Digital Dojo: *A Serious Gaming Approach for Supplementing Martial Arts Education Using Kinect Motion Capture Analysis*

A Thesis

Submitted to the Faculty

of

Drexel University

by

Michael John Dixon

in partial fulfillment of the

requirements for the degree

of

Master of Science in Digital Media

June 2012

Dedications

This thesis work is dedicated to my mother and father, Sue and Mark Dixon, who always taught me to finish what I start.

Acknowledgements

I extend a great deal of gratitude and credit to my thesis advisor, Jervis Thompson. His continuous advice kept this thesis work precise and focused. Through his council, I was able to scope my time and workload effectively and reach goals consistently.

I would also like to express my thanks to my committee members, Troy Finamore and Dr. Will Dampier. Their suggestions, critiques, and demands pushed me to produce a high quality project that surpassed my own expectations.

My fellow students, Glenn Winters and Matt Haas, have been incredibly supportive in developing a reliable motion capture workflow at Drexel and deserve my sincere thanks. Without their continued interest and eagerness to experiment, I would surely not have been able to realize my vision to its fullest potential.

I thank Dr. Will Dampier once again, for allowing me to utilize his talent and record his motions for the Digital Dojo application.

I give my dearest thanks to Christine Adams for her love, support, and patience during the past year of intense production and late hours.

Finally, an enormous debt of gratitude and respect goes to my Kempo Karate instructor of nearly 10 years, Frank Manzelli. Frank pushed me mentally and physically to become the best athlete, friend, brother, son, and person I could be. He was instrumental in instilling me with a tremendous work ethic, self-discipline, and an indomitable spirit. He taught me about the importance of teaching others, and is ultimately the inspiration for this thesis work.

Table of Contents

LIST OF FIGURES

ABSTRACT

Digital Dojo: A Serious Gaming Approach for Supplementing Martial Arts

Education Using Kinect Motion Capture Analysis

Michael John Dixon

The potential for learning through serious video games is something that has been examined closely in recent years. Teaching students through well-designed games has several potential benefits, such as immersing them in the subject material, boosting interest, comprehension, and motivation, and developing strong problem solving skills. While serious games have been heavily researched and applied to subjects such as economics and mathematics, physical tasks have traditionally been difficult to teach through computers for a number of reasons. Currently, there are no computer applications that supplement martial arts training by analyzing and critiquing students without the use of expensive and cumbersome peripherals. With the advent of compact, affordable natural human input devices such as the Microsoft Kinect, the user's real-world body position can now be accessed and analyzed in real-time. Using current education theory, I have developed a serious game that utilizes the motion-tracking capabilities of the Microsoft Kinect device to supplement a traditional Judo education.

I. INTRODUCTION & THESIS OVERVIEW

While methods for serious gaming are rapidly evolving, few, if any, effectively deal with the instruction and training for physical tasks, specifically martial arts. The term *serious game* refers to a game that's primary function is not entertainment, but to deliver a message or relay information or skills to the player. This type of learning has been an expanding area of study over the past decade, and has many potential benefits for students [1]. Popular subjects for serious gaming include mathematics, and physics [2].

Due to the current limitations of computer hardware, bandwidth, artificial intelligence, and human-computer input devices, martial arts is a difficult subject matter to communicate through a computer. For the purpose of narrowing the focus of this research, I have identified three main challenges that make teaching and learning physical activities and skills via a computer interface such a challenge:

1. *It is difficult to demonstrate physical actions using standard 2-D media.* Studying the way a martial artist performs a technique in a video, animation, or diagram only provides a fraction of the content. One reason for this is that these forms of media are often created from a singular vantage point. As a result, the pieces of the technique opposite the camera become occluded, making it difficult for the student to discern the full-body intricacies of the technique. One common issue with video in particular is poor resolution that makes it impossible to interpret the details of the technique [3].

---

1 T. I. Perttu Hamalainen, Johanna Hoysniemi, Mikko Lindholm, Ari Nykanen, "Martial Arts in Artificial Reality," presented at the Enhancing Virtual Spaces and Large Displays, Portland, Oregon, 2005.

2. *Independent online learning of physical behavior is not structured and has no defined progression.* Current online martial arts instruction provides students with no clear starting point or goals. Within martial arts in particular, this could lead to dangerous situations where naive or over-ambitious individuals attempt maneuvers above their competency level. This unstructured, self-directed learning can also lead to loss of interest in the material and demotivation of the student. Without defined goals or progression, it is difficult for students to chart their progress and stay motivated [4].

3. *Previously, there has been no affordable, unobtrusive input method for gathering and analyzing real-world motion of the student.* Using an online computer game to teach martial arts, it is difficult for an instructor to evaluate a student's progress and provide prompt feedback. Using current Web 2.0 techniques, the student would have to capture still images or video of their own actions and send them to the instructor for analysis; he or she would then need to view them and reply with feedback. In the process, the ability of the students to capture themselves using photographs or video could have skewed the accuracy of their true movements. That movement would then be interpreted by the instructor, and then analyzed against the feedback by the student. This is an unacceptable amount of distortion and subjective interpretation for effective teaching or learning to take place. Furthermore, even with above-average Internet speeds, this method of feedback could never hope to achieve the immediacy of a real classroom environment.

Current online martial arts E-Learning is a one-sided process; the students are expected to absorb information through pictures, text, audio, or video, but have no way to demonstrate that they have truly learned what is being taught. With the development of affordable, unobtrusive hardware for gathering human motion data, such as the Microsoft Kinect, serious games may begin to support the instruction of real-world physical skills, specifically martial arts. The Microsoft Kinect device is a piece of hardware that contains a camera, depth sensor, and an infrared light projector. The camera collects RBG color data, similar to a standard web camera. The depth sensor interprets the infrared light thrown from the projector. This infrared light data allows computer programs to collect the user's body position and movements.

A brief sample of former strategies for teaching students to perform physical actions include a martial arts trainer comprised of a full-room installation [5] and a software for learning Tai Chi that relies on a motion capture suit and head-mounted display (HMD) [6] . These solutions are expensive, cumbersome, and are not informed by current education theory. Rather, the martial arts training application encourages the user to free-style and invent their own techniques, while the Tai Chi program provides little more than a copy-cat mechanic where the student is shown a movement and attempts to replicate it without feedback.

Game-based learning provides important potential opportunities to supplemental martial arts education. For instance, automated sessions with dynamic feedback provide students with personalized assistance and 24-hour-a-day access to material. This helps students retain information and stay focused when they are not under the supervision of an instructor. As a supplementary tool, a game-based learning application could provide an incredible aid to students who are simultaneously pursuing martial arts in a traditional class.

The martial arts techniques used in this study were limited to beginner-level Judo maneuvers. The relatively recent martial arts style of Judo was chosen for several reasons. First, at its roots, Judo has traditional ties to education and problem solving which connect it thematically to situated learning and serious gaming. Secondly, Dr. Will Dampier, a member of my thesis committee, has extensive experience and expertise in the art of Judo; he was the point of contact for supplying the curriculum, and approving the accuracy of each technique. Finally, the compact nature of Judo techniques forced the user's body into a variety of extreme positions that helped to define the technical limitations of the system. The lessons were also limited to only one student user at a time. Although, training multiple students using the same Kinect device is theoretically plausible, it was far beyond the scope of this study.

1. Definition of Terms

Aim Constraint

Constrains an [digital] object's orientation so that the object aims at other objects, as defined by the Autodesk Maya 2011 help documentation.

Autodesk Maya 2012

3D graphics application. In the case of Digital Dojo, 3D modeling, texturing, and some correctional processes were handled within Maya.

Autodesk MotionBuilder 2012

3D graphics application specifically designed for manipulating motion capture animation.

Dojo

In this study, the Japanese word *dojo* will refer to the physical space where martial artists study.

E-Learning

Any computer or internet-based model for learning information or skills.

Game Object

Base class for all entities in Unity scenes, as defined by Unity 3D scripting reference. It is the term for any number of different objects within Unity including geometry, lights, cameras, and particle systems. Game objects can hold scripts and often have many editable components.

Intrinsic Value

Gratification that comes with performing a certain behavior for individual satisfaction. Exercising skills, sharing with others and bettering one's self are examples [4].

Microsoft Kinect device

A piece of "natural interaction" hardware developed by Microsoft, which consists of a camera and depth sensor which perceive RBG color values and near-infrared light respectively. It can be used as an input device for human body position and has a USB plug that connects to all modern laptops and desktop computers.

Motion Capture

The process of collecting the movement of a live actor for the purpose of applying that animation data to a digital character.

Orient Constraint

An influence on a driven [digital] object's rotation, matching it to a driver object's rotation.

Pipeline

The system by which a process is completed. The pipeline includes every step, from beginning to end.

Situated Learning

A theory for learning where students learn by becoming immersed in the culture of the subject matter.

Tai Chi

A style of martial arts that is performed very slowly with emphasis on focus, minute details, and balance.

Vicon Blade

A software package that is used to collect and polish optical motion capture data.

Zigfu Plugin

A plugin that allows Unity 3D to interpret data from the Kinect using C#.

II. REVIEW OF RELEVANT RESEARCH & THEORY

1. History of Judo Education

Judo is a form of martial arts that highly emphasizes health, education as well as physical and mental technique; it was first taught by a man named Kano Jigoro from Japan in the 1890's [8]. By the start of the twentieth century, it was a well-established style of competitive martial arts. In fact, the Japanese government made it a part of the mandatory curriculum in schools and universities. After WWII, bans were put on Japan concerning the militarization of the people; this ultimately caused Judo to become more emphasized as a sport than a method of self-defense. The commercialization of the sport changed lots of the original traditions, such as switching the gi (traditional Judo uniform) color from white to blue. The sport even caused some of the original techniques to change. However, despite the numerous iterations that Judo has gone through, its founding ideals of maintaining strong mental health and education along with physical conditioning has persisted [8].

2. Education Theory

With the start of the modern education system in the United States, knowledge and the learning process have been thought of as two separate entities [9]. However, the idea of situated learning has begun to gain in popularity. Situated learning is a method by which the students immerse themselves in the culture of the particular subject matter. Students involved in situated learning perform what are called authentic activities. Brown defines authentic activities as the ordinary practices of a specific sub culture [9]. Miller and Gildea describe two stages of situated learning: the first as a quick acceptance of a concept, the second being the further understanding achieved only after repeated usage and thought.

In the chapter “Language and Identity at home” of his book, Situated Language and Learning, James Paul Gee argues that games have great potential for to evoking situated learning. He claims that students will not only learn the techniques, but also the traditions and respect [10].

“Educational Technology: Media for Inquiry, Communication, Construction, and Expression” is a paper that seeks to categorize educational media can be categorized based on John Dewey’s model of impulses: Inquiry, Communication, Construction, and Expression [11]. Using this method for identifying educational media as a guideline, I can be sure that Digital Dojo will serve its purpose as a serious game.

In “From Content to Context” Squire suggests that games are shifting us towards the culture of simulation. ”Participants learn through the grammar of doing and being [12].” The challenge with serious gaming is making sure that the user actually learns something. Some focus points include challenge, curiosity, fantasy, interactive narratives, collaborative problem solving, and game players as producers. Players must be made aware of what they can do. Players set up projective identities within the game world.

Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation [4], by Matthew Lee, et al., is a look at how intrinsic value is now being looked at as the motivator for online education. The paper lays out guidelines for creating Internet Learning Medium (ILM). Lee defines the critical aspects of Internet Learning Medium as having varied content, being fun to use, providing immediate feedback, and encouraging user-interaction.

In *Here Comes Everybody: The Power of Organizing without Organizations [13],* Clay Shirky explains how peoples’ inherent desires to share and communicate are shaping modern technology. Rather than espousing technological deterministic causes for social change, he examines how social changes drive technological advancement. Shirky analyses why people are willing to expend their time, effort and skills online without financial compensation. He attributes participation in online social activities in part to intrinsic value. Intrinsic value refers to the personal motivations that people inherently have to do work, such as bettering themselves or exercising their skills. Digital Dojo will provide intrinsic value to potential students by offering them an opportunity to learn new skills, improve their physical condition, and boost their confidence by learning self-defense techniques. These personal benefits will motivate people to dedicate time to the site, learning and practicing the material. In accordance with Shirky's assertions, the content of the site will need to be managed, so that students are not overwhelmed by sub-par or redundant lessons. To accomplish this type of filtering, the users will be charged with the responsibility of monitoring the content of the site by using the website's built-in rating system.

3. Case Studies: E-Learning used for Physical Tasks

3.1 E-Learning Martial Arts

E-Learning Martial Arts [14], a paper by Taku Komura, et al., outlines the potential for motion capture technology as an aid to learning martial arts. The study identifies several important downfalls of traditional martial arts education. The first issue is the large class sizes; with one instructor over-seeing a large class of students, it is likely that each individual will receive little personal attention for feedback and critiques. Komura argues that "the students can still interpret the advice in a wrong way, and again repeat the wrong motions until the coach comes to see them the next time." [14] The lack of one-on-one attention in traditional martial arts settings may discourage some students from participating and cause others to fall behind. Personal issues and scheduling conflicts are also offered as explanations why people cannot commit to classes. These students are left with the option of watching instructional videos and reading books. However, without the real-time critiques and suggestions of an instructor, it is difficult to master technique. The setup that Komura uses includes an optical motion capture system with 7 cameras and a HMD. A virtual coach, who is seen through the HMD, evaluates the student on several aspects of martial arts: Total movement for defense, minimum pre-action for attack, speed, feints, and combinations. While the study came up with positive results pertaining to the effectiveness of motion capture in martial arts training, there are some glaring limitations. For example, their methodology requires the use of an optical motion capture system, as well as a specific outfit and the head-mounted hardware. By using the Microsoft Kinect, I will eliminate the need for any of the wearable peripherals and drastically lower the cost of the system. Additionally, I will be using a different set of criteria, based on effective evaluation of martial arts *techniques*, rather than various aspects of combat.

3.2 Interactive Video Mirrors

Interactive Video Mirrors for Sports Training [15] the researchers experiment with different ways of hands-free control for their video mirror device. The first being automatic play back, with no true interaction. It used a simple pixel-tracking algorithm and was helpful, but limiting, also using spoken commands with mixed results. The third technique was a floating graphical user interface (GUI) that could optionally follow the user. The researchers tested the three systems on martial artists with different levels of experience. The audio scored the best, while the automatic method scored the worst. Ceasing movement in an unguarded standing pose is actually a fairly uncommon stance in martial arts. The over-lay menu was more intuitive than speech. The tracking buttons were preferred over the static ones.

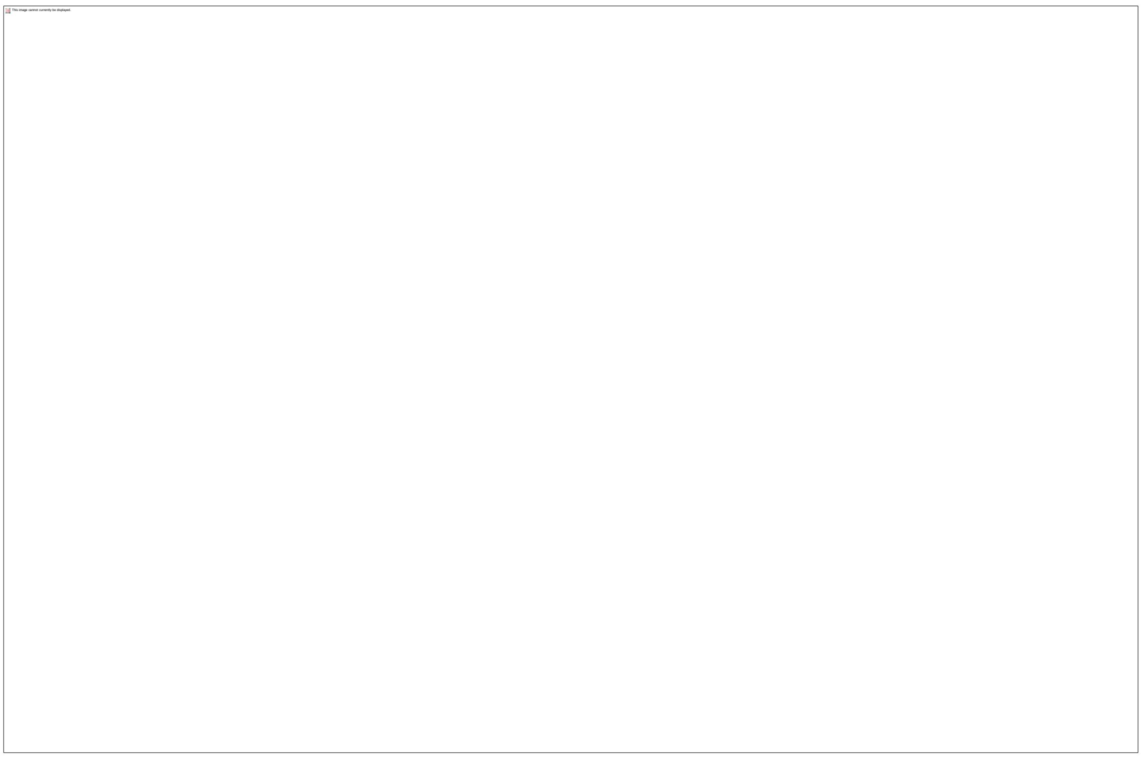

Figure 1. Hamalainen's setup for capturing and analyzing video for martial arts training.

Martial Arts in Artificial Reality [5] is a paper by Perttu Hamalainen et al. that describes their work using computer vision software and two large projection screens to create an artificial reality martial arts training experience (Figure 1 [5]). For their application, the team used computer vision technology to remove the area behind the user and composite their body into a virtual environment on each screen. The user was able to interact with the program by fighting virtual enemies with their body without any additional hardware or gear, such as motion capture cameras, HMDs, or battery packs. The virtual environment also had the ability to warp time and exaggerate actions, allowing the user to appear in slow motion and jump over NPCs. However, although the team stresses the potential of their system for serious martial arts training, it currently lacks any sort of structured material or progression for the student. Furthermore, the system is enormous and not practical for home usage. I hope to benefit, as they did, from allowing students to participate without external peripherals, but my system will use the Microsoft Kinect device to reduce the scale of the setup dramatically.

3.3 Training for Physical Tasks in Virtual Environments

Training for Physical Tasks in Virtual Environments: Tai Chi [6], a paper reporting on a study done by Philo Tan Chua et al., explored some of the possibilities for representing the student and instructor in a virtual environment, specifically pertaining to the instruction of Tai Chi.

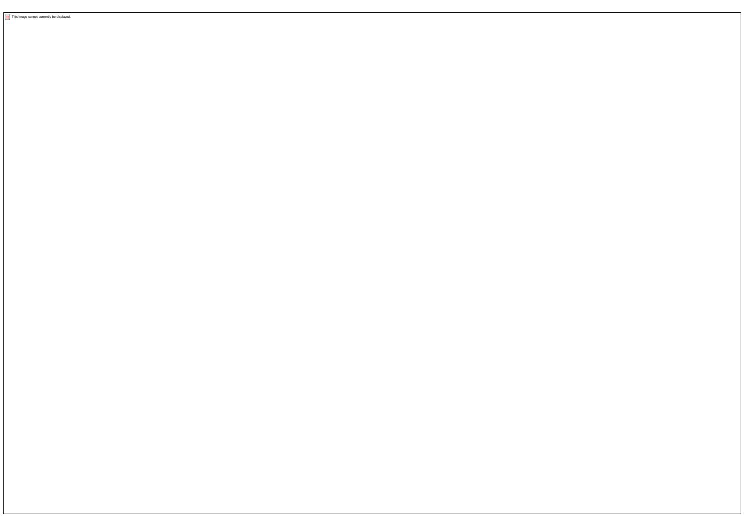

Figure 2. The user must wear a head mounted display and motion capture suit.

While the system they developed was cumbersome and impractical for the average person (shown in Figure 2. [6], the process required a body suit, 12 motion capture cameras and a HMD), they did have some interesting findings relating to representing avatars in virtual space. There system was tested on 40 volunteers with various combinations of positioning and duplicating both the instructor and the student (shown in Figure 3. [6]). Their results concluded that the traditional one-on-one approach to teaching martial arts was just as effective as some of the seemingly more comprehensive, view-every-angle type setups. [6] This will have an impact on my work, as I will implement the one-on-one approach with confidence that I will not be missing an opportunity to enhance the experience.

Figure 3. Different configurations for representing the student and the instructor in a virtual space.

3.4 A Serious Computer Game to Assist Tai Chi Training for the Elderly

Recognizing the health benefits of martial arts, Wagner O. de Morais and Nicholas Wickström developed an E-Learning system for teaching the martial art to elderly users. In order to track the user's position, they used small, wireless motion trackers attached around the wrists and ankles. [16] In this example, the software tracked the students' silhouettes, without considering depth. The data was broken into critical chunks before analysis, which helped their software to decide whether or not the player has done certain key poses.

4. Case Studies: Mainstream Kinect Martial Arts Games

4.1 Ubisoft's Self Defense Training Camp

Self Defense Training Camp [17] was released on November eighth, 2011, making it one of the earliest Xbox Kinect games dealing with self-defense on the market. It was received with terrible reviews from critics and users alike; however its similarity to Digital Dojo makes it worth examining, in order to define which aspects of this educational experience makes it so ineffective. Below, I have compiled my personal observations of this learning game and organized them into areas of focus that align with challenges faced in designing Digital Dojo.

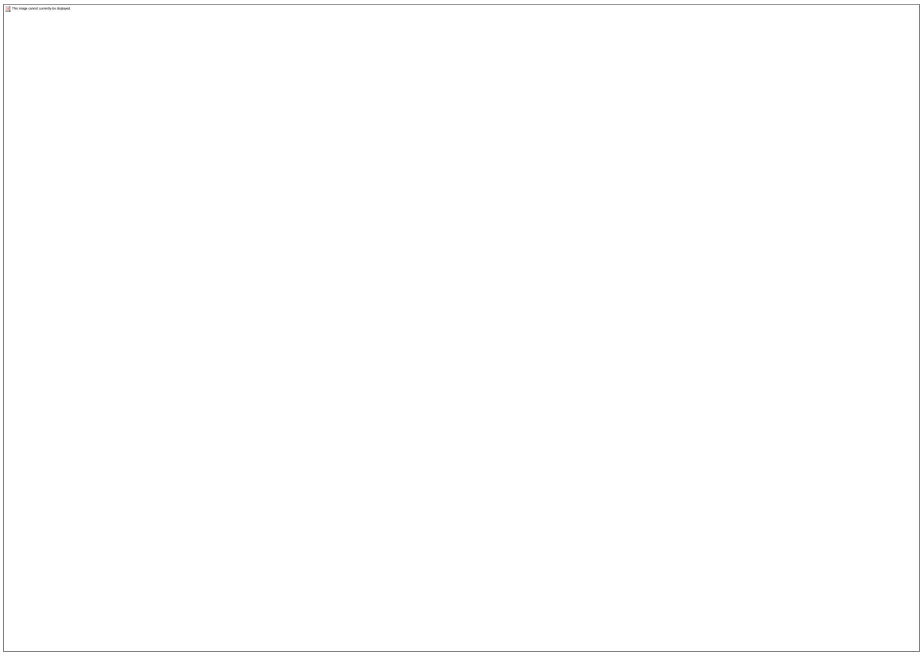

Figure 4. ‘Swipe and punch’ based navigation menu. [Screen shot]

Navigation

The navigation system for Self Defense Training Camp is based upon two user actions: swipes and punches. The ‘swipe’ command is triggered when the user passes either their left or right hand in front of him or her from right to left or left to right. While the movement feels natural to perform, it rarely works as expected. The functionality of the swipe breaks if the user is standing at certain distances from the Kinect. It is completely unreliable and frustrating.

Avatars

The player’s virtual avatar in Self Defense Training Camp is female by default, and the assailant is male (see Figure 5 [18]). Also, textual advice during loading screens lends tips about avoiding purse snatching. While this may adhere to the demographic that the game was marketed to, it reduces the usefulness of the game as an educational experience. [2]

Figure 5. The default player avatar is a female, pulling male participates out of the game world.

Student Evaluation

The game evaluates the user during two distinct phases. First, the user must perform actions step by step along with verbal commands from the trainer. Next, the player is told to perform the move on his or her own at 'full speed.' I found that the system held the player to almost no standards; flailing one's arms and legs arbitrarily in front of the Kinect receives an "excellent" rating. During the second phase of user-analysis where the player is prompted to practice the move at 'full speed,' the game cannot keep up with. Even performing the technique half as quickly as 'full speed', the system could be easily left behind.

Motivation & Progression

One of Self Defense Training Camp's strengths was its progression system. From the title menu, I was allowed to see several game play modes that were disabled until I had accomplished certain goals. Also, within one of the sub-menus, there were many defense techniques that were disabled until I had completed the previous ones. This is valuable for two main reasons: motivation and safety.

4.2 UFC Personal Trainer

THQ's UFC Personal Trainer [19] is more focused on leading the user in exercise than teaching martial arts techniques; however, it shares many similarities with Digital Dojo in its approach to instruction and is worth analysis. As before, I have categorized my observations to correlate with issues that are dealt with as well in Digital Dojo.

Navigation

Much like a traditional computer game, UFC Personal Trainer's navigation system revolves around a cursor. The position of the cursor is determined by the user's outstretched arm. Selecting menu buttons is done by hovering over them for a specific amount of time. While this system is intuitive to the average computer user, it is often jittering and unreliable, especially near the edges of the screen.

User Feedback

One positive aspect of the UFC Personal Trainer was all of the audio cues used to enhance the user feedback both in the menus and during gameplay. The sound adds an additional piece of sensory confirmation when the user performs a task. Also, this allows for the game to signal to the user, even if he or she is in a position where they cannot see the screen, which can often be the case in martial arts.

Finally, one thing that lacked within the training program was a virtual representation of the player. On the side, as part of the onscreen heads up display (HUD), there was a flat cluster of pixels that displayed the player's silhouette. However, there was no avatar for the player to relate to within the lesson.

III. APPROACH

Digital Dojo was produced in four major phases, which are outlined below. The first piece of the process was to gather motion data that would represent the 'master' technique (the technique that the student would try to emulate). Next, was to develop a stand-alone application that would receive user input and compare it to the master data. Third, I created a lesson based on current education theory to implement in the program.

1. Research: Selecting the ideal Kinect platform

1.1 Research Overview

Finding the ideal platform for this project was a crucial part of the process. I was looking for a program that could process input from the Microsoft Kinect reliably and accurately. It needed to have 3D graphics capabilities, and be easily scriptable and customizable. Furthermore, it should be simple for the end user to install and operate. I analyzed various possible solutions (described below) before determining that Unity 3D, in conjunction with the Zigfu plugin would be the most robust solution.

1.2 Brekel Kinect

Brekel Kinect is a program that was release for free by a software developer named Jasper Brekelmans [20]. The software combines Open NI, and NITE Middleware to achieve 3D skeletal tracking of multiple users. This skeletal animation can be recorded within Brekel Kinect and exported as a .C3D or .bvh file. It also has the ability to create geometry and acquire photographic textures using a combination of the RGB camera and depth sensor built into the Kinect device. Brekel Kinect has several promising features and, despite not being selected, still stands to benefit Digital Dojo in the future (see RECOMMENDATIONS & FUTURE APPLICATIONS section).

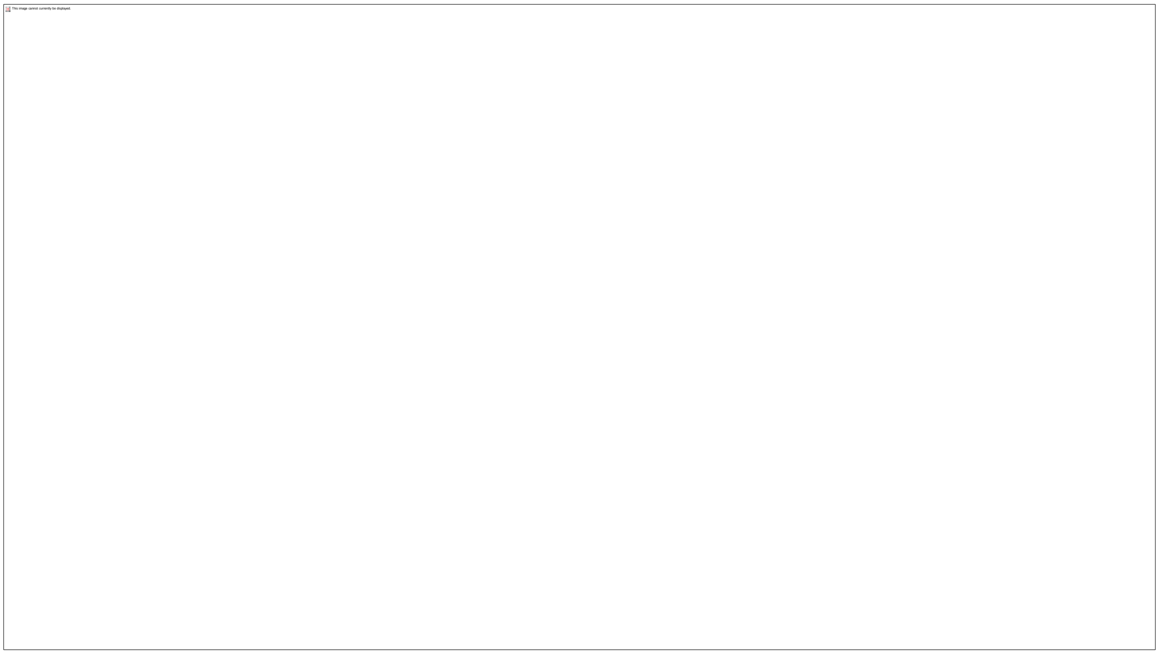

Figure 6. Brekel Kinect

The biggest disadvantage of Brekel Kinect is that it is not customizable. I could not add in my own buttons, scoring system, character models, or pre-recorded animation.

1.3 Processing

Processing is a popular programing language and application that is can be exported and run on the web using Java. Open source scripts allow interaction with the Kinect for both Mac and PC. However, Processing scripts do not currently support skeletal tracking. Also, while

Processing does have some 3D graphics ability, it does not support imported 3D animation data, such as .fbx.

1.4 Microsoft SDK

It's worth mentioning that I did, of course, look into Microsoft's official software development kit for Kinect. However, it was ruled out, largely based on the timing of its release. By the time the Microsoft SDK was released in the summer of 2011, I had already interfaced the Kinect with Unity 3D using Amir Hirsch's 'Tinkerer' wrapper.

1.5 Unity 3D with Zigfu wrapper

Unity 3D was ultimately chosen as the platform for developing Digital Dojo. Unity 3D is a popular, free gaming engine that is approachable for designers. It features full support for .fbx files from Maya, and maintains character animation and skinned geometry. Unity 3D also has built-in blending features for animation that help polish the look of the animation. In addition, there is a huge community of users, which aids greatly in learning and troubleshooting. Unity 3D has the ability to communicate with the web using Javascript. In combination with the free Zigfu plugin (formerly the Tinkerer Unity 3D wrapper) Unity 3D has a multitude of built-in abilities to handle Kinect input [21]. Zigfu supports skeletal tracking for one or two players, and the latest version is very stable. Lastly, the future developments from Zigfu show enormous potential for expansion of Digital Dojo, and even in-browser play online.

2. Development of an intuitive workflow for capturing and importing usable, 3D motion data into Unity 3D.

2.1 Motion Capture Pipeline Overview

To insure that the master techniques are generated to the highest standard possible, the instructor avatar's animation was recorded using an optical motion capture system. Each move was captured multiple times and the best takes were selected. These animation clips were then stitched together and imported into Unity 3D. Small modifications were made to insure that the animation's accuracy was held as the data was transferred across different software. The entire process is detailed in the following sections.

2.2 Capturing the Live Motion Performance

Although the Kinect is the motion capture device that will be used within Digital Dojo at run-time, I used the in-house Vicon optical motion capture system at Drexel University to insure the most accurate results. The system includes 16 cameras, 14 Vicon MX40's and 2 Vicon MX. Our volume for capturing usable animation data is approximately 9ft by 12ft.

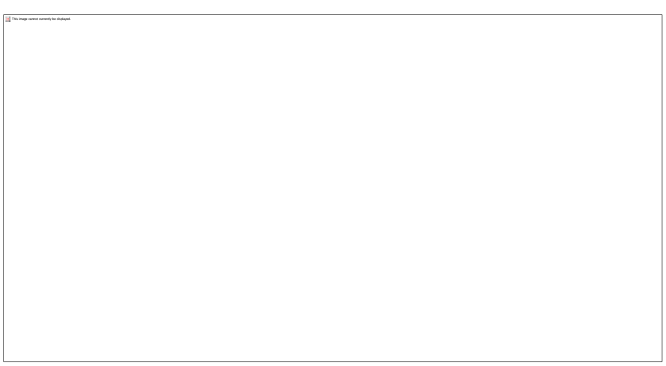

Figure 7. The optical motion capture stage at Drexel University.

To assist on the day of shooting, Glenn Winters, a graduate student of digital media at Drexel University, was in attendance to operate the Vicon Blade software. His research of the motion capture pipeline has been invaluable for my own knowledge and has dramatically eased my workflow. Dr. Will Dampier and Molly S. lent their skills in the art of Judo as the actors for the shoot.

Our first task was to calibrate the camera to ensure the best results possible; this involves three steps. The first is an automated process referred to as Create Threshold, during which the Vicon Blade software masks out any reflective surfaces within the room that are being detected by the cameras. Next, we run a calibration process known as Wave Wand. This step involves entering the capture space with a reflective wand and making circular motions in front of each of the 16 cameras, enabling the software to calculate their exact positions in relationship to one another. Lastly, we placed a small apparatus with reflectors in the center of the room that allowed the cameras to define an origin for the X,Y,Z-axis in 3D space.

Figure 8. Dr. Will Dampier and Molly S. don motion capture suits in preparation for capture session.

Dr. Will Dampier and Molly S. were each fitted with a motion capture suit. 53 reflective markers were placed at precise points on their bodies that would help the system accurately estimate their movement. We first ran the process known as the range of motion. During this process, each actor individually performed exaggerated movements that would help define the widest possible range of movements that they could perform. Example movements involve rolling the neck, circling the arms, rolling the wrists, jumping jacks, toe touches, and jogging in place. All of the data for the range of motion was captured by the cameras and processed in order to construct an accurate skeleton that matched each actor. This skeleton is important for real-time visualization while capturing, and so that when the actors where together in the capture volume, it could distinguish them from one another.

The subjects went through a series of specific moves, ranging from beginner to intermediate. These moves included formal bows, defensive blocks, hand strikes, kicks, a forward roll and two different throwing techniques. Some techniques were shot with both actors at once and others with only one at a time.

It will be possible for future instructors to use Brekel Kinect [20] with Autodesk Motion Builder as a cheaper, more user-friendly alternative to optical motion capture.

2.3 Cleaning and Stitching the Data

While the raw motion capture data was quite accurate, there were a number of instances where certain markers were occluded from the view of every camera. In these cases, the resulting animation of the digital skeleton would contain mechanical ‘pops’ in the animation that look artificial and unnatural. To correct this, I first used Vicon Blade's built in post-processing tools. The software is able to automatically locate the gaps in the data for each marker, trim the key frames on either end of these gaps, and fill them using a simple algorithm that estimates the most likely path between the known positions using their trajectories. Manual tweaking was needed for some of the more 'compact' body positions such as the forward roll, and the kneeling bow. When the data had reached an acceptable level of completeness, I exported the 8 moves, each as separate .fbx files.

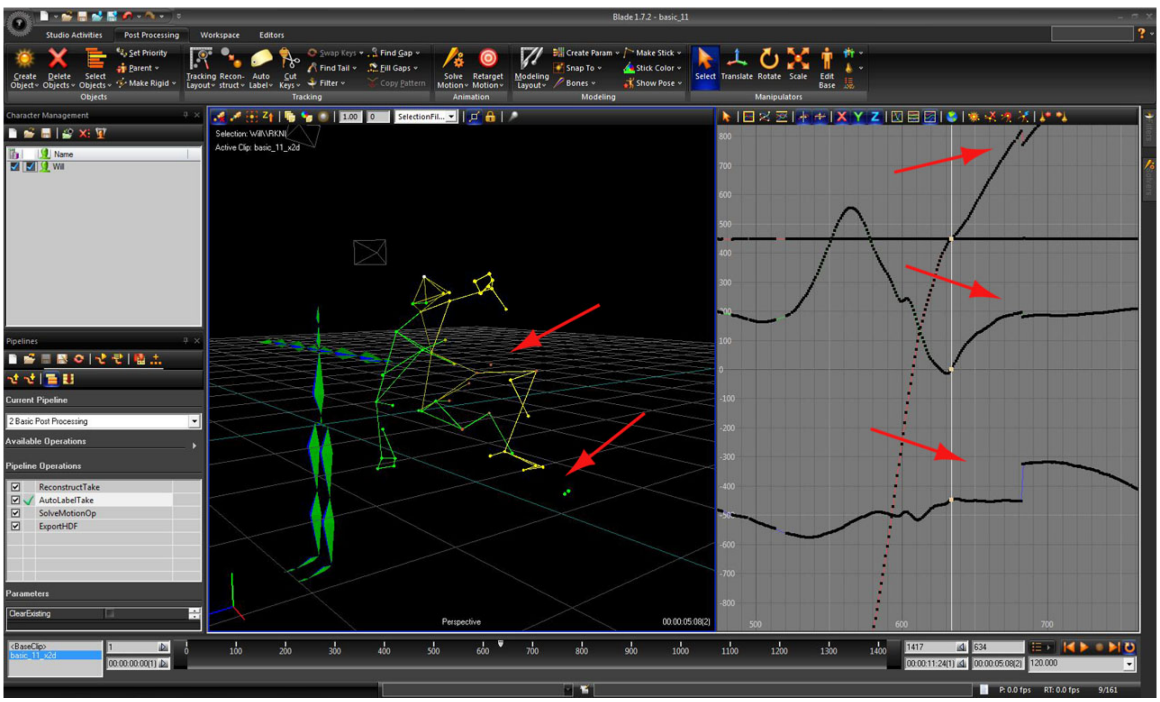

Figure 9. When markers are occluded, they leave gaps in the animation that cause 'popping.'

In order to efficiently store, activate, and blend these animations in Unity 3D, they had to be combined into one, consecutive sequence of animation. Achieving this added another level of complexity to the workflow, and involved another program: Autodesk MotionBuilder 2012.

The most efficient way to stitch motion data together in MotionBuilder 2012 is using the Story window. However, the marker data from Blade could not be stitched until it was translated to a characterized virtual skeleton. To accomplish this, I first created an actor in MotionBuilder, and imported the markers as an .fbx. I deformed the actor mesh until it lined up with the markers, and defined the markers as a marker set in MotionBuilder. Once a maker set was defined for one of the movements, I was able to use that same marker set to quickly connect each animation to the actor. I imported a default characterized skeleton named Mia, provided in MotionBuilder 2012, and set it to receive its movement from the actor. I created a control rig and skeleton for Mia before baking the animation data into the skeleton. At this point, everything could be removed besides the characterized skeleton with the data baked into it.

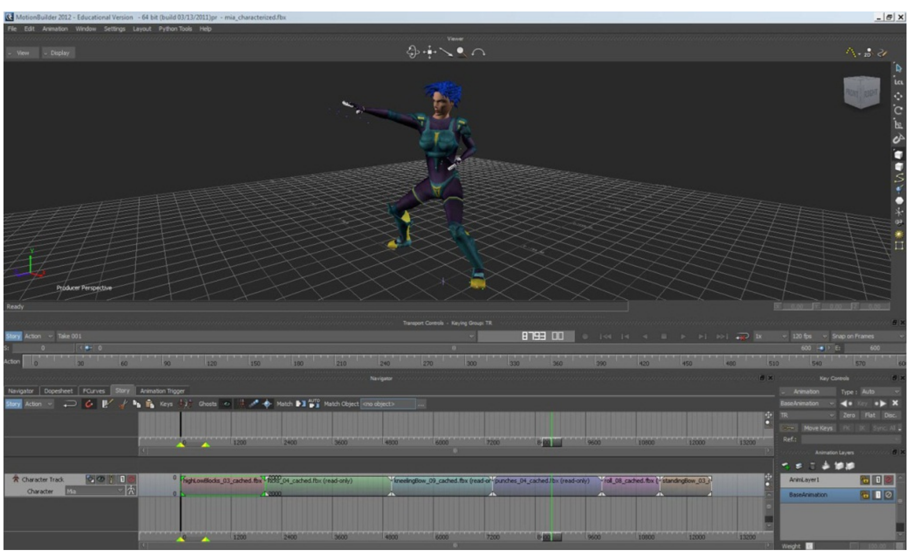

Figure 10. The Story tools in MotionBuilder 2012 are used to stitch together animation clips.

I did this process for each of the judo animation clips, and brought them all into a single MotionBuilder project where I used the Story tools to stitch them together. This was achieved by simply dragging the .fbx files into a timeline and placing them back to back. Finally, I plotted the data to a single skeleton and exported an .fbx file will all of the animation baked into it.

This data was brought into Autodesk Maya 2012 where it was attached to a character model that I had created. The weights were adjusted slightly to prevent any unusual mesh deformation.

2.4 Extracting the Rotational Values of Primary Joints

An unforeseen issue that arose upon analyzing the combined movement data was that the rotational values for joints such as the elbow and the knee (which are restricted to rotating on a single axis) where not displaying the proper rotation. While the NITE skeletons restrict the rotation of these joints for the player's data, the instructor's skeleton was displaying values as if these joints had free movement in all three axes.

This issue was corrected for in Autodesk Maya. I snapped basic cube geometry to the elbow joint on the skeleton and applied an orient constraint so that the cubes axis would line up with the bone, rather than the world space. I parented the cube to the upper arm bone (representing the human humerus bone) and then applied a 'freeze transformations' function, which redefines the cube's position, rotation and scale as 0,0,0. Next, I applied an aim constraint to the cube that caused it to always point at the hand, but only affected its Z-axis. In this way, I was able to extract the angle between the upper arm bone and the forearm bone along the same axis that a real human bone can physically bend.

2.5 Importing the Data into Unity

The data was imported into Unity 3D in as two parts: the skinned character and the cubes with the elbow and joint rotation values. The skeleton with the skinned mesh serves the purpose of demonstrating the moves to the user and allowing them to match their own avatar to the instructors. The cubes are where the actual calculations come from, being that their rotation values are actually more accurate than the skeleton itself. While the four cubes remain visible for debugging, they will ultimately be hidden from the user's view.

