HOCHSCHULE DER MEDIEN

Masterarbeit im Studiengang Elektronische Medien

Virtual Production

Possibilities and Limitations of Virtual Production Environments
Optimization through Implementation of Innovative Interfaces

vorgelegt von

Kai Götz
(26143)

an der Hochschule der Medien Stuttgart

am 26. Februar 2015

Erstprüferin:	Prof. Katja Schmid
Zweitprüