Virtual Cinematography: Beyond Big Studio Production

A Thesis

Submitted to the Faculty

of

Drexel University

by

Girish Balakrishnan

in partial fulfillment of the

requirements for the degree

of

Master of Science in Digital Media

2013

Dedications

When I was younger, my grandfather told me, ”The past is history. The future is a mystery. But, this is the present because it’s a gift. Don’t waste it, yet enjoy it.” Since then, I have tried to live by that mantra each and every day. While the road from then to the completion of this paper has been an uphill climb, none of life’s challenges were ever designed to be easy, after all.

I would like to dedicate this body of work to my loving and supportive family: Bala (Dad), Usha (Mom), Nalini (sister), the extended family within the states and back in India, as well as all of my close friends [who might as well be considered family]. This paper is the culmination of 23 years of maturity, dedication, and focus, of climbing the hill, and for choosing the path less traveled by. From the very beginning, the decision to pursue a degree in this field was unorthodox for someone in this family, but one driven by a dream. This paper is the manifestation of my dreams materialized, and a vision to pursue an education learning about something I am passionate about and to succeed in it.

In no better words than of the late Walt Disney, ”It’s always fun to do the *impossible*.”

Acknowledgements

I would like to first and foremost give thanks to my thesis committee, Dr. Paul Diefenbach, Jeremy Fernsler, and David Schwartz for not only supporting and encouraging me through this journey but also pushing me to reach beyond my initial vision. Each of you brought a unique insight into approaching this project, from game development to on-set filmmaking to post-production. Your feedback was instrumental in the construction of my thesis and I would not be at the stage I am at without all of your wisdom. Thank you for believing in my crazy ideas.

Thanks to the Digital Media faculty for providing me the platform to create, build, and innovate in sandbox of Computer Graphics over the past five years at Drexel University. Michael Wagner, Jichen Zhu, Glen Muschio, Ted Artz, Jervis Thompson, Troy Finamore, Dave Mauriello, Nick Jushchyshyn - each of you provided me with the foundation to learn and gave me the tools (and the necessary motivation) to achieve my maximum potential. I would also like to give thanks to Mike Glaser who coached me through developing a successful product design and manufacturing ones vision utilizing the latest 3D printer technology.

Thanks to the resources at Carnegie Melon University: Entertainment Technology Center and graduate student Xun Zhang for developing a framework to access the Playstation Move.Me server within Unity.

Sincerest appreciation for Clint Reagan, Previsualization Director at Halon Entertainment LLC, for all of his support in my project, critical feedback on my design, and insight into the previz process in the film industry. In addition, I would like to send my sincerest thanks to Eric Carney, Previsualization Supervisor and CFO of The Third Floor for taking the time to provide me with feedback of how a system like the one I developed could have potential in a production environment.

Most importantly, I would like to give a toast to my fellow graduate students Glenn

Winters, Nate Lapinski, Kevin Gross, Bobby Speck, and Yujie Zhu. Together over the past three years (some even longer) we have learned from each other and triumphed in our individual areas of focus in Digital Media. My sanity would not have been in tact if it weren't for your constant motivation and support throughout each of my personal ventures. I would also like to give thanks to the Digital Media students both past and present that have laid the foundation and continue to set the standard for what artists in this field are capable of. My creative and personal growth is attributed to the mind-numbing amount of hours we spent together dedicated to our work and continuously challenging each other as artists.

And finally, I would like to give thanks to you, those handful of individuals that are taking the time to read through my magnum opus. This document is a testament to the work I achieved by just tinkering with an idea and dreaming of its possibilities, of which I hope you become inspired to do the same in the future.

Table of Contents

List of Figures

Abstract

Virtual Cinematography: Beyond Big Studio Production

Girish Balakrishnan
Advisor: Paul Diefenbach, Ph.D., Jeremy Fernsler, and David Schwartz

In the current production environment, the ability to previsualize shots utilizing a virtual camera system requires expensive hardware and large motion capture spaces only available to large studio environments. By leveraging consumer-level technologies such as tablets and motion gaming controllers as well as merging the cinematic techniques of film with the real-time benefits of game engines, it is possible to develop a hybrid interface that would lower the barrier of entry for virtual production. Utilizing affordable hardware, an intuitive user interface, and an intelligent camera system, the *SmartVCS* is a new virtual cinematography platform that provides professional directors as well as a new market of amateur filmmakers the ability to previsualize their films or game cinematics with familiar and accessible technology. This system has potential applications to other areas including game level design, real-time compositing & post-production, and architectural visualization. In addition, this system has the ability to expand as a human-computer interface for video games, robotics, and medicine as a functional hybrid freespace input device.

1. Introduction

Traditional film and game production presents a linear approach to development, moving from initial conceptualization to producing to post-editing. The notion of virtual production introduces an idea in which the convergence of advanced technologies allows directors to iterate on the production and performance of their film or game with ease. While the pipeline for virtualizing the entire production process will require a paradigm shift in the way current media is developed, there are opportunities for core pieces of the pipeline to be improved upon, notably the concept of virtual cinematography. This process allows for the virtual "filming" of computer-generated performers within a 3D environment using cinematographic techniques that have already been well established in language of film.

George Lucas, one of the proponents for this new production, wished for "a tool simple enough that directors could sit on their sofa watching CSI and mark up shots for the next day." [15] There exists a gap in the market for virtual cinematography outside of the studio environment. With the advent of advanced motion controllers and tablet interfaces, this project seeks technological convergence by offering "off the shelf" hardware and software solutions that will provide small independent film producers and game developers the means to utilize visualization techniques which, up until recently, have been beyond the constraints of their budget. Effective use of visualization techniques can drastically reduce production costs, which will then lead to a more finely tuned final production output.

Both film and game industries offer unique methods of storytelling. Jenkins proposes an idea of technological convergence as a phenomenon that occurs when multiple disciplines collaborate to develop a new product [13]. However, as similar as the production processes may be, there still exists a conceptual and perceptual gap between storytelling in film and storytelling in games. The film industry is beginning to adopt game technologies within their production pipeline in order to visualize shots in real-time, but often struggle to work within the limitations of the game engine. Seth Kendell, Cinematics Lead at Carbine Studios, expressed concerns that cinematics in todays games are used to fill in context that cannot be explicitly shown in gameplay. Modern games attempt to offer cinematic experiences, but often find it difficult to replicate cinematic qualities for their audience. As these two industries begin to find overlap, a new visual language will develop as a hybrid of both game and cinema [10]. By leveraging overlapping technologies from both industries, new toolsets will be developed to enhance the way artists conceptualize their work.

Through the innovation of motion capture technology, which can be seen in studio productions such as James Camerons *Avatar* and 343 Studios *Halo 4*, filmmakers and game designers are actively seeking to develop virtual production technologies that simulate, in a virtual environment, the benefits of being on a live action set. “Through the virtual production process, we are able to optimize the entire film (or game), to make it more creative, and to insure that more of the budget ends up on screen,” [2] says Chris Edwards, CEO of pre-visualization studio *The Third Floor*. With the technology of virtual camera rig, cinematographers of both media are now given back the compositional control of maneuvering and manipulating a virtual camera within a 3D world.

In recent years, the entertainment industry has challenged companies to develop new user interfaces that attempt to close the gaps between the user and the virtual world. Sonys Playstation Move controller, coupled with the Playstation Eye camera, allows a users exact spatial location to be tracked in 3D space. Additionally, with the latest generation of iPad hardware, Apple has offered a fluid gesture-based interface that encourages a natural user interaction design through touch. The combination of both of these technologies will facilitate the development of a unique virtual camera system that can allow directors to control the composition, staging, and timing of their shots by interacting with their virtual world in real time.

This is where the *SmartVCS* (Smart Virtual Cinematography Solution), a new intelligent, collaborative, freespace capture platform differentiates itself from traditional systems. By utilizing consumer hardware and easily accessible software, amateur filmmakers and game designers have the opportunity to replicate the production conditions of professionals in the comfort of their own living rooms. Developing a platform that embraces the language of cinema in an easy-to-use, intelligent interface presents the "prosumer" audience with an opportunity to take expedite production pipelines. Directors, both novice and professional, will no longer require expensive hardware and large motion capture volumes to accurately achieve their vision on every single shot.

2. Background

2.1 Navigating Virtual Environments

From the late 1960s, there spawned an idea where one can navigate and interact with a virtual environment while immersed within a real environment. This notion of creating the perceptual and spatial illusion of a 3D world grew into the concept known as Virtual Reality. While this field is still heavily in development, through the years several researchers played a role in developing key components that contribute to virtual reality including head-mounted displays, tracking systems, and space navigation devices.

Figure 2.1: Ivan Sutherland's "The Ultimate Display"

Morton Heilig proposed the idea of an immersive theater, one where the viewer would be able to see three-dimensional projections, hear stereographic sound, and feel the sensation of motion. [8] In order to develop such a theater, there existed a need for a head-mounted display, first developed by Comeau and Bryan, who utilized magnetic tracking to determine the direction of the head. MITs Ivan Sutherland furthered this system by developing The Ultimate Display, which utilized a mechanical tracking system in which the 3D camera in a virtual worked moved in tandem to the users head via a machine-driven arm. While rudimentary, these systems pioneered the concept of visualizing a virtual world from a user-perspective, whether it was for medical, military, or other visualization application.

In addition to head-mounted displays, there existed a need for more precise tracking and object manipulation in the 3D space. The early 1980s saw a rise in instrumented gloves utilizing optical fiber cables and 3-space sensors to accurately track joint rotations in a virtual environment. This technique of virtual finger tracking, from systems such as the Sayre glove to the Commodore 64s Z-Glove to the Dataglove, became an efficient technique for multidimensional control. [8]

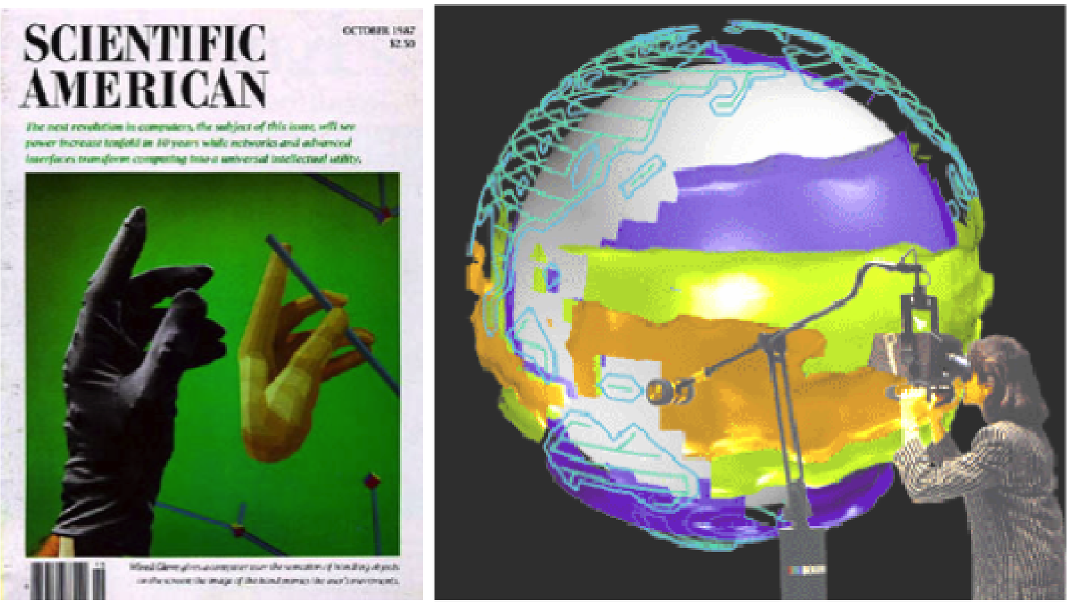


Figure 2.2: Glove-Based Input and Fakespace's Virtual Environment Workstation with BOOM arm

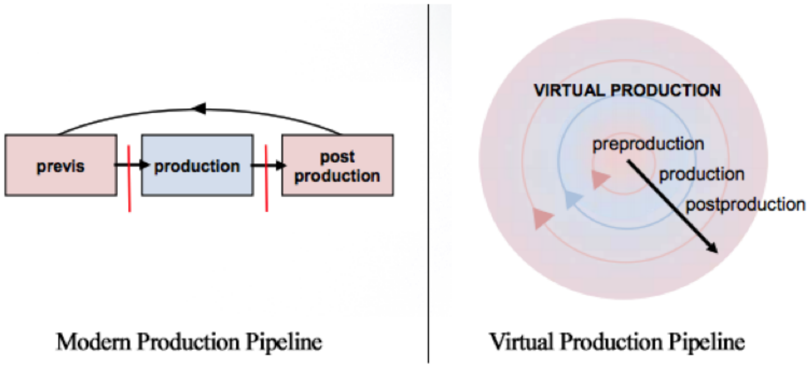
Combining the two technologies, a UNC research group created the first wearable virtual-world workstation. Researchers were able to develop a heads-mounted display that coupled with an ARM gripper to manipulate, rotate, and examine molecular structures. The gripper not only tracked the spatial position of the hands in the virtual world but also provided haptic feedback so that the researcher could "feel" the forces of the object as it was operated with. In order to further this sort of camera manipulation system, Fakespace, in 1988, built a telepresence camera system utilizing a BOOM (Binocular Omni-Orientation Monitor). The system utilizes a monitor fixed onto a mechanical arm to track the perspective change in the virtual cameras view. With Fakespaces technology, researchers were now able to observe virtually created environments with full 360-degree freedom. CAVE Projections, a technology which grew from this notion of freespace motion, developed in

UICs Electronic Visualization Lab, allowed the user to be surrounded by displays in order to achieve a heightened sense of immersion in the virtual world. Utilizing multiple projector setups and often coupled with hand and head-tracking systems, CAVE provides the illusion of being completely engrossed in the virtual world experience, enabling multiple people to exist within the same virtual environment and assume a sense virtual reality. [8]

2.2 Virtual Production in Cinema

Advances in digital technology, allow filmmakers and game designers the freedom to develop new interfaces for producing and interacting with creative content. The process of developing creative content has largely evolved over time, especially due to the convergence of game technologies, and thus has created a bourgeoning new field called virtual production.

Production for films are broken into three core sections, previsualization, production, and post-production. Previsualization is the process in which storyboards of a film are rendered with 3D imagery and audio. The process provides a sandbox for the director to experiment within a “low-cost” virtual environment. It serves as the blueprint for the film and gives the director a sense of how certain shots will affect the narrative. [11, 14] The goal of previz is to establish all the creative decisions of the director, establishing all the shots and leaving little to chance for the production and post-production phases.



Modern Production Pipeline

Virtual Production Pipeline

Figure 2.3: Film Production Pipeline

The techniques for previsualization in film developed along side its medium. The cinematic vocabulary developed over an extended period of time as filmmakers experimented with different lenses, camera moves, and compositions to best deliver their message through a series of shots. “As the language of cinema developed, certain rules of composition have become standardized in the way they are applied to certain shots.” [17] While concepts like the rule of thirds, depth cues, and 180-degree rule have become standard compositional practices over time, these rules of cinematic composition are not law and are subject to being broken and modified to best deliver the shot. In order to develop well thought out shots, rich with compositions that reflect meaning to the story, filmmakers required a technique to “find the frame” and visually map out their story before shooting called storyboarding.

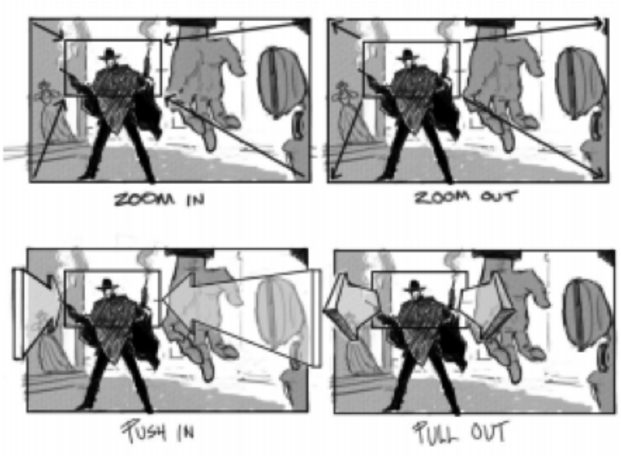


Figure 2.4: Schematic Storyboard Annotations

In its origins, film ideas were translated into storyboard form in order to best put the directors vision down on paper. In order to communicate with other members of the film crew about moving compositions, directors required a special type of diagram, the schematic storyboard, which accurately annotates the film language onto a drawn frame. Storyboards are typically used to lay out key moments that take place within a particular shot and embody the following qualities: “extrema of motion; “representative” poses of a subject, which are in some sense similar to a large range of the observed poses; clarity of expression or pose; and dynamic balance, suggesting the motion in progress.” [8] Within each storyboard, the use of motion arrows & zoom lines, for example, are utilized to best communicate the progression of the camera and/or actors in the frame. In addition, camera diagrams are often utilized to plan the logistics of each camera shot. Using simple shapes and lines, the diagram allows directors to know the physical limitations of the set and helps lay out the movement of the actors and cameras before they shoot.

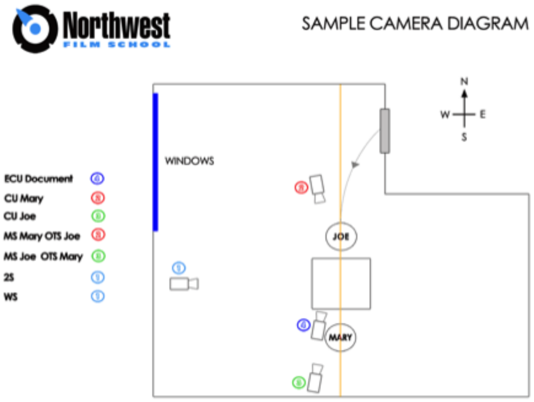


Figure 2.5: Camera Diagram Example

Filmmaker Alfred Hitchcock was famous for being able to storyboard so efficiently and to extreme detail that he often "considered the work of actual shooting to be pure drudgery." [4] Disney was the first to take these storyboards and string them together into an animatic called a "Leica Reel", complete with recordings of the voice actors from the movie. This allowed the directors to view crude renditions of the movie in order to get a better sense of the final films look and sound.

In the 1980s, Francis Coppola developed the notion of "Electronic Cinema" with the very first live-action film that utilized elaborate storyboards and recorded scripts to compose his vision. Film production has been constantly evolving with the introduction of the non-linear editing system in the late 1980s. From this process, the non-linear editor was invented and editors were given more freedom with how they interacted with their film content. This digital edit became the very first virtual representation of a film. [2]

Previz through the 90s grew as more of technical process for visual effects and began to take its own form, separate from storyboarding. During this time, game companies began using the advanced graphical technologies to tell rich visual stories through games. Hackers began experimenting with different game engines through mods, or modifications of the main game, in order to tell their own stories. This movement led to the growth of "Machinima" productions, which are user-generated narratives that are recorded and distributed by the player using in-game graphics and gameplay. [4, 18] Will Pope uses Bolter and Grusins notion of remediation to describe video games as a platform which emulates the techniques of cinema. Machinima is the first defined crossover between games and film industries. [12]

Figure 2.6: Star Wars Episode III Previsualization

Futurists began to predict that one day all rendering for film would be real-time and within a virtual world. [4] As the concepts of visualization improved, filmmakers looked into developing innovative approaches to achieving their visions utilizing both media. George Lucas asserts the latest advancements of the early 21st century in real-time digital technologies moved aside all barriers of imagination presented by analog visual effects. In developing the Star Wars film prequels, Lucas was the first to develop the “modern previsualization process” [11]:

1. Previsualization is created for the director
2. Develop fully edited sequences with narratives of every shot
3. Work with the art department to build digital sets built to scale
4. Utilize textured and modeled characters and sets with basic visual-effects

His goal was to create an iterative system to experiment non-destructively with shots without the stress of making the directorial decisions on a live set. As film budgets grew ever larger, Lucas looked to use previsualization as a tool to save his studio money by effectively building the movie before shooting the movie. By front-loading the art department to work in conjunction with the newly formed previsualization department, collaboration is encouraged in developing assets that work for each shot. From the previz, he was able to output information such as lens choice, angle, and camera path to the live shoot, which thus became a precise process. These templates would then be sent to Industrial Light and Magic, Lucass post-production studio, to handle the final visual effects for the piece. [22] With this process, there is little left to chance in deviating from the directors vision without his/her knowing essentially following a WYSIWYG philosophy.

Figure 2.7: Simul-Cam on the set of *Avatar*

From this, filmmaker James Cameron felt that the previz process was lacking one final component in order to truly bring the director into the virtual world. Through his development on the film *Avatar*, he developed the technology of the Virtual Camera Rig called the *Simul-Cam* [7]. This physical rig was fashioned with an LCD screen with accelerometer, compass, and gyroscope technology to essentially become the “mouse” which controls the CG camera in the virtual world. Everything about the CG camera can be mimicked on the rig such as lens type, field of view, depth of field, etc. as well as camera move techniques such as zooming and dollying. Using a motion capture studio, the position of the rig in real-world space drives the CG camera position and can be recorded in the capture application.

This technology coupled with Lucas’s previz process opened up a new field of complete Virtual Cinematography and Production. Everything from the actors to the environment could be represented in 3D and seen through a virtual viewfinder controlled by the director. [9] The benefits of virtual cinematography lie in its ability to transport the virtual camera around the scene without physical constraints. With this new production methodology, filmmakers now have direct access and hands-on creative input into the movie making experience.

Rob Powers, head of the first Virtual Art Department for *Avatar* states, “Virtual production can liberate filmmakers from the often stifling visual effects and animation process by allowing them to experience the production in much the similar way they do while shooting live action. They are immersed in the world of the film and the entire production benefits immensely from the kind of immediate creative discovery and problem solving.” [2] Even at the FMX 2012 & SIGGRAPH 2012 Conference, virtual production has become a focus within the film and gaming industry worldwide, such as Giant Studios, Digital Domain, and Electronic Arts, for discussion on how it has the potential to change the future of these industries. Douglas Trumball believes that this new era of production, which he coins as “Hyper Cinema” opens up the ability to tell stories the way they were intended to be told due to the ability to “debug the whole movie and iterate change long before building real sets or hiring a real cast.” [1]

In this area of Virtual Space Navigation study, researchers have been studying how to effectively traverse digital landscapes through a variety of visual cues as well as intuitive interfaces. At the “Virtual Production Branches Out” talk at SIGGRAPH 2012, Wayne Stables, Visual Effects Supervisor from Weta Digital, made the bold claim that “the worst place to use a 3D camera is in a 3D package.” [23] This is because the current set of standardized control inputs, the mouse and keyboard, are not designed as devices that can maneuver 3D spaces. In fact, the mouse itself is a peripheral that is, in design, limited to a two-dimensional flat surface plane for motion. While it is possible with the conjunction with keyboard hotkeys, there is a push for more natural user interfaces.

Spatial input, as a term, describes a class of interfaces that are designed for free-space 3D interaction, one that is specifically designed for three-dimensional manipulation, such as the Virtual Camera System. While there exists a perceptual gap between understanding and experiencing 3D space, devices such as the Microsoft Kinects natural body-tracking via a camera sensor and the Oculus Rifts immersive head-mounted & tracked display provide unique approaches to engage with virtual environments at a consumer level. Studies show that "it can be awkward and fatiguing to repeatedly switch between spatial input devices and traditional input devices" independently as well as potentially distracting from the user experience [15]. Observations conclude, however, that spatial interface tools with touchscreens seem natural and enable the user to perform direct 2D manipulation tasks independent of their 3D position. These emerging hybrid 2D & 3D interaction devices provide a unique, unified, and largely unexplored framework for new methodologies of user interface design.

2.3 Virtual Production in Games

The advantage of working within a game engine simply lies in its ability to, in real-time, give feedback on how the game will operate in the final product. While this allows for iterative game design, there is more to be considered when developing a virtual production pipeline for game development. Prototyping and iterative design is key within games, more so than film, because unlike film where virtual worlds are primarily developed to fill a particular shot/camera, games require full virtual worlds to be developed for the player to navigate in.

In developing a game level, for example, designers aim to guide the player through various visual narrative elements through an environment effectively, without holding the players hand through the experience. Grey boxing, a technique developed by the Metal Gear team, allows their designers to rapid prototype their level in a 3D package using nothing more than grey box primitives. This technique allows the designer to rough out their basic level elements without the need for assets. [21] World builders and set designers will then take the grey box elements and swap them with textured assets to populate a particular scene. In developing assets for games, unlike for film, greater optimization in terms of models, textures, lights, and effects is required to run at full FPS on the game engine.

Cinematic Lead Seth Kendell stated, “Content designers make the story. Virtual production in game design requires an active pipeline where, for example, the character designers influence the gameplay as much as the world builders do.” [16] In this case, collaboration on a greater scale is required across multiple departments. Dominic Ciaciolo of Netherrealm anecdotes a scenario where the cinematics department has to continuously modify their camera layouts based on each level design update, which is a live process until the end of production. [6] Since all simulations and actions occur in real-time in the game engine, without an organized iterative production pipeline, all designers, programmers, and cinematographers must continuously stay vigilant to all gameplay updates throughout the production.

Cinematic game development has taken much inspiration from its film counterpart. Most triple-A games of all genres, from Madden to Halo, utilize motion and performance capture systems coupled with virtual camera rigs to create more engrossing and filmic experiences. Adhering to the language of film, game studios often recruit traditional cinematographers and filmmakers to “simplify and accelerate the camera animation and layout process.” [6] These individuals bring their understanding of the language of film and their knowledge of the production pipeline to the realm of video game cinematic.

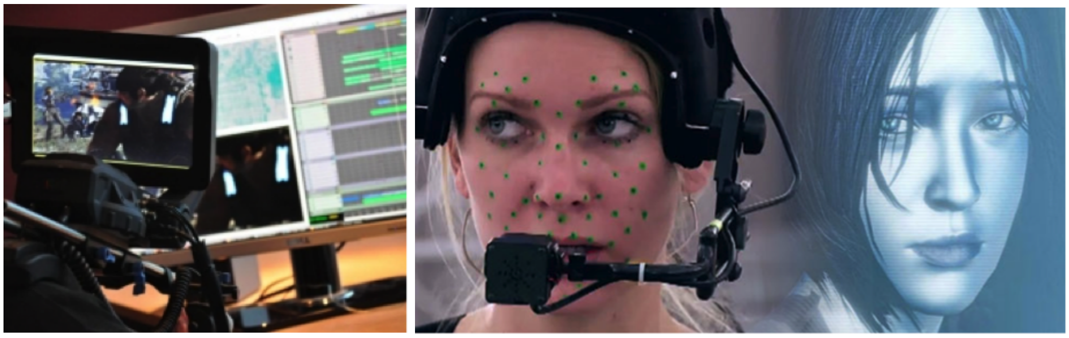


Figure 2.8: Virtual Camera System in Epic Games’ *Gears of War 3* and Performance Capture in 343 Studios’ *Halo 4*

Film previsualization in a game engine is inevitability, according to Habib Zargarpour of Microsoft Game Studios, once the transitional pipeline from the film to game has been established. At FMX 2012, Zargarpour reiterated that having come from film, he feels that there are tools still to be developed that could produce faster iterative design results on the game side while also providing previsualization opportunities on the film side. [6]

3. Related Work

Several academics have expanded upon the concept of developing cinematic tools for previz pipeline for both film and games. Through the following work, I discuss the contributions and limitations of each system and elaborate on how future systems can be built to resolve any shortcomings.

Cambot is an application designed to model that understands the real process of filmmaking. With a script and 3D character, Cambot is able to develop all the possible shots available. The software understands cinematic principles such as staging, blocking, line of sight, one-third rule, etc. in order to properly compose and vary the shots. Through its interface the director is able to modify and record shots within its engine for output to a final video file. [18] In its current form Cambot is limited by the lack of user control in composing the camera freely, something that is crucial to cinematographers. Seeking to primarily be used for resolving camera blocking issues through a rapid database of standardized shots, Cambot does not have the capability to be a full previz tool without returning the control of the camera to the cinematographer rather than relying solely on a computer.

The Virtual Cinematographer is a camera control and directing application that automatically generate camera situations for capturing specific events based on the scene. The application acts as an intelligent camera system that employs basic filmmaking heuristics into the system to enable those unfamiliar with the techniques of cinema to be able to compose shots without breaking cinematic conventions. This real-time application utilizes a 2D layout to build the virtual scene and then allows for the Virtual Cinematographer to automatically generate cameras sequences that correlate appropriately with each shot. [12] This 2D-3D hybrid camera setup is the basis of the intelligence within the SmartVCS. Like Cambot, the intelligence in the system lies in the 2D layout and allows for shots to be blocked, but removes the advanced natural control cinematographers require to adequately previsualize shots.

NUCCI is a package allowing freedom of camera control within a game engine. The user can import models and animation cycles from Maya and operate a camera within a 3D scene. Users have the ability to record and play camera motions as well as set up multiple cameras within the scene. While the interface was simple, the package was limited in being unable to modifying the animations when filming, thus not taking advantage of the real-time potential of the game engine. [19] The ability to import and export camera and scene items is a feature in NUCCI that allows designers to work within their own virtual environment. However, real-time asset modification is a key feature lacking in NUCCI, especially considering it is built within a game engine. The previsualization process encourages experimentation of camera shots, something that could benefit from modifying scene assets in order to better frame the camera.

Cinemotus is a proposed virtual camera solution utilizing game controllers. The system utilizes a magnetic motion detection controller to navigate a camera within a virtual space. In addition, Cinemotus provides robust tools for dollying, tracking, and modifying the camera properties within an existing animation package. This product is currently in development and is set to launch by the end of 2012. [3] Cinemotus provided the foundation of the hardware design of the new virtual camera system. The dual analog control with a large center screen, similar in form to the Nintendo WiiU controller, allows for the viewfinder of the camera to be always visible while having the tactile, analog feedback that allows for precise tweaking of camera properties. The design of my new virtual camera system borrows key elements from this setup, namely its hybrid interface format. This format benefits from its ability to have the


Figure 3.1: Cinemotus

freedom of freespace motion while using the analog sticks and buttons to make finer adjustments. The SmartVCS builds on Cinemotus design by replacing the center screen with a touch screen tablet interface, enabling another level of input with the system for tasks that are best handled on a touch screen.


Figure 3.2: Zeus Camera Tracker & Markers

Zoic Studios proprietary ZEUS environment unification system allows for the real-time compositing of live actors on a green screen onto a virtual 3D set. This system utilizes a special tracking camera placed on top of the set camera that is pointing at an array of tracking markers on the set-roof. As the camera moves about the green screen set, the cameras position is tracked in 3D space and a powerful computer is able to key and process the green screen data in real-time, allowing directors and actors to get a sense of their environment on the studio floor instead of in post. In addition, the system allows for a Unity game engine-based iPad app to modify 3D camera and lighting details in real-time, giving the director complete control of what is happening on set. While the technology is currently being tested on episodic TV productions, Loni Prestere, co-founder of Zeus Studios, predicts that with this next generation of previsualization, “movies that would once cost $120 million could now be remade for $30 million.” [24] By utilizing an open game engine like Unity, the Zoic system potentially allows external plug-ins to be developed for the iPad to build functionality to the system. While this freespace system was ideal for large green screen studios, the feasibility of implementing this tracking system in a smaller volume would be unnecessary. At smaller volumes, basic camera tracking systems such as the Microsoft Kinect and Playstation Eye provide the same degree of tracking fidelity achieved with this studio-sized system.

Zviz is a previsualization solution built internally at Industrial Light & Magic that is aimed on becoming a consumer level product that allows filmmakers to block shots outside of the studio environment. The system consists of three modes consisting of building a set, animation & shooting, and editing. The director be using a system that utilizes a game controller-like input to build a set using premade assets as well as the ability to draw in 2D cards for custom props. The system is built within a 3D game engine environment and supports in-engine camera rigs, multiple lenses, and film backs as well as basic real-time lighting. This intelligent system is still in the early stage of development but is designed with the hopes of enabling unfamiliar amateur filmmakers the same tools and assets used in studio previz using consumer accessible hardware and easy-to-use software. [20] While little is known about the look of this user interface, the sort of ease-of-use and widespread accessibility that is aspired for with Zviz is something that is a core pillar I hope the design of my system will achieve.

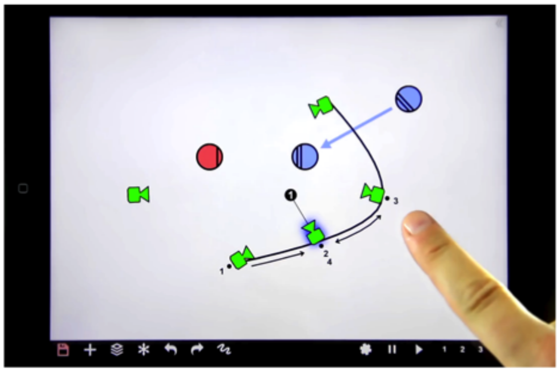


Figure 3.3: Camera Blocking in Shot Designer

Lastly, Shot Designer is a new cross-platform camera blocking and shot-planning application for desktop and mobile that allows for directors to lay out basic assets, camera moves, shot lists, and annotations using prebuilt and custom assets. The application uses a camera diagram-like interface to build and layout the core character/scene assets and cameras. From there, the director can sequence the shots via a shot list and animate the action in real-time. In addition, the director can pin storyboards and reference images to each camera as well as annotate each shot with notes for reference. This app is being used by directors and cinematographers across the industry and shows promise as an off-set planning and on-set referencing tool. [5] The intelligent use of touch where needed to draw and lay out camera positions is furthered by the hybrid nature of the SmartVCS, where physical controls could add tactile feedback to produce the perfect shot composition. While Shot Designer is built on the Flash platform, the SmartVCS is built within the Unity game engine, which allows the easy modification of camera and scene properties through an overhead while allowing the ability to drop the camera into the 3D world to view the desired shot in-camera and in real-time.

4. Methodology

The following outlines the exploration in developing a virtual camera solution under the limitations of consumer level hardware. While under the constraints of today's technologies, there exists no single purpose solution for developing an intelligent, freespace, and collaborative virtual camera system. Thus, an evaluation of both the requirements of virtual camera technology and the vast array of consumer level products and platforms was essential. As a result, it was necessary to marry several pieces of hardware together under a common platform in order to develop this system.

The SmartVCS stands as proof of concept of the larger idea, that consumer level technologies can provide an equivalent, if not greater function to augmenting the virtual camera capture experience. In this chapter, we will discuss the system architecture of the SmartVCS, explore the user experience of the system, and delve into its practical applications. We close with a comparison of the SmartVCS to existing solutions and overview its limitations and future development.

4.1 System Requirements

Motion capture is the process of recording positional and rotational motion of tracked point within a designated capture volume. The process enables rapid, real time animation results that encourage an iterative workflow. An aspect of the motion capture process is camera capture, the ability for camera movements to be recorded within the capture volume to augment the motion of a virtual camera in the digital space. Like real cameras, a virtual camera system is often driven by a camera operator who has the ability to pan, tilt, and dolly around the stage while the actor is performing or after the fact.

Modern virtual camera systems, like the Optitrack VCS, provide real-time feedback through four key tracking points which generate the position and orientation of the system by the motion capture camera system. The physical system provides the camera operator with a visual of the virtual camera position à la an LCD display on the system which is tethered via a display cable to a computer streaming the motion capture software. In addition, the system has precision virtual camera controls mapped to analog joysticks, buttons, and scroll-wheel input. With the system's flexible design, operators can modify the form factor of the device through shoulder, tripod, dolly, and hand-held mounts.

Figure 4.1: Tradition Virtual Camera Capture in a Motion Capture Studio with the Optitrack Insight VCS

While the above outlines the motion tracking and interface fundamentals required for a virtual camera system, the SmartVCS strives to be a cinematography solution. In order to support the intelligence required for a ”smart” system, a computing device with logic is necessary. The device required needs to be easy to develop on and support a dynamic user interface to adapt to different cinematography scenarios such as scene development, shooting, and lighting. In addition, the device requires strong graphical processing power to display complex models, animations, lighting, etc. within a virtual scene.

To leverage collaborative cinematic development, a networking infrastructure is necessary to allow for multiple systems to be used in tandem with desktop hardware. This would facilitate several users to interact within a singular virtual world via a wired or wireless signal. Further extensions of the system such as augmented reality, requires a backfacing camera in order to capture the environment in which virtual objects are to be projected on to. Lastly, the system needed to be flexible and mobile, reflective of the consumerist limitations placed on the device. This means a system that can easily be put together and taken apart utilizing hardware devices accessible to a mass audience.

4.2 System Architecture

Utilizing a combination of game and production technologies, it is possible to develop a robust real-time virtual camera system. In development of this system, both the hardware and software must work in harmony in order to build a natural hybrid interface.

4.2.1 Game Engine

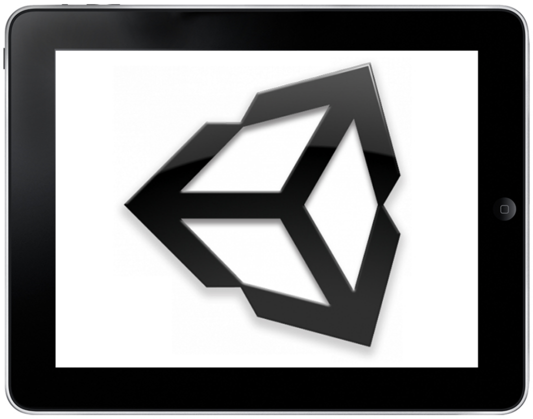


Figure 4.2: Apple iPad running Unity

This virtual cinematography solution utilizes a touch tablet computing system running a 3D game engine as the display along with a motion gaming controller for spatial tracking. The choice of game engine was made based on the real-time benefits granted by the engine such as iterative design, live feedback, and visual fidelity. Through real-time rendering, complex models, environments, effects, lighting/shadows, etc. can be displayed on the screen and modified without the need to pre-render each frame.

Unity [1] was chosen as the core development platform for the SmartVCS due to its cross-platform flexibility and ease of adoption. Being such a wide and open platform, Unity has gained the support of thousands of game designers, from independent to studio production

---

[1] `http://www.unity3d.com`

teams. Unity is also a prime platform for mobile development due to its rich support for a native mobile 3D graphics engine and support libraries for developers. That, along with its vibrant community, broad consumer-base, and rich developer resources, made Unity the platform of choice for the SmartVCS.

Earlier tests included desktop data to be streamed directly onto the tablet device in real-time. This proved successful as a desktop extension and allowed the core processing to be handled by a computer interface. The primary drawback was the lack of true native multi-touch interaction as well as core tablet functionality such as the camera for AR or local storage for data export. By utilizing Unity's native mobile framework, while the limitations of a mobile operating system were persistent, display latency was less apparent since the tablet was not streaming high-framerate visual data across a wireless network. This allowed for snappier results that are often evident in native tablet applications.

4.2.2 Input

Taking a look at the consumer product marketplace, the Sony Playstation Move family of products (the Eye camera, Move controller, and Navigation sub-controller), offered true one-to-one tracking for the Playstation 3 game system. Other motion devices such as the Nintendo Wii Remote, Microsoft Kinect, and Razer Hydra offered strong motion control solutions, but lacked the real-time wireless tracking fidelity that could only be achieved with Playstation's solution. With the Playstation Eye camera's 75-degree field of view, the Move controller can achieve a max range of 12.5 ft. width by

Figure 4.3: Playstation Eye Camera, Move Motion Controller, and Navigation Analog Controller

10 ft. depth. Much like the Zeus camera system, the

Move system relays to the game system the position of the controller in real world coordinates based on its relation to a camera, in this case the Playstation Eye. This level of tracking fidelity allows for the ability to have true low-latency spatial interaction within a 3D environment.

The Playstation Navigation controller is a wireless input controller used in tandem with Move motion controller that enables more sophisticated control with its analog stick, buttons, and pressured triggers. Initial interface explorations of the SmartVCS (Figure 4.4) included a full touch-based system for movement, focal-length modification, scene adjustments, and recording/playback.

Figure 4.4: Touch-Based Virtual Camera Test

Even with the accuracy afforded with touch interfaces, a greater level of precision is required to control a virtual camera. In addition, being a cinematography solution, a design decision was made to remove the user’s fingers from the tablet display of the virtual camera viewfinder. Analog stick input is a natural user interface that is widely supported in traditional cinematography for controlling motion-controlled camera systems. By supporting the analog input afforded by the Navigation controller, the user is now able to compose shots with accuracy and with no visual obstruction of what they are filming.

4.2.3 Tablet

The controller is bound to a multi-touch tablet via a case and is be calibrated to stream the position of the tablet in 3D space in real-time. Tablets, namely the Apple iPad, were chosen as the core input device based on research of natural human-interface interaction and an understanding of their ubiquity in the consumer spectrum. Beyond simple touch input, tablets allow for fluid user interface design, one that does not relay on a static set of buttons but rather on context sensitive input design. Inspired by the intuitive design of the Shot Designer iPad app, gestural input control provides users a playground of possibility not limited by analog input. Tablets offer consumers, at an affordable price point, the ability to interact with images through touch while utilizing the evolving 3D graphical and processing ability of the system.

4.2.4 Networking

Sony released an an app for the Playstation 3 in conjunction with an SDK called ”Move.Me”, which enables the data of the Move controller tracked by the Eye camera to be streamed across a server via direct IP connection. Utilizing the Carnegie Melon University Unity-Move package, developed by graduate student Xun Zhang, the data streamed by the Playstation 3 can be read into the Unity game engine on a desktop PC. With modifications to make the package mobile-friendly, the Playstation is able to stream position, rotation, acceleration, and button-input data of each Move and Navigation controller directly to a mobile build of the Unity game.

With the SmartVCS iOS app opened, the user is requested to connect the iPad to the Move.Me server in order to begin the real-time stream of data to the app. With the Move.Me app open on the Playstation 3 and the controllers are calibrated to the game system through the Playstation Eye, the user can connect directly to the server via direct IP address. Now moving the tracked iPad within the capture volume determined by the Eye camera will move the virtual camera within the iOS app.

4.2.5 System

The combination of these individual pieces of technology with the software is the backbone of the SmartVCS. Through a mobile build of the application, the user is able to drive the software using both the touch and analog inputs through the device. The controllers are registered to the Move.Me server with a simple calibration step and are now accessible remotely via IP connection. The SmartVCS app on the iPad connects to the server and receives the necessary spatial coordinates of the Move controller as well as button/trigger input of the Navigation controller. The Move data is driving the position and orientation of the in-game camera in real-time. The Navigation controller is mapped to specific functions including swapping camera lenses, global camera position modification, and record/playback control.

Figure 4.5: The SmartVCS: Smart Virtual Cinematography System

Pre-made assets are loaded into the tablets mobile game engine through a standalone app and the user is be able to build a rapid prototype of a desired scene utilizing the touchscreen. The Unity game engine is used as the framework for this standalone app development system. The robust mobile engine has a low barrier of entry for consumers, allowing for accessible development and extend-ability within the target market.

Intelligence is core to the user experience of this virtual camera system. The very nature of an update-friendly software-based camera system affirms the intelligence potential of the SmartVCS. Being able to add further content and functionality to the application, tweak settings based on user preferences, and customize the experience for the system are all experiences that can only happen with such a device. The system allows for basic camera intelligence functions that abide by the language of cinema. Building off of the research on the intelligence framework set by NUCCI and the Virtual Cinematographer, the system informs users of basic film principles such as the 180-degree rule and one-third rule while allowing them so select from a set of commonly used shot compositions such as extreme close-up and long shot. The user is be able to select cameras based on a set of real world lenses and aspect-ratios in addition to having the option to modify the individual access sensitivity of the camera.

As the user begins to prepare to shoot, they will be able to drop the camera into a 3D environment and move freely with physical tablet to record several takes of the camera move. The user will be able to review and save only the takes that have been approved. The intelligence of the system will prevent the user from making film decisions and camera moves that are not ground within what is possible in the real world.

With this system, as the user moves around the space, the virtual camera in Unity moves in conjunction, with the iPad acting as a virtual viewfinder into the 3D world. This immersive experience allows the user to compose and frame shots in real-time shot to block out their camera angles. The system utilizes custom recording scripts modified from the Object

Motion Studio package to capture the camera data for review & playback. Users have the ability to approve a take and save it directly onto the device as a recorded sequence that can be shared or as RAW keyframe data that can be ported to an external program. Currently, the SmartVCS supports recorded sequences to be saved and imported into Autodesk Maya for further motion editing.

4.3 User Experience

The SmartVCS, in its current prototype form, is a feature rich system for consumer level virtual moviemaking. The feature set chosen for the device are by no means all the capabilities and potential of the system, but rather carefully crafted and polished aspects of functionality that best showcase this virtual camera solution. Furthermore, all user interface input controls are designed for the purposes of this prototype and can be customized using an on-screen inputs settings menu based on the user's preference. This section emphasizes the five pillars of the SmartVCS - virtual camera control, scene & lighting modification, record & playback review, intelligence, and collaboration as well as provides the grander vision of the system beyond its prototype state.

Figure 4.6: The SmartVCS User Experience Flowchart

4.3.1 Virtual Camera Functionality

The following section overviews the core camera functionality of the SmartVCS. Being a cinematography tool, camera development was a core area of focus for this system.

The camera controls built for the SmartVCS include freespace and tactile input, sensitivity controls, focal length adjustment, and aspect ratio considerations.

Freespace, Analog, and Touch Motion

With the SmartVCS, the user is free to maneuver the virtual camera naturally within the tracking volume. The Eye camera's wide-angle field of view permits a spacious capture volume limited only by the camera's ability to ”see” the Move tracking sensor. With freespace motion, the user is able to pan, tilt, and roll the camera naturally while moving about the space. When users move outside of the camera's visible range, the system automatically switches to gyroscope and accelerometer control for the virtual camera until the tracking sensor is again in the camera's view. Motions made to the system are tracked in the real world and viewed on the tablet display in real-time with no noticeable lag.

In order to extend the capture range of the system, like commercial virtual camera systems, users may use the analog joysticks in tandem with freespace control to move about the 3D environment. The system supports camera tracking, dollying, panning, and booming (Figure 4.7). In addition, in order to aid with camera moves such as whip-pans, which require the system to be spun a full 180-degrees, the SmartVCS supports analog orientation control.

Figure 4.7: Analog Joystick Controls for the Virtual Camera: Track, Dolly, Boom, Pan

With these controls, the user is able to navigate across large 3D spaces without physically moving to that location. This is helpful for establishing the starting point for a particular shot. The analog control is designed to enhance the freespace control with precise input that works similarly to motion-controlled camera rigs utilized on a live-action set. With both modes of input enabled, the user is able to achieve both handheld (freespace) and steady (analog) shots.

In the event the user does not have access to a Playstation 3 or the Move family of products, on-screen touch regions are available for the tablet in order to facilitate input. With a combination of touchpads and sliders, the user is able to move about the 3D world with the same standard controls as in first person action games for tablets. The user also has the ability to jump so specific sections of the virtual set by simply tapping on the desired location on an overhead map layout. While functional as a touch-only solution, the SmartVCS is, at its core, designed to be a hybrid device that marries freespace motion, analog input, and touch to maximize its capability.

Tracking Sensitivity

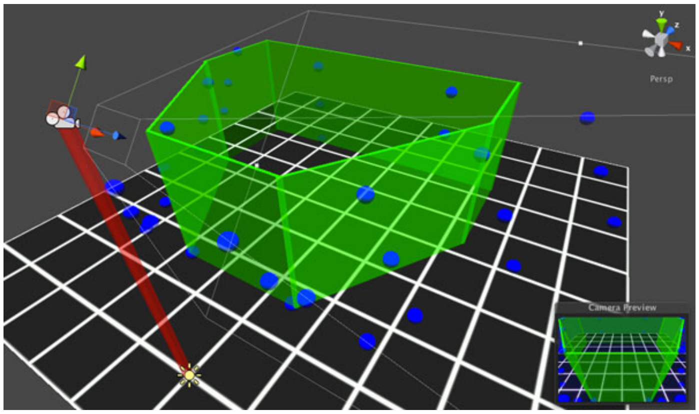


Figure 4.8: Tracking Volume of the SmartVCS

In designing a mobile system, it is important to establish the relationship between the physical system and the virtual camera. The tracking volume is considered to be the area within the virtual environment with which the system is able to move within the physical space. With the system's tracking volume peaking at a 12.5 ft. by 10 ft. playspace (with respect to the Playstation Eye's visible capture range) (Figure 4.8), the SmartVCS supports controls to modify the planar and vertical tracking sensitivity independently. With the Navigation controller's directional pad, users can increase or decrease the sensitivity to each axis depending on the desired camera motion. This allows for one step in the real world to become one virtual step under standard conditions, one inch under low sensitivity, and one football field's length under high sensitivity (Figure 4.9).

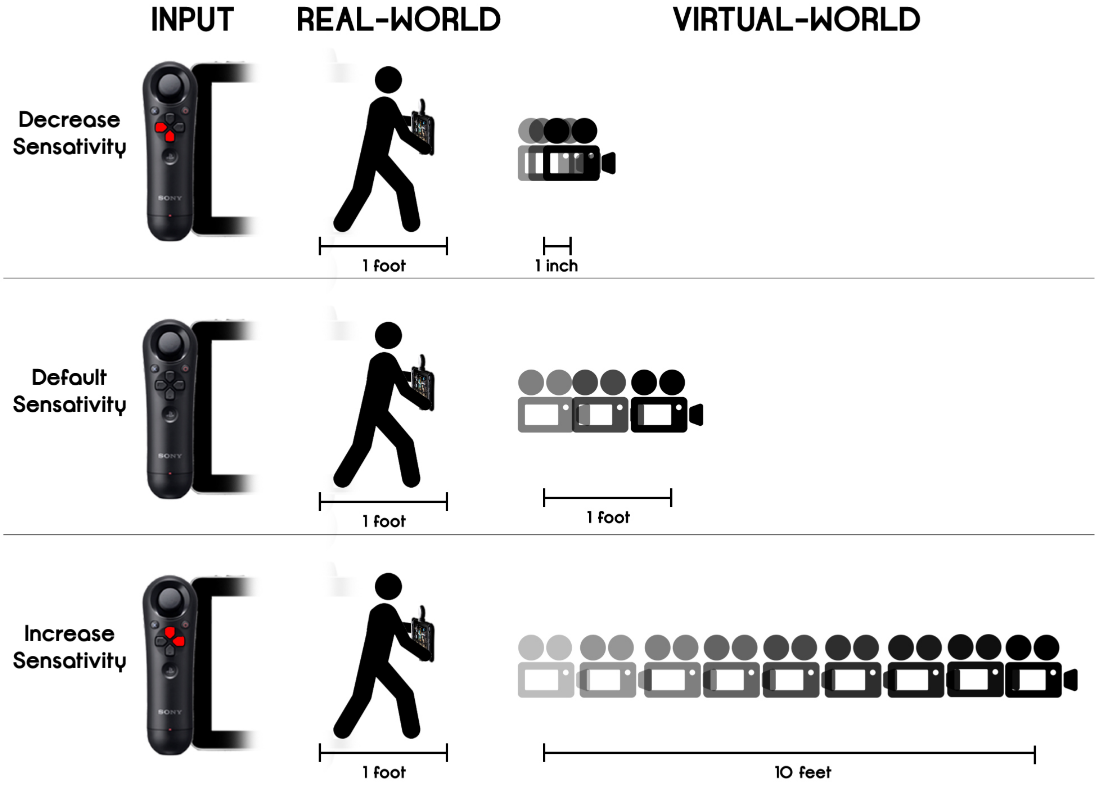


Figure 4.9: Tracking Sensitivity Real-to-Virtual World Comparison

For example, if you wish to accomplish a tracking shot of a virtual character walking down a hallway, you would be able to decrease the vertical sensitivity of the camera in order to smoothly follow the character's horizontal motion. Techniques like that are useful in establishing the range of motion of the device for a particular a shot, a practice commonly used in traditional motion capture in order to pre-plan the motion of the shot. With sensitivity control, the user is now able to achieve a greater range of motion within the limitations of the capture volume, and most importantly, is given the freedom to modify these parameters live between & during takes.

Focal Length Modification

On a live-action shoot, camera operators prefer to have a specific set of lenses to swap in and out depending on the shot required, from macro to telephoto. The SmartVCS has made the process of swapping virtual lenses simple with its built-in supply of standard prime lens types ranging from 20 mm - 300 mm. Utilizing the Navigation controller's face buttons, the user can toggle between these standard lenses with ease and see the real-time feedback through the virtual camera's viewfinder on the tablet display (Figure 4.10). In addition, the user has the ability to dial in the focal length using the controller's soft-triggers or enter a specific lens value using an on-screen keypad. While the system currently supports prime lens types, it can be expanded to feature a full array of industry standard and camera-specific lenses.

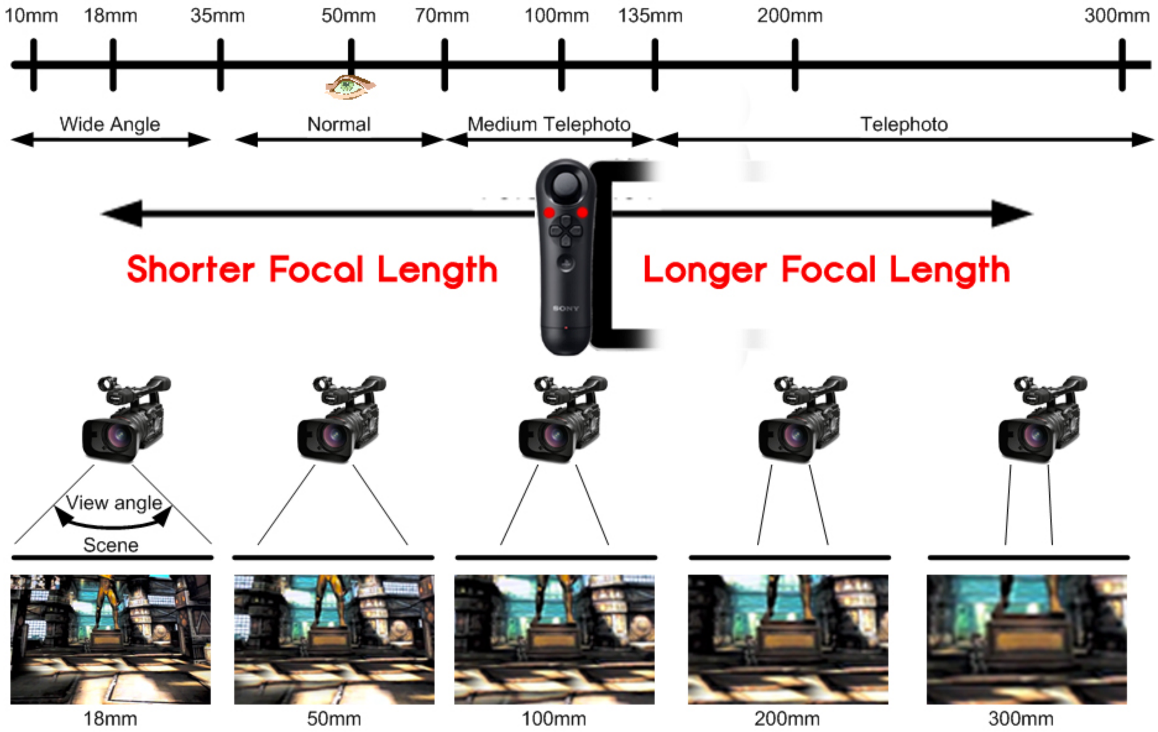


Figure 4.10: Modifying Virtual Camera Lenses

Film Aspect Ratio

Being able to view through the camera lens with guidelines highlighting the film's aspect ratio is a powerful compositional and framing tool for cinematographers. With the SmartVCS, the system is able to simulate a variety of standard film aspect ratios including *4:3*, *16:9*, *1.85:1*, and *2.35:1* (Figure 4.11). Beyond simply having access to common aspect ratios, the system allows for an on-screen keypad for custom entry of the desired aspect ratio required for the particular virtual shoot.

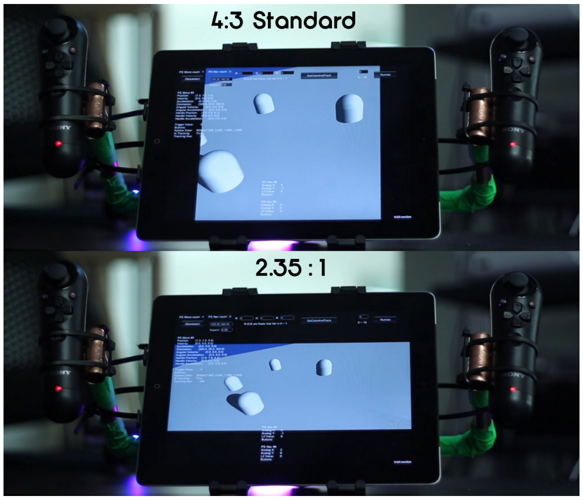


Figure 4.11: Aspect Ratio Modification for the Virtual Camera

4.3.2 Scene & Lighting

Scene Modification

Inspired by Shot Designer's scene modification toolset, the SmartVCS support real-time modification of scene elements that exist in view of the virtual camera. To move objects within the scene, utilize the touchscreen to point and drag objects from their original to new position in camera-space. This is helpful for directors who want a 3D sandbox to manipulate objects around in order to lay out a scene. In addition, the touch-drag function could be used as a compositional tool to better frame foreground and background objects within a desired shot. This feature could be expanded to include a more robust scene layout builder where new objects could be inserted or removed from the scene.