3. Build an external program, capable of motion analysis and feedback

3.1 External Program Feature Overview

There are several elements that make the 3D application work as an E-Learning tool. Primarily, I strove to create an experience that was conceptually clear and delivered the message effectively, while also developing a 3D environment that was simple and intuitive to use. The user interacts with the program solely through the Kinect, which was a major challenge throughout the development process. I developed various strategies for interacting with tutorial 'pop-up' instructions, menu and sub-menu navigation, as well as navigating a 3D viewport using hand gestures. Another crucial element of the program was analyzing the player's movements in real time and cross-referencing them with those of the instructor. I built scripts that would track certain features of the body and determine whether the student is performing the technique correctly. If so, the program will notify the student and allow them to advance deeper into the material.

3.2 Body Trackers

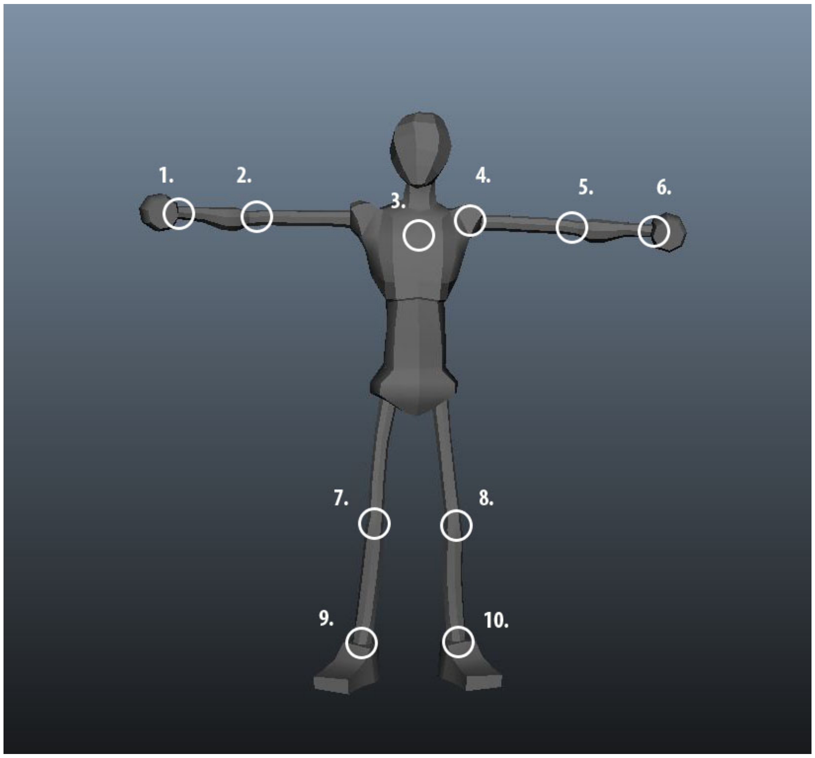

Figure 11. Digital Dojo avatar with numbered body trackers.

The idea of using empty game objects in Unity 3D to track the x, y, and z coordinates of the avatar's key body parts in 3D was crucial to the application. Primarily, I was interested in tracking the user's hands and feet (see Figure 11: #1, 6, 9, 10), as well as the hands and feet of the instructor avatar. However, I also had position trackers attached to the user's avatar's chest and shoulders (Figure 11: #3, 4). One of the primary reasons that this data is so valuable is that 3D coordinates are some of the most accurate readings that the Kinect can provide. Based on my own experimentation and tests, the translation data (of the hands particularly) is quite reliable and consistent, as opposed to the rotational value of joints on the skeleton (Figure 11: #2,5,7,8) [22]. Collecting the positions of the hands and feet at all times allowed the development of systems for in-game navigation, as well as increased the accuracy of the player analyses.

The downside of using x, y, and z coordinate positions is that the readings are not scale-invariant. This means that a player's real-world proportions will result in slightly different results. Informal tests with peers and faculty suggest that the accuracy of the trackers will hold up despite this fact. There is, however, the potential that someone in the shortest or tallest height percentiles could experience inaccuracies with the system.

3.3 Displaying and Navigating the Tutorial Menus

'Pop-up' tutorial menus are used throughout Digital Dojo to communicate with the player. In some instances, they help users learn how to navigate the program, while others display tips about using the Kinect. Each tutorial screen is made up of five parts inside of Unity 3D: A plane with a .png texture on it, two dynamic text 3D boxes, a background plane with a solid colored material, and a script. The plane geometry displays a diagram that illustrates the instructions. This plane swaps materials at run-time in order to display the appropriate diagram. The dynamic 3D text boxes contain written instructions to the user. All of the possible messages are stored as strings in an array within a script. The second text box is activated when the particular string contains too many characters to fit in the first text box. In this case, the extra characters are isolated from the string and printed in the second 3D text box. The background plane serves to separate the 3D world of the application and the 2D help screen in the foreground as well as to contain and organize the instructions.

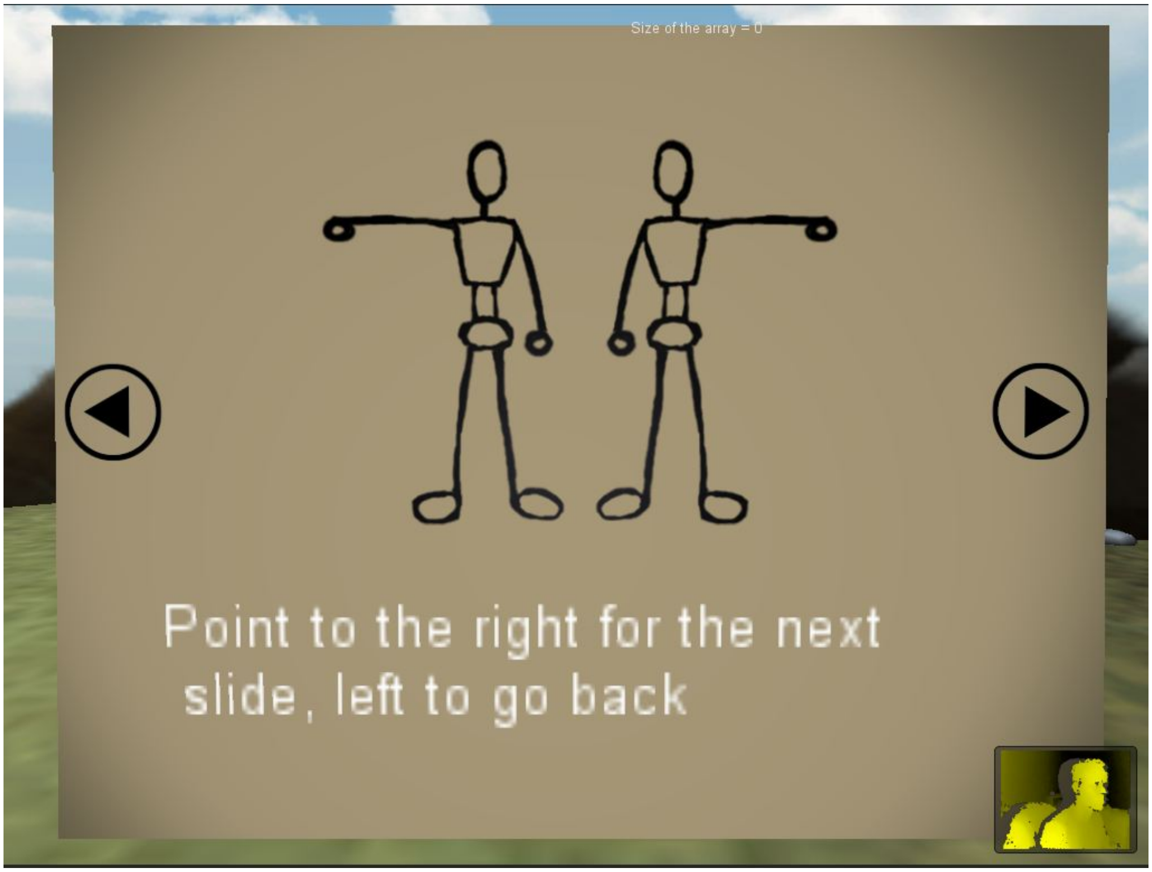

Figure 12. Dynamic tutorial menus help guide the user through the lesson.

The player is prompted to navigate through the tutorial menu screens using two simple gestures: pointing right with the right arm to move forward, and pointing left with the left arm to view the previous instructions. To avoid confusion, if both arms are pointed opposite directions, or both arms are down by the user's side, nothing is activated. This functionality is accomplished with a script that uses the x and y translation values of each hand to determine when to move back, when to stay on the current menu, and when to move progress to the next instructions page or the game.

