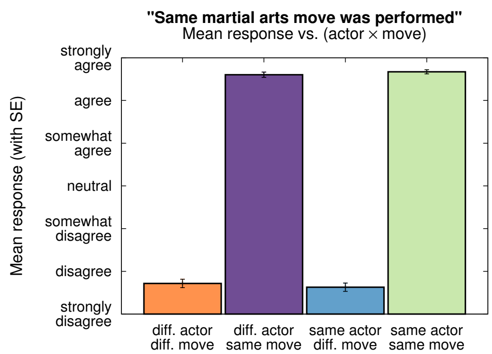


Figure 6.1: **Question 1** These results show that participants were able to distinguish between movements of each technique.

6.2 Results About Actor Recognition

For *Question 2* participants are asked if the moves performed in both videos are done by the same actor. The results from this question were not as high as hypothesized and not as conclusive as the results from *Question 1*. Figure 6.2 shows the output graph for *Question 2*. In the discussion chapter, we will talk about this question further to determine what caused these results to be less than expected, and how can more conclusive results be achieved in future studies.

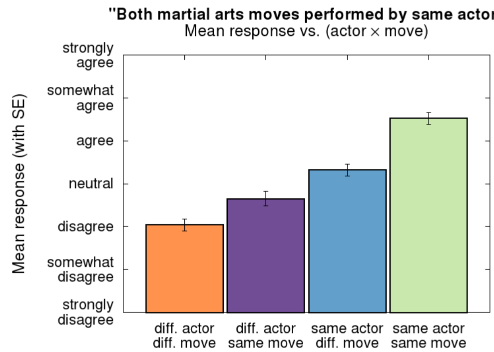


Figure 6.2: **Question 2** These results support our proposed hypothesis.

6.3 Evaluating the Confidence Level of the Participants

The goal of *Questions 3* and *4* is to determine the participant's confidence level of distinguishing sex between the left and right actors. Based on Figures 6.3 and 6.4, the results are similar for both the left and right videos. The data from this figure is inclonclusive. Participants are neither confident nor unconfident in their ability to distinguish sex. We will discuss these two questions in the next chapter to offer an explanation on the outcome of these results.

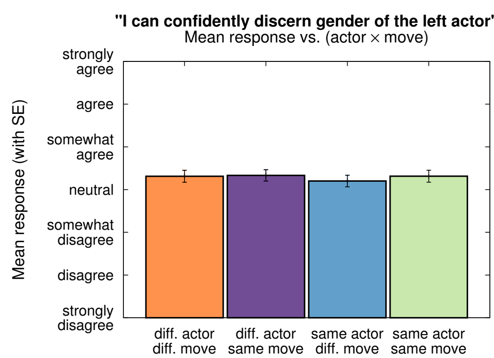


Figure 6.3: **Question 3** These results show the confidence level of the participants trying to ascertain the sex of the left actor.

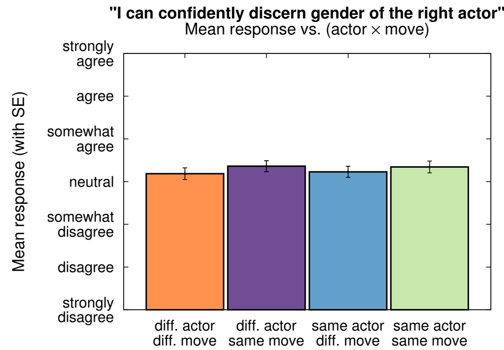


Figure 6.4: **Question 4** These results show the confidence level of the participants trying to ascertain the sex of the right actor.

6.4 Results About Sex Recognition

Since *Questions 5-8* relate to the recognition of the sex of the actors, we discuss them together in this section. In *Questions 5* and *6* participants are asked about the sex of the left actor, where in *Questions 7* and *8* participants are asked about the right actor. In these four questions we ask if the participant agrees if the actor is male, and then we ask if the participant agrees if the actor is female. In the discussion chapter we will note how the wording of these four questions can create an opportunity for error, and we will talk about rewording these questions for future studies. Figures 6.5, and 6.6 only show the output for *Questions 5* and *7* since only two are needed to determine the sex of the actor. The results show that it is challenging to distinguish the sex of each actor in the videos. In the next chapter we look at reasons for why this is difficult and what can be done to improve results for future studies.

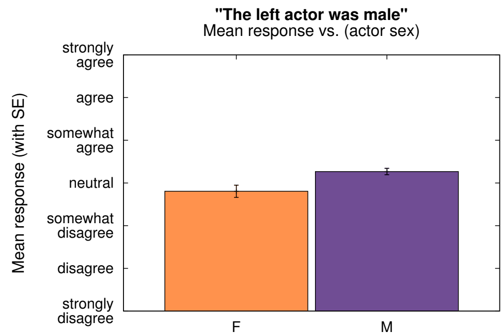


Figure 6.5: **Question 5** These results show how well participants were able to recognize the sex of the left actor.

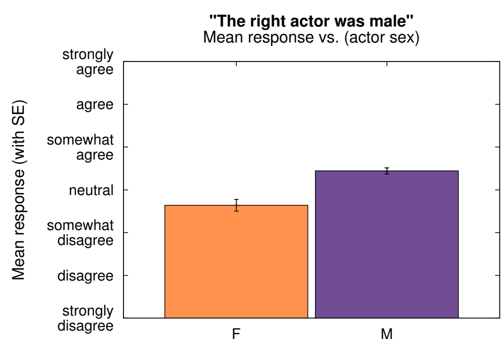


Figure 6.6: **Question 6** These results show how well participants were able to recognize the sex of the right actor.

Chapter 7

Discussion

In this section we discuss the findings from the results section. We address the questions and provide some rationale for the outcomes and suggest improvements for future studies in this area.

7.1 Discussing Technique Recognition

Question 1 is intended to be a control question. From the graph in Figure 6.1 we show that with regards to an actor being independent of another, people are able to tell the difference between the three sets of techniques. They are able to distinguish when the left actor is doing a different technique from the right actor, and they are able to distinguish when both actors were doing the same technique. This first question illustrates that our study is valid since participant can easily distinguish the differences in the techniques.

7.2 Discussing Actor Recognition

The responses for *Question 2* have varied levels, as seen in Figure 6.2. The results show that the participants are better able to recognize an actor when said actor is doing the same move in both video. This shows that even with minor variances between someone repeating an action, participants can still recognize the actor's overall performance as being from the same person. The results for the rest of the question are not as high as we hypothesize. When the same actor performs a different move in each video, participants are not as confident. The graph shows that the mean response is close to being neutral, indicating that overall it is more difficult to see the differences in an actor when the moves are different. However, we can still consider this valid data since the mean responses are above neutral and close to the *Somewhat Agree* level. When different actors are performing the same technique the mean response is also close to the neutral answer. However, since it is below the neutral line and close to the *Somewhat disagree* line, we can see that the results are going in the general direction of the proposed hypothesis. These two parts of *Question 2* should be studied more in the future. Problems that arose in earlier sections of this paper may effect the results, especially for this question, and now that these problems have been recognized, most of them can be fixed in any future studies. In regard to the wording of the question, perhaps an explanation at the beginning of the study might indicate that only one style of martial arts is used, so all techniques would derive from that style.

7.3 Discussing Confidence Levels of Sex Recognition

Questions 3 and *4* show the participants were neutral in their confidence levels. We can offer conjecture that the results from these two questions could have occured for two reasons. One reason being that particpants who were unsure or uncofident chose the *Neutral* answer because they were not confident. The other reason for the neutral outcome could be that there is an even spread of confident and non-confident participants. Of these two conjectures the former is the most likely due to the answer *Neutral* having similarity to the phrase *I don't know*, thus displaying a low level of confidence. These two questions may have been more detrimental to the study than helpful. Future studies may revise *Questions 3* and *4*, or remove them and base confidence levels on the answers for the questions directly addressing sex distinction.