Traditional systems require an external operator to move objects on a remote desktop station, disparate from the virtual camera operator on the motion capture stage. Scene modification directly on the SmartVCS system provides directors greater hands-on control of the 3D environment while they are planning their shot. That being said, the SmartVCS prototype does support remote scene modification within a networked sandbox environment (See 4.3.5 Collaboration: Networked Collaboration).

Lighting

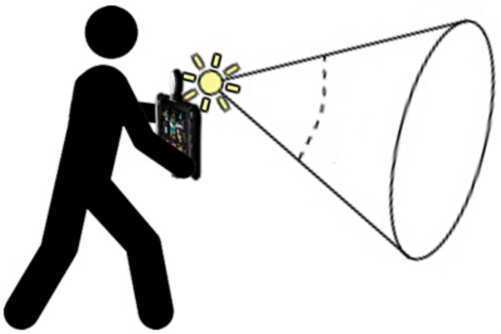


Figure 4.12: Freespace Spotlight Placement in Virtual Scene

Lighting, being an integral component of the cinematic process, is a key aspect of the SmartVCS that is not found in traditional systems. Users have the ability to choose from a variety of standard virtual lights including point lights and spot lights. Users can select their light type and utilize on screen sliders to control cone-angle and penumbra-angle values. With the freespace control, they can actually place the lights in the scene through the physical motion with the system. The desired light is parented to the camera allowing you to move the lights with the same control and flexibility as a Director of Photography would on a live-action set. This precision control of lighting combined with the real-time feedback of in-game shadows provided by the Unity game engine, make the SmartVCS a unique platform for lighting a scene. Further intelligence can be built directly into the system to support standard light kits, colored gels, and more complex lighting setups.

4.3.3 Capture, Playback, & Edit

Being able to record, review, and approve multiple camera takes is a crucial component of the virtual camera pipeline. Existing virtual camera systems require an external operator to, in MotionBuilder, set the animation clip as well as begin/end the recording. With the SmartVCS, all the recording functionality is built directly into the software.

Recording

With the pre-animated scene loaded on the tablet, using the system's shoulder buttons, the user can scrub through the entire animated sequence in real-time. Once the user has determined the appropriate start point of the animation, the scene is locked and the user is now free to frame the camera for that particular shot. With the camera framed, by pressing the designated record button on the Navigation controller, the user is able to use a combination of freespace motion within the capture volume as well as analog joystick input to record a take of the shot.

Playback Review

After the user has completed the recording, they enter the Playback Review interface (Figure 4.14). In this view, the user has the ability to review the take they had just recorded.

Using the same shoulder buttons and touchscreen playback controls, they are able to scrub through the recorded camera & scene animation. If the take is approved, the user can use the touch screen slider controls via the Playbar to designate the start and end keyframes of the camera animation. The user has two options to save the recording, as a locally stored rendered video file on the tablet itself and as a text file of keyframe data that can be imported into production software such as Autodesk Maya. If the take is not approved or the user has completed saving their camera data, they may return back to the 'Live' mode and begin framing their shot for another take.

Figure 4.13: Playback Review Interface

Exporting Camera Data

While visualizing camera moves on the SmartVCS allows users a quick way to iterate through camera moves, advanced users will want to take a captured camera motion path and work with it in post. Within the app, the user has the ability to save multiple approved takes onto the tablet device for motion editing in an external application. Connecting the iPad and launching iTunes enables users access to each camera take via iTunes App File

Sharing. Each take is saved as an individual text file and, by default, is named ”ShotNumber_TakeNumber_Date-Timestamp”. The text files contain the positional and rotational values of the virtual camera at each frame of the recorded animation. Additional data can be loaded into the text file including, focal length and film aspect ratio.

Industry standard animation applications such as Maya, 3ds Max, Lightwave, and MotionBuilder support custom written scripts to read external files. The following is an example of a text parser in Maya which will read the inputted camera data file received from the iPad, create a new camera with the correct shot code, and assign a keyframe to the camera for each frame of SmartVCS recorded animation (4.15). After loading in the existing scene file, the user is able to see the same camera move achieved on the SmartVCS translated within the desktop 3D application. From there, users can apply the same advanced motion editing techniques such as keyframe interpolation and frame reduction techniques that would be used with data received from the traditional virtual camera system.

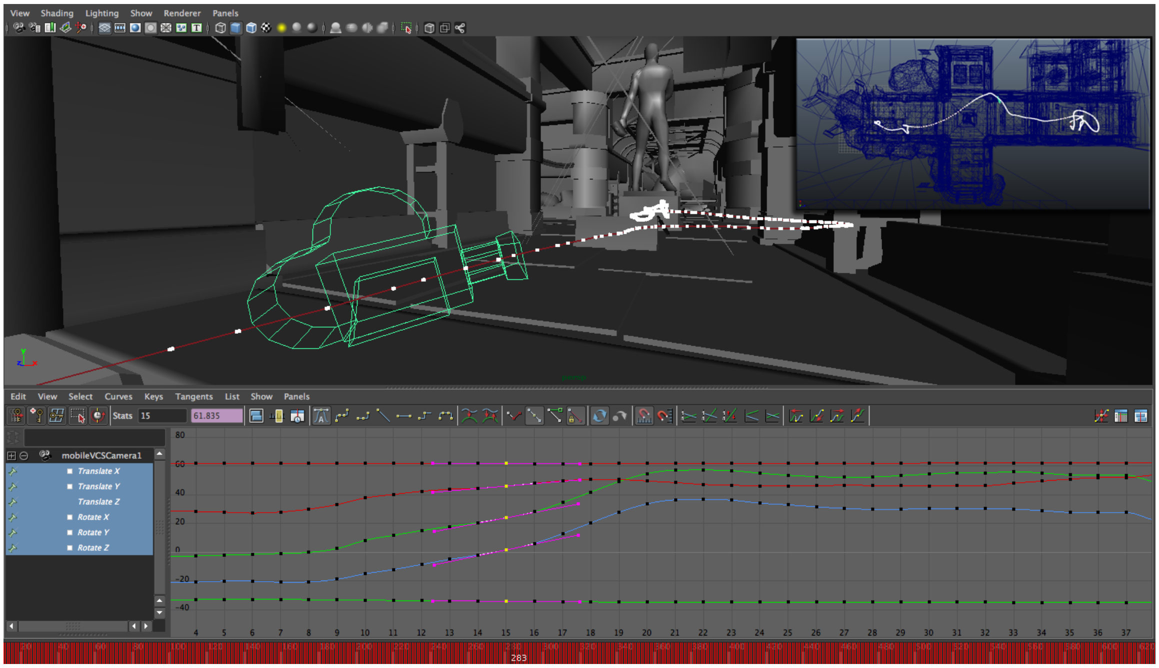


Figure 4.14: SmartVCS captured camera data imported into Autodesk Maya for motion editing

The SmartVCS prototype only supports full recording of the virtual camera’s position and rotation information. All animated objects within the scene, in-scene modifications, and additional camera properties are not recorded. For performance reasons, it was decided that streaming camera position data from the Move and capturing the virtual camera was a priority. As more complex scene assets such as characters and dynamics need to be recorded, the user may experience a performance decrease of the game engine on the tablet, In addition, as needed, the aforementioned properties can be scripted to be added to the recorded text file for external software parsing. In order to support a better two-directional import & export pipeline of camera data between the SmartVCS and desktop software, a parser is required to convert software created animated camera motion to a comma-separated text file that can be loaded back into the SmartVCS app.

This is significant because it now extends the captured data from simply existing on the tablet to existing in the user's preferred animation software. With this edited data refined on the desktop, the SmartVCS can load the new keyframe data into the app for review or further alteration.

4.3.4 Intelligence (Expert System)

As a device that is targeted not only to independent developers but novice cinematographers, the SmartVCS supports true virtual camera intelligence with its knowledge of standard cinematic principles. Game engines, uniquely, provide custom scripting opportunities and built-in intelligence that allow the SmartVCS to become a more intelligent and flexible system. Unlike traditional virtual camera systems, the software intelligence can evolve and adapt based on the user's interaction. Through a series of overlays, presets, and head-up displays, amateur users are able to shoot more interesting camera moves with the SmartVCS.

With its proposed overlay system toggle, the user can receive a ”Smart Grid” over similar to the one-third rule grid displayed on most consumer cameras. Beyond a simple GUI overlay, the Smart Grid is able to identify if you have placed characters or environmental landmarks in a visually interesting composition (Figure 4.16). Since key elements within a scene can be tagged with a certain flag such as ”Main Character” and ”Landmark”, the Smart Grid attempt to align these flagged elements within the camera’s view to an area of the Smart Grid that abides by the laws set by the one-third rule.

The SmartVCS also has an under development lock-on feature, which allows the virtual camera system to intelligently follow a tagged object within the scene. Beyond simply parenting the camera to the object, this intelligent system provides a pre-set list of camera shot types to choose from to best compose the desired shot. These shot types include, but certainly not limited to, wide-shot, medium shot, low/high angle, and close-up. Instead of manually moving the camera about the tagged object, these shot presets provide a starting point to jump start the creative shot compositions for the camera operator. In addition to shot types, the user has the ability to swap between a set of standard camera setups including tripod camera, steadicam, dolly track camera, and crane camera. Each of these constrict the virtual camera in a certain way, like turning off positional motion in tripod mode, to allow the user to focus on composing & shooting the scene.

Smooth Follow Camera Distance: 10

Smooth Follow Camera Height: 5

Playback Speed: Normal

Time 00:13:01

Figure 4.15: Examples of ”Smart Grid” & Camera Path Smoothing

Other examples of educational intelligence within the system that are currently under development include dolly track layout and heads-up display warnings. When entering the dolly track camera mode, the user will be brought to a bird eye view of the scene. In this view, the user is able to use the touch screen to draw a dolly track onto the scene floor. Based on the chosen camera type, the system allows the user to build a dolly track that is based in real-world standards of dolly rails. This prevents the user from building unrealistic paths that would cause the camera to travel on a jarring path. Leveraging the affordances of a game engine such as collision detection, with the heads-up display system, a warning GUI element will appear on the screen when the system notices an issue with the camera such as if it is intersecting with any geometry or moving in an unnatural motion. While traditional virtual cameras utilize fuzzy logic to handle shake, the SmartVCS is able to compensate shaky camera motion with a slider that allows the user to stabilize the motion path.

Future implementations of the systems intelligence will enable the virtual camera to begin making cinematic decisions on behalf of the user. Beyond a simple tutorial system that will evaluate the user's setup, play space, and cinematic expertise, the software could adjust to work with the user. In addition, as previously mentioned, a smart camera will provide recommendations on pre-sets such as accepted camera motions, shot types, etc. giving the user a starting point in which to compose a shot. The goal is for the intelligence designed to aid in the creative process of the user. This grander vision, inspired by the research of the Virtual Cinematographer, will automatically generate intelligent camera compositions and motion paths as well as adjust to the user's spatial limitations as they shoot.