3.4 Creating and Navigating the Main Menu

This navigation menu is built procedurally at run-time using trigonometry. The menu items are placed along the circumference of a circle with radius *r*. However, to make the selection experience smoother for the player, I only use 120 degrees of the circle. This segment of the curve was chosen because it was easy on the user, and because it is evenly divisible by 1, 2, 3, 4, 5, 6, and 10, depending on how many buttons are needed. The variable *r* can be changed in the script to make the circle smaller or larger depending on the situation. In addition to the radius, I also needed an angle in order to solve for the *x* and *y* coordinates. In order to determine the angle, I used the horizontal X-axis as zero, and calculated the angle above and below x. Sin and Cos were then used to calculate the exact x and y coordinates for each menu item around the circumference of the circle.

3.5 The In-Game Navigation Menu

In order for the player to have control over the pace at which they learn and practice, a menu is available during gameplay. By assuming the “Menu” gesture that is illustrated in the bottom-left corner of the screen during play, the player can call it up (see Figure 13). This menu provides options to resume the lesson, change the camera angle, and return to either of the previous menus. If the player chooses to change the camera angle, they will have the ability to manipulate their view of the instructor in order to obtain the best vantage point.

Figure 13, In-game menu provides extra options once in demonstration or replication phases.

3.6 Navigating the 3D Scene in Unity 3D

A 3D navigation feature is available to users while watching the instructor go through the technique. A script, attached to an empty game object, is activated when the left and right hand trackers are within a *defined proximity.* This script allows the viewer to use his or her hand to pan around and dolly into the instructor's performance. As a result, the viewer may review the maneuver thousands of unique angles and distances. Once in navigation mode, holding the hands within the defined activation distance will trigger the script a second time, relinquishing the user's control of the camera.

3.7 Real-Time Feedback

One of the essential elements of the Digital Dojo application is that it provides the user with real-time feedback. There will be two different ways in which the user receives feedback on their performance at two separate parts of the lesson. The first is by color indicators on four segments of the body: the left shin, right shin, left forearm and right forearm. These colors will let you know how well you're matching the original data while you are still going through the technique. The second instance of feedback will be presented on a pop-up menu and take place after you have successfully completed a lesson. The program will store the number of seconds you were successful and divide that value by the total number of seconds to figure out your percentage of accuracy.

3.8 Level Progression

There are two opportunities for progression in the game. First, the player must complete certain prerequisites within a single belt level. Then, when he or she has completed all of the lessons for that level, they may progress to the next belt rank, unlocking more lessons with more advanced techniques.

3.9 Special Consideration for Hands-Free Operation

My first attempt at a main navigation menu was a vertical column of 3D buttons that appeared next to the user's avatar. The user could hover over them by lining up his or her hand horizontally with the button. When the button was in-line with the player's hand, it would light up. When the player held their hand in position for a specified amount of time, it would activate the button and perform a function. However, after early user trials, it was clear that comparing the user's hand's height with the buttons was an impractical, awkward, and limiting method. Shorter users had a hard time reaching the higher buttons without raising their hand uncomfortably high, and tall users experienced a similar issue when bending over to try and active the lowest menu item. The second attempt at fluid navigation was a modified version of my original design, informed by Hamalainen's work with interactive video mirrors [15]. The two main differences are that the menu is built in a hemisphere around the user and that button selection is now based on the slope of the arm, rather than the y translation value of the hand. With this improved system, the player must simply point at the menu item they want to active, regardless of their physical build.

Figure 14. The new menu design is comfortable for many body types.

4. Design the lessons

4.1 Overview

The lessons will be built based upon Bruce and Levin's guidelines for educational technology [11] and strive to emulate the successful aspects of traditional martial arts training. The material will be verified by consulting a martial arts expert. The lesson will begin with a *demonstration phase* where the student will be shown the technique. The student will then be asked to replicate the movement in the *action phase.* During this portion of the lesson the Kinect will analyze their body and they will be give real-time feedback to help them improve their technique.

4.2 The Avatar

Digital Dojo uses virtual avatars to represent both the player and the instructor. The design and interaction of the avatars were informed by several sources. Both avatars are highly symbolic, faceless, and minimalistic in color, with simplified extremities. In his book, Understanding Comics: The Invisible Art, Scott McCloud demonstrates that simple, iconic characters are often more effective for communication than highly realistic characters [23]. As a learner, a generic, iconic character can help to engross one's self into the material. In fact, if the user cannot relate to their avatar, their "emotional involvement" can create "an obstruction for learning…" [2]. McCloud argues that it is advantageous for the teacher's avatar to be symbolic as well, lest the viewer become "far too aware of the *messenger* to fully receive the *message*." [23]

Some functional aspects of the avatars' design aims to optimize them for Kinect-driven animation. Figure x. shows the avatar design. Limitations of the Kinect that affected the avatar's design are that it does not currently have the ability to track fingers, toes, or facial expressions. The character's (1) simplified extremities prevent the user from becoming confused by arbitrarily positioned fingers. Also, the (2) blank face of the avatar helps to detract attention from the limitations of the hardware in facial animation. Another common issue with Kinect-driven animation is the character geometry colliding with itself. The (3) thin arms and legs help to reduce self-collision caused by inaccurate readings.

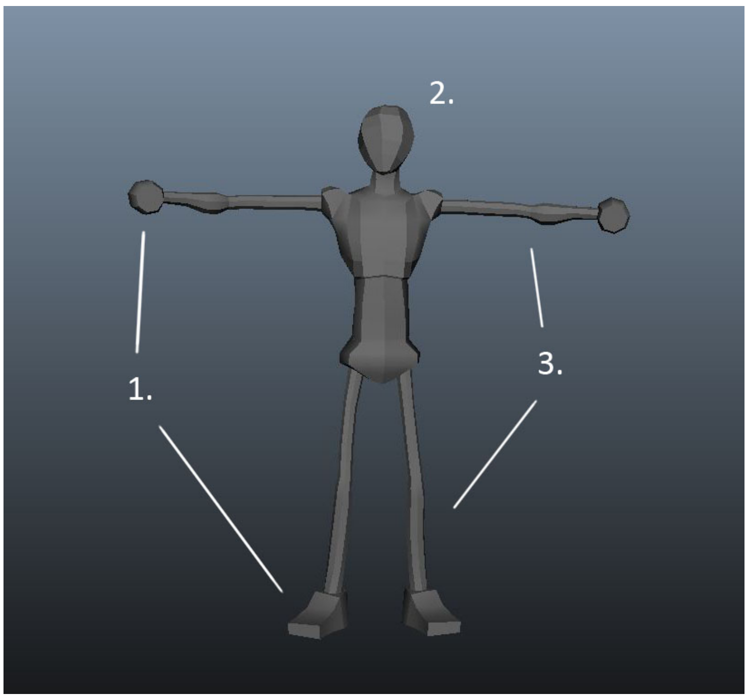

Figure 15. Avatar design.

4.3 Getting Input from the Kinect into Unity 3D

The design and implementation of a user-end motion capture pipeline based on Microsoft's Kinect facilitates ease of operation, even for the least tech-savvy martial artist. In order for a player to use the Microsoft Kinect as a motion analysis device in Digital Dojo requires just two steps:

1. Connect the Microsoft Kinect device to the PC.

2. Download 'All-in-one OpenNI Kinect Auto Installer'

I have produced a detailed tutorial that illustrates the process of setting up the Kinect in order to receive motion data. This tutorial is available on the Digital Dojo website [http://www.animationfarmer.com/DigitalDojo/] as a .pdf document. In addition, all of the necessary hyperlinks are provided on the website to facilitate the users' setup process. By following this initial, basic tutorial, students will be equipped to learn.

4.4 The Belt Menu

The first selection screen that the user is presented with is a menu displaying belt ranks. Upon first starting the program, only the white belt rank is available, with the menu buttons for all of the other belt levels grayed out. Designing the game to limit the content that the student has access to is done for two reasons. The first is safety. The second reason for limiting the player's access to intermediate content is goal setting. [4]

Figure 16. The menu items are disabled to guide the player in a logical manner.

After the belt level menus, there is a sub menu where users can select what specific move they would like to study.

4.5 Demonstration Phase

The user is shown the technique performed by the instructor avatar. The user is first prompted to follow along with the trainer, step by step. This demonstrates situated learning by allowing the student to actively participate while the maneuver is being explained. After the student is guided through the move once, they are prompted to perform the technique by themselves.

4.6 Replication Phase

During the replication phase, the user may practice as much as they want. The instructor cycles through the movement, allowing for as much observation as required in order for the player to understand the basic movement. By performing the technique at full speed, in sync with the instructor, the user will have completed the lesson.

In this stage, the student will receive real-time feedback about how their actions compare to that of the professional. This visual feedback is delivered in the form of colored cylinders overlaid on top of the player's avatar's forearms and shins. When the player is within an acceptable range of accuracy, the cylinder becomes green. If the player's body is in the incorrect position, the cylinder becomes red. During my own trials, I found that an acceptable margin of error for the player was just less than 20 centimeters. However, the script allows for this value to be modified easily for more advanced students or to suit more accurate motion tracking hardware.