7.4 Discussing Sex Recognition

The results for *Questions 5-8* are combined into 2 graphs. Figure 6.5 depicts the responses for the sex of the left actor and Figure 6.6 depicts the responses for the sex of the right actor. Since *Question 5* asks if the left actor is male, and *Question 6* asks if the left actor is female, some confusion may have occured in the participants. To eliminate redundancy in the graphs, we decided to only use one question from each pair. The mean values are marginally representative of participants being able to distinguish sex between the actors. The mean values for both graphs show that for both sexes the graphs are above or below their respective neutral areas. This might lead us to conclude that participants may be able to distinguish sex through complex motion captured movements like martial arts techniques. However, because of the low margin it is not conclusive, so more research should be done in the future. One solution to help future studies might be to reevaluate the last four questions so that only two questions are needed, and they will be less confusing and allow less room for error. Another fix might be to reevalutate the Likert scale and determine if it is needed for every set of questions.

Chapter 8

Conclusion

In this thesis we investigated motion capture and some of the different types of motion capture currently available. We used the passive marker system that utilizes infrared lights to capture the movements of an actor wearing small reflective dots or orbs on a velcro suit. The type of martial arts used for this thesis, called Cuong Nhu, is a varied style of martial arts incorporating moves from other styles. This martial arts style allowed the movements and techniques for the experiment to look different from one another without being mistakenly identified as a different style. The primary goal of this study was to determine if participants watching videos of motion captured actors could distinguish two actors doing the same move. Our results indicate that participants cannot easily distinguish two different actors based on their movements; however, participants are able to distinguish the same actor performing the same move. The evidence in the results propound that future studies should be done involving motion capture with more complex movements, such as martial arts, in order to further validate this study. Distinguishing the sex of the actors was a secondary goal of this thesis. Determining the sex was more difficult than general actor distinction and the data shows marginally that people might be able to tell the sex of an actor; however, due to certain problems that arose during this study, future studies related to this thesis must address the problem of question organization and delivery, and make sure that all terminology is understood by all participants. This study and any future studies related to this one should help validate the use of motion capture in video games that use movements particular to their characters; games such as *Jet Li Rise to Honor* and the *MLB the Show* series baseball games are examples of these types of video games.

Bibliography

[1] Point light display image, October 2011. `http://www.differenceblog.com/2008/09/point-light-figures.html`.

[2] Vicon Blade, 2012. `http://www.vicon.com/products/blade.html`.

[3] Cuong nhu oriental martial arts, 1965-Present. `http://www.cuongnhu.com`.

[4] Motion capture world records. `http://www.motioncapturesociety.com/resources/world-records`.

[5] Jet li: Rise to honor. `http://jetli.com/jet/index.php?l=en&s=work&ss=articles&p=risetohonor`.

[6] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Attention, Perception, and Psychophysics*, 14:201–211, 1973. `http://dx.doi.org/10.3758/BF03212378`.

[7] B.A. King and L.D. Paulson. Motion capture moves into new realms. *Computer*, 40(9):13 –16, sept. 2007.

[8] Jacquie Kubin. Merging trends in the world of gaming, October 2003. `http://www.awn.com/articles/gaming/merging-trends-world-gaming`.

[9] Fani Loula, Sapna Prasad, Kent Harber, and Maggie Shiffrar. Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1):210–220, Feb 2005.

[10] Rachel McDonnell, Sophie Jörg, Jessica K. Hodgins, Fiona Newell, and Carol O'Sullivan. Virtual shapers & movers: form and motion affect sex perception. In *Proceedings of the 4th symposium on Applied perception in graphics and visualization*, APGV '07, pages 7–10, New York, NY, USA, 2007. ACM.

[11] Meredith McLendon, Ann McNamara, Tim McLaughlin, and Ravindra Dwivedi. Using eye tracking to investigate important cues for representative creature motion. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 85–88, New York, NY, USA, 2010. ACM.

[12] A. Menache. *Understanding Motion Capture for Computer Animation and Video Games*. Morgan Kaufmann, 2000.

[13] John A. Pyles, Javier O. Garcia, Donald D. Hoffman, and Emily D. Grossman. Visual perception and neural correlates of novel 'biological motion'. *Vision Research*, 47:2786–2797, 2007.

[14] Mike Snider. Minnesota Twins Catcher Joe Mauer.