The key differentiator of this system is the ability to layer on different levels of intelligence. Traditional systems lack this level of intelligence and resort to simply replicating just one-to-one motions captured in the motion capture volume in the 3D environment. SmartVCS's intelligence, as a proof of concept, takes this a step further by providing a toolset for novice cinematographers to take advantage and learn cinematic principles from.

With the use of a game engine in the SmartVCS, intelligence can be scripted to provide guidance and feedback to the user, offering them suggestions for improving their traditionally recorded camera motions.

4.3.5 Collaboration

While the system on its own is functional as an intelligent virtual camera system, being a consumer device aimed at a indie market, developing unique features to encourage collaboration among small production teams was key. The decision to include collaboration as a pillar of the system was due to observations made with the way cinematographers operate. In production environments, cinematographers often work with directors, layout artists, and supervisors when developing shots. In order to support this level of multi-user interaction, the SmartVCS supports three types of collaborative environments - a multi-camera, directorial, and networked collaboration.

Multi-Camera Collaboration

In its current form, two SmartVCS systems are supported for simultaneous multi-camera control. This opens up possibilities of having two camera operators working together to shoot a single sequence. In addition, this encourages greater experimentation and collaborative exploration within the virtual world coming from two operators who can work together to complete a shot. In a dynamic action sequence, for example, cinematographers will have the luxury of testing out multiple angles of a scene within a single take through this system. Together, the two operators can more effectively review many shots and save the best takes for post-motion editing.

Directorial Collaboration

With the Playstation's support for multiple Move controllers, this opens up possibilities for other controllers to act as characters, props, lights, or elements within a scene. Much like how James Cameron used the Simul-Cam to remotely capture the motion capture performance occurring on an independent stage, SmartVCS operators can use the camera independent to the function given to other Move controllers.

For example, two actors can play the role of fighter jets flying through the air, each controlling a Move controller and using the controller's trigger and face buttons to fire an array of weapons. The jet models would be parented to each controller independently. While the two actors are choreographing and performing a dogfight scene, the SmartVCS operator can record the scene and direct the performance of the two prop actors. This collaborative experience is unique to this particular system and was previously reserved to larger studios with capture volumes that can fit both actors and VCS operators.

Networked Collaboration

The SmartVCS is already a networked device, connecting and streaming data directly with the Move.Me server. In addition to receiving data, the system has the ability to stream content directly from a desktop PC running the Unity Game Engine. Since the SmartVCS was built on top of the Unity engine, it is able to leverage the myriad of networking functionality supported by the platform and optimized for mobile development. What this enables is the ability for content within a scene to be streamed in real-time with what is displayed on the iPad. This allows for a more familiar production environment to what already exists within motion capture spaces where camera operators are separate from the technicians. Camera operators can call out instructions to the technician to, for example, add additional assets into the level or modify the lighting, all using the familiar mouse and keyboard controls within the game engine interface.

How SmartVCS differs from traditional AR systems is two-fold: on the graphical performance front and from remote collaboration. Since the system streaming data to the iPad is a completely independent system, that PC has the ability to utilize all of the graphics horsepower of a desktop to render more advanced scene assets & lighting all within the game engine. This tiered approach allows the SmartVCS to run at a high framerate and low latency with lower resolution scene assets, while the PC can adjust the quality settings of assets based on the system's specifications.

One key feature that would benefit from this is real-time depth of field. Early experiments of tap-to-focus depth of field modification on mobile proved to lack in realtime performance with camera motion. Since rendering depth of field is a very processor intensive feature, networked collaboration could be utilized to stream depth information to a more powerful desktop machine for the real-time processing. This would allow the SmartVCS to act as a ”focus puller” or first assistant camera operator and lock the focus to an object within the scene, even without the live display preview. The data could then be recorded as a locator and used to drive the depth of field when streamed or exported to a desktop.

Figure 4.16: Modifying scene elements across a wireless network

Another benefit of streaming with an independent PC system is the potential for remote collaboration with a technician. Since all the data is being streamed, access to the information required connection to a direct IP address, provided by the camera operator. With this information, a remote computer running the Unity client of the scene is be able to have access and work with the camera operator to help film the shot. In this scenario, a camera operator could be working on a remote studio and a director, who may be in another country filming, will be able to remote connect to the system and view the camera movies and make scene modifications on the fly (Figure 4.17).

While only the aforementioned networking techniques have been explored through the development of this system, SmartVCS was designed to be an open camera capture platform for developers to build an infrastructure that works for their particular studio environment.

4.4 Application

Through this section, we will explore the potential cinematography applications of the SmartVCS from basic cinematic composition to advanced applications such as augmented reality and match moving.

4.4.1 Virtual Composition & Shooting

The core function of the SmartVCS is to allow cinematographers, regardless of expertise, a tool to develop their desired shot just as they envisioned it. The following discusses the process in which users of the system will be able to compose and record their scenes solely on the device.

Users begin with loading the desired scene within the SmartVCS iOS app and connecting to the Move.Me server. With the system calibrated, the user is now able to make global modifications to the scene, adding and manipulating the objects, characters, etc. based on the needs of the shot. Once the scene is developed, the user can utilize the freespace and analog input of the system to compose the desired shot. Utilizing the built-in intelligence features such ”Smart Grid” or the recommendation system, dialog options will appear on the display to suggest shot types, lens configurations, and framing of the shot.

With the core shot framed, the user will now perform a ”Range of Motion” to practice the camera motion. In this exercise, the user will, within the tracking volume, perform the physical maneuver required for the shot. In this, the user will be able to make modifications to the tracking sensitivity of the system in order to achieve the desired motion. On-screen messages will appear if the system detects motions that seem ”unrealistic” and make suggestions to best approach the camera motion. Once the move has been finalized and the animation has been cued, the user is now ready to record a take of the shot.

While recording the camera motion, the user still has access to all of the analog controls such as panning and dollying to aid in the final camera motion. After the recorded motion has been completed, the user enters the playback review mode. In this mode, the user has the ability to review the take using the playback controls (play, pause, stop, fast-forward, rewind) as well as the ability to scrub across a playbar. If the shot is approved, the user is permitted to save the take within the app. If not, the user can return back to the scene to reshoot the take. There is virtually no limit to the number of takes the user can take, thus encouraging the user to experiment with many shot takes.

4.4.2 Augmented Reality

SmartVCS, being a iPad-centric device, is able to leverage all the AR innovation occurring on the iOS app space. With the iPad’s high resolution camera combined with the tracking capabilities of the Playstation Eye camera and the Unity game engine, new opportunities for markerless augmented reality exist with this new capture platform.

The current state of augmented reality requires a tracking marker to be placed in view of the camera and for software to track the position of the marker and replace it with a layer of 2D or 3D augmentation. While this method is effective, the need to place a physical marker in the environment is a hindrance to the design experience. In addition, latency concerns are common with mobile augmented reality application, often lagging due to having to track the position and render the data onto the screen in real-time. Other solutions that use GPS tracking in combination with a gyroscope to layer on information over a camera is novel, but lacks the freespace positioning accuracy of other devices.

With the SmartVCS, all of the 3D tracking of the device occurs on the Playstation 3, freeing up mobile processor space for advanced rendering and interaction. In addition, there is no longer a need for a tracking marker due to the system knowing the exact position and orientation of the device in real-time and in real world space. Possible applications of augmented reality on such an intelligent device include 3D Integration into Live Video and 3D Tracking Information in Post.

Another possible extension of augmented reality on the SmartVCS is the integration of traditional AR marker recognition and manipulation to affect objects in a 3D scene. In this scenario, the iPad camera would capture markers placed in the real world. These markers would represent 3D assets within the virtual world. As these markers were moved, the camera will be able to track the marker's motion and translate the virtual asset respectfully on the iPad display. With this, the user is able to lay out a virtual set with characters as tangible markers in the real world, only to be able to easily move each around to support iterating on a camera shot. - With data being streamed and recorded directly onto the iPad, an opportunity exists to overlay aspects of the virtual world within the real-world. An option on the SmartVCS interface can enable a video overlay of the 3D world over-top a live stream of the camera output on the iPad (Figure 4.18). With this, the user is able to tell exactly how far to move in order to complete the shot. This level of immersion marries the real and virtual world in a unique way to aid virtual cinematographers to best maneuver about the world.

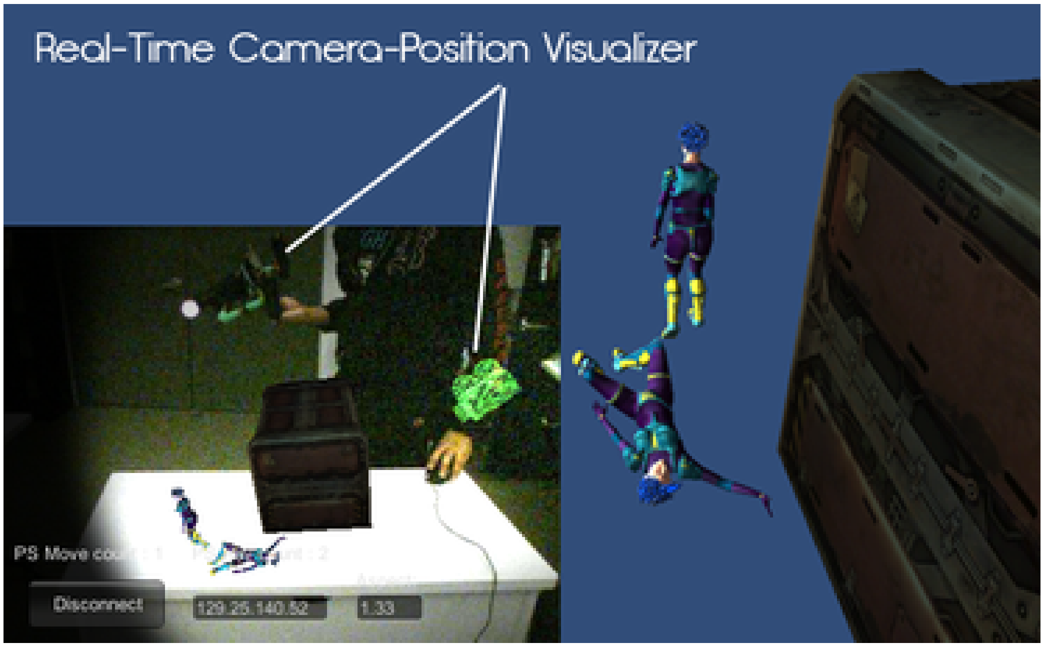


Figure 4.17: 3D Scene Assets and Camera Visualizer overlaid on top of a Live Camera View from the Playstation Eye

4.4.3 Match Move

With the growth of iPad videography as a popular mode of capture, the SmartVCS supports the ability to act as a virtual camera for a Match Move solve. Since camera data is captured in real-time in conjunction with the video stream, cinematographers can use the SmartVCS as a 3D Tracking Tool, similar to the Zeus platform developed by Zoic Studios. With this camera information, the SmartVCS removes the need for any camera tracking/matchmoving software and operates similar to the several-thousand dollar motion controlled cinema camera rigs.

Considering the camera data includes position, rotation, focal length, etc., bringing this data into any compositing software will create a precise track of 3D content within a video back plate recorded from the tablet. With the rise of mobile developed films, the integration of tracking technology will enable a new, accessible level of 3D production previously reserved to Match Move Artists.

4.5 Comparison to Existing Solutions

The following is a comparison of the systems discussed in Section 2 to the SmartVCS. I have chosen six key aspects that capture how my SmartVCS is different from the other existing systems developed:

Intelligence: To have a system-built understanding of shot composition, cinematic principles, and film language. Techniques such as line of sight and one-third rule and camera terminology such as film-back, aspect ratio, and lens type are both aspects of intelligence expected in such a system.