In addition to the real-time feedback the player is being presented with, the game is recording information during this phase to determine whether the user has mastered the technique or not. A percentage of accuracy is determined using the body trackers mentioned in section 3.2. Different body trackers are accessed depending on the specific lesson in order to collect the most accurate data. For example, the 'bow' lesson records the coordinates of the chest, both feet, and both arms, whereas the 'punches' lesson only records the chest and the hands. This is not a feature, but a work-around to account for specific inaccuracies in the Kinect's analysis of the player. Once the technique has begun, a timer is activated that records the entire duration of the move. A second timer is activated whenever the player's body trackers are within the acceptable tolerance. These two time values are then divided to find the appropriate percentage of accuracy for that particular round.

4.7 Assessment Phase

During the final stage of learning, the player is presented with a screen displaying the results of their latest training exercise. This screen contains a message, explaining whether the student passed or failed the evaluation, and a corresponding image. In addition, a line of text lets the player know how close he or she was to matching the instructor using the percentage of accuracy discussed in the previous section.

IV. CONCLUSION

I was able to utilize the novel ability of the Microsoft Kinect to interpret a user's body position and create a situated learning experience built on current education theory. The animation data collected from the motion capture shoot was clean, accurate, and engaging. More importantly, the method for comparison between the instructor data and the user's input appears to show promise. In addition to the user analysis, multiple strategies for interfacing with the Kinect were explored and early trials show that they are effective and intuitive for beginners. The user controlled camera navigation is affective and improves the viewing experience for the student. Informal testing of willing participants shows that first time users generally want to keep trying, which infers that this approach could be promising for motivating students outside of the physical training space.

My own personal exploration during this project led me to experiment with what has to be considered when creating a serious game with motion control, enabled by the Kinect. I have explored different gestures and positions that can be collected with the Kinect and tested some of their possible uses in order to discover the current controls.

While Digital Dojo shows great promise as an educational tool for physically based material, it is important to mention the many limitations that became clear through this project. One of the biggest limitations in creating a training application for home use is that the students available, real-world space makes many maneuvers impractical or unsafe.

Figure 17. Drexel Judo Club student performs an exercise that may not be translated to the home.

Also, similarly to Padilha's findings with motion tracking assisted rehabilitation [22], I found the rotational values of the user's joints to be fairly inconsistent. Additional data (for example the hand-tracking technique that was implemented in Digital Dojo) is required to judge the user's real-world position with greater accuracy. Of course, we can predict that the depth resolution and human-tracking software will improve in time.

Another limitation that can be anticipated from this research is the problem of including an entire catalog of martial arts techniques into the application. For accuracy to be preserved, advanced moves that include forward roles and grappling require intensive post-processing work. Although much of this work is automated, the creator must have expensive software and at least an intermediate understanding of the motion capture animation post-processing process. In addition, these same types of movement (that cause much of the body to be hidden from view) are equally as troublesome on the user end interacting with the Kinect.

A limitation of the Digital Dojo process, as pointed out by Ogawa and Kambayashi [24], is that prerecorded animation data does not provide the student with the ability to ask questions. Further research with the in-browser Kinect capabilities of the Zigfu Development Kit (ZDK) [21] may yield a method for students to communicate with each other, or an instructor online while they train.

Finally, after informal testing of the application, it became apparent that the viewer is often put into a position where they cannot easily view the screen. In order to remedy this, Digital Dojo could use some type of additional sensory feedback, such as auditory [19] or physical [25]

The Digital Dojo application is just the base of one component of a potentially huge network of resources for learning martial arts. However, the future of this thesis work lies not only in martial arts, but also in a vast range of physically based subjects such as dance, physical therapy exercises, yoga, or juggling. The Digital Dojo framework that I have created could easily be repurposed to suit any number of topics. As educators begin to seriously adopt educational games and the accuracy and resolution of natural human input devices, such as the Kinect, improve, we can expect to see an explosion of physically based serious games.

Beyond simply re-skinning and repopulating the framework that I have developed, I believe that the true future of this work is expanding Digital Dojo into an interactive, social web space. While Digital Dojo is already a promising model for desktop application based learning, incorporating the Internet would allow users to meet and discuss about the subject matter. Players could maintain profiles that track their progress, and share their accomplishments or challenges with their friends and family through social networking platforms such as Facebook and Twitter. In fact, motion capture capabilities such as the Brekel Kinect software [20] suggest that instructors from around the world could potentially contribute their own lessons to the community. In these ways, connecting this application to the web could lead to an enormous user base for sharing resources and information. The addition of a social web space would not only keep the student connected to the information outside the dojo, but to the community as well.

LIST OF REFERENCES

[1] J. P. Gee, *What Video Games Have to Teach Us About Learning and Literacy*, 1 ed. New York, NY: Palgrave Macmillan, 2003.

[2] A. F. Mamta Shah, "Inquiry, Communication, Construction and Expression (ICCE) Experiences for Transformative Learning in Digital Games," presented at the AERA, Vancouver, British Columbia, Canada, 2012.

[3] S. S. H. M. Masry, "An Analysis of Subjective Quality in Low Bit Rate Video," presented at the International Conference on Image Processing, 2001.

[4] M. K. O. Lee, C. M. K. Cheung, and Z. Chen, "Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation," *Information & Management,* vol. 42, pp. 1095-1104, 2005.

[5] T. I. Perttu Hamalainen, Johanna Hoysniemi, Mikko Lindholm, Ari Nykanen, "Martial Arts in Artificial Reality," presented at the Enhancing Virtual Spaces and Large Displays, Portland, Oregon, 2005.

[6] P. T. Chua, "Training for physical tasks in virtual environments: Tai Chi," presented at the Virtual Reality, 2003.

[7] E. Z. Katie Salen, *Rules of Play*. Cambridge, MA: MIT Press, 2003.

[8] "JUDO," in *Martial Arts of the World: An Encyclopedia of History and Innovation*, ed. Santa Barbara: ABC-CLIO, 2010.

[9] A. C. John Seely Brown, Paul Duguid, "Situated Cognition and the Culture of Learning," *Educational Researcher,* vol. 18, pp. 32-42, 1989.

[10] J. P. Gee, "Language and Identity at Home," in *Situated Language And Learning*, 1st ed New York, NY: Routledge, 2004, pp. 21-38.

[11] J. A. L. Bertram C. Bruce, "Educational technology; Media for inquiry,communication, construction, and expression," *Journal of educational computing research,* vol. 17, pp. 79-102, 1997.

[12] K. Squire, "From Content to Context: Videogames as Designed Experience," *Educational Researcher,* vol. 35, pp. 19-29, 2006.

[13] C. Shirky, *Here Come's Everybody: The Power of Oganizing Without Organizations*. London, England: The Penguin Group, 2008.

[14] B. L. Taku Komura, Rynson W.H. Lau, Howard Leung, "E-Learning Martial Arts," pp. 239-248, 2006.

[15] P. Hamalainen, "Interactive Video Mirrors for Sports Training," presented at the NordiCHI '04, Tampere, Finland, 2004.

[16] N. W. Wagner O. de Morais, "A Serious Computer Game to Assist Tai Chi Trainging for the Elderly," 2011.

[17] "Self-Defense Training Camp," ed: Ubisoft, 2011.

[18] J. Fletcher. (2011, May 23). *Ubisoft takes Kinect to Self Defense Training Camp*. Available: http://www.joystiq.com/2011/10/05/ubisoft-takes-kinect-to-self-defense-training-camp/

[19] "UFC Personal Trainer: The Ultimate Fitness System," ed: THQ, 2011.

[20] J. Brekelmans. (2012). *Brekel*. Available: http://www.brekel.com/

[21] A. Hirsch. (2012, April 1). *Zigfu*. Available: http://zigfu.com/en/

[22] L. B. Antonio Padilha, Mitsuhiro Hayashibe, Philippe Poignet, "Joint Angle Estimation in Rehabilitation with Inertial Sensors and its Integration with Kinect," presented at the International Conference of the IEEE EMBS, Boston, Massachusetts USA, 2011.

[23] S. McCloud, *Understanding Comics: The Invisible Art*. New York, NY: HarperCollins Publishers, 1994.

[24] Y. K. Tomoaki Ogawa, "Physical Instruction Support System Using Virtual Avatars," presented at the ACHI 2012: The Fifth International Conference on Advances in Computer-Human Interactions, 2012.

[25] J. Z. Kenichiro Fukushi, Ramesh Raskar, "Second Skin: Motion Capture with Actuated Feedback for Motor Learning," presented at the SIGGRAPH, Vancouver, British Columbia, 2011.