Editor: To have the ability to make real-time modifications to scene and camera properties. This could include the ability to import models and animation cycles form an external program, draw out camera tracks before recording, and build basic scene scenarios using a set of pre-built assets for rapid experimentation.

Game Engine: To use a game engine to facilitate in the real-time aspects of the film-making process. Leveraging fast computational and graphical systems often attributed to games to make cinematic decisions on lighting, object placement in scene, and camera lens attributes.

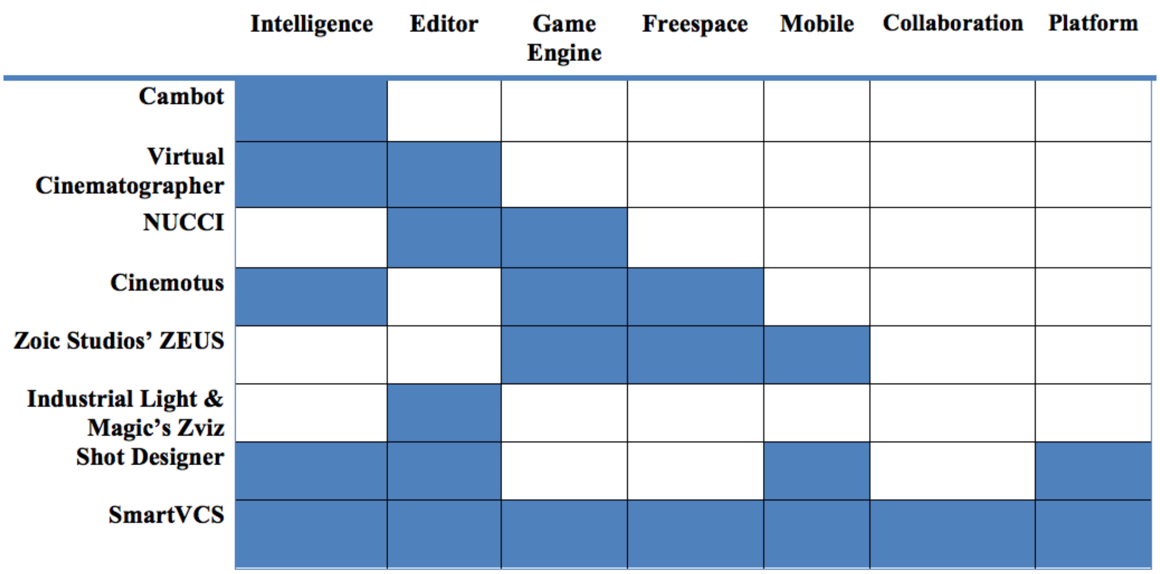
Freespace: The ability to move in full six-degrees of directional & rotational freedom. This enables the camera to operate in a capture volume allowing true one-to-one tracking of the system within a computer.

Mobile: To use the latest mobile hardware solutions to develop new touch interfaces for cinematic purposes. Some advantages of utilizing such hardware include portability, network compatibility, dynamic user interface design, and mass market availability.

Collaboration: Taking advantage of the networking of multiple devices and hardware, the system could extend to be utilized by multiple members of a team. This fosters creativity and encourages the iterative pipeline supported by the virtual production ideology.

Platform: To develop tools that exist as an open framework that can be built upon by developers. No longer does a virtual camera system's software remain static. Every piece of cinematographic functionality described can be modified and updated based on production needs.

The SmartVCS leverages all six of the aforementioned properties in a new hybrid user interface. The system is built on an iPad using the Unity game engine to enable freespace motion via the Playstation Move tracking system. In addition, the interface of the system includes an editor to build basic scenes from a pool of premade assets and plot out camera moves that abide by film conventions. Based on the research, no system developed thus far is able to capture all these six aspects in one harmonious device. By leveraging the latest consumer technologies, it is possible to build such a rapid previsualization platform that could target a market of both professional and amateur filmmakers.



	Intelligence	**Editor**	**Game Engine**	**Freespace**	**Mobile**	**Collaboration**	**Platform**
Cambot	■						
Virtual Cinematographer	■	■					
NUCCI		■	■				
Cinemotus	■		■	■			
Zoic Studios' ZEUS			■	■	■		
Industrial Light & Magic's Zviz		■					
Shot Designer	■	■			■		■
SmartVCS	■	■	■	■	■	■	■

Figure 4.18: An evaluation of the SmartVCS in comparison to existing virtual camera systems

4.6 Known Limitations & Future Development

While several of the aforementioned features offer a grand vision for the current capabilities of the SmartVCS, they currently exist in prototype form to establish the possibilities of this platform. That being said, in its current form and with the constraints of the current state of consumer hardware and software, there are certain limitations of the system that need to be addressed.

4.6.1 Tracking Volume

One of the major constraints of using the Playstation Eye system is the maximum tracking volume achieved with the camera. As previously mentioned, an area of 10 ft by 12 ft is adequate for limited physical camera motions. Along with this space issue, is the limitation of having the hybrid system always be positioned in an orientation seen by the Eye camera.

While the SmartVCS has taken these issues into consideration with its inclusion of tracking sensitivity, fallback gyroscopic controls when the tracking signal is lost, and camera rotation functionality on the analog joystick, this is something that could be improved with the use of more cameras covering the tracking volume.

Companies such as iPi Soft offer markerless motion capture solutions that allow up to six Eye cameras to be used in conjunction to expand the volume. In the future, one can expect more camera-less tracking solutions such as the Leap Motion Controller, local GPS and wireless magnetic motion sensing to provide effective alternatives. While systems like these could be integrated within the SmartVCS platform, the design of the system in its current form was really to challenge what could be accomplished with the bare bones of accessible consumer technology.

4.6.2 Processing Power

With every generation of tablet hardware and game engine update comes a significant boost in graphical processor power. At the time of writing, the iPad 3, which was used in the production of the SmartVCS, was the first Retina display tablet on the market to support the Unity game engine. Running a game engine at 2K resolution offers several benefits in terms of clarity and precision control, but does take a significant performance toll on the device. With every update in the Unity software, optimization is made to better run game content onto the mobile hardware. In addition, mobile authoring techniques for optimization are key considerations to be made when developing game engine ready content for this platform.

4.6.3 Form Factor

With the advent of the Nintendo Wii U tablet controller, Microsoft SmartGlass, and the motion-tracking Playstation 4 controller, it seems that the industry is headed in a direction where touch screen devices and tracking technologies are the future of gaming interfaces. The SmartVCS was a hybrid interface that combined touch and analog control, something the new Wii U controller embraces. Similarly, the Playstation 4 controller combines analog control with the same tracking technology found in the Playstation Move. While disparate technologies in their current consumer form, one can anticipate that a singular multi-touch tablet device with analog input and motion-tracking technology exist in the future.

4.6.4 Expansion to Other Fields

With cinematography as the SmartVCS's core function in its current form, it is important to understand that this new input device could be expanded to apply to other fields such as:

- *Video Games:* Interactive Level Design Creation, Hybrid Freespace-Touch Game

Controller

- *Film:* On-Set Planning Tool, Real-Time Compositing
- *Engineering/Architecture:* Visualization Tool for CAD Models, Machine Input Controller for Robotics
- *Medicine:* Robotic Surgery Input Device, Medical Visualization

Each of these fields open up an area of possibility to leverage affordable freespace input devices and iterative development provided with game engines. Motion capture is currently utilized in all of the above listed fields for research and development. Any field that requires the level of precision, visual feedback, and intelligence could find a use for this system. As a visualization tool, the SmartVCS stands on its own as a functional prototype platform for development using consumer hardware.

5. Implications of Research

The possibility of a consumer based previsualization platform would have tremendous impact within the industry. The film and game industries are already in high demand for tools that will lower the budget of their production. The SmartVCS is a demonstration that such tools can be developed to help expand the industry. By developing intuitive, accessible, and intelligent systems, there exists an opportunity to build a device that will make production more efficient. In addition, the promise of such a system would introduce a new, iterative production pipeline which allows directors and cinematographers to experiment with their shot compositions throughout the production process instead of simply at the previsualization stage. Being an open platform device, developers can begin to build on top of the existing virtual production experience, tailoring the software to their production needs. Lastly, with intelligence at its core, this device has the potential to open up a new market of amateur content makers ranging from students to small-production teams that cannot afford the cost of studio previsualization tools.

The industry is starving for innovation in the production pipeline and my goals in this thesis research are to develop a compelling system that will facilitate the existing pipeline and help push virtual production to a brand new audience. “As the technology continues to become more scalable and more accessible a broader range of creative professionals will able to adopt virtual production techniques.” [2]

Bibliography

[1] Visionary filmmaker douglas trumbull fuels new filmmaking paradigm with nvidia quadro gpus. `http://www.nvidia.com/content/PDF/success-stories/maximus-case-study-douglas-trumbull.pdf`.

[2] The new art of virtual moviemaking. *Autodesk Inc*, 1, 2009.

[3] Cinemotus: A virtual camera solution. `http://www.university-technology.com/opportunity/218/cinemotus-a-virtual-camera-solution`, 2011.

[4] Is realtime real? part 2. `http://www.creativeplanetnetwork.com/node/43448`, 2011.

[5] Hollywood camera work: Shot designer ipad app. `http://www.hollywoodcamerawork.us/sd_index.html`, 2012.

[6] Virtual production in games, 2012.

[7] Alex Billington. A brief look at the technological advancements in james cameron's avatar, 2008.

[8] Wayne Carlson. A critical history of computer graphics & animation - section 17: Virtual reality. `http://design.osu.edu/carlson/history/lesson17.html`.

[9] Mark O. Riedl David K. Elson. *A Lightweight Intelligent Virtual Cinematography System for Machinima Production*. PhD thesis, 2007.

[10] Raffael Dickreuter. Interview with colin green. `http://www.xsibase.com/articles.php?detail=133`, 2007.

[11] Bill Ferster. Idea editing: Previsualization for feature films. `http://www.stagetools.com/previs.htm`, 1998.

[12] Li-wei He, Michael F. Cohen, and David H. Salesin. The virtual cinematographer: a paradigm for automatic real-time camera control and directing. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '96, pages 217–224, New York, NY, USA, 1996. ACM.

[13] Henry Jenkins. Convergence cuture. *New York University Press*, 2006.

[14] S. D. Katz. Charting the stars v.3, 2005.

[15] John C. Goble Neal F. Kassell Ken Hinckley, Randy Pausch. A survey of design issues in spatial input. 1994.

[16] Seth Kendell. Private conversation, 2012.

[17] Gustavo Mercado. *The Filmmaker's Eye: Learning (and Breaking) the Rules of Cinematic Composition*. Elsevier Inc., 2011.

[18] Brian O'Neill Mark Riedl Michael Nitsche Nicholas Davis, Boyang Li. *Distributed Creative Cognition in Digital Filmmaking*. PhD thesis, 2011.

[19] Michael Nitsche. *Experiments in the Use of Game Technology for Pre-Visualization*. PhD thesis, 2008.

[20] Barbara Robertson. Animation world network: Zviz: Ilm goes interactive with previs. 2006.

[21] Jessie Schell. *The Art of Game Design: A Book of Lenses*. Elsevier/Morgan Kaufmann, 2008.

[22] Steve Silberman. G force. `http://www.wired.com/wired/archive/7.05/lucas.html?pg=1&topic=&topic_set=`, 2009.

[23] Wayne Stables. Virtual production branches out, 2012.

[24] Jim Thacker. Fmx 2012: Where next for virtual production?, 2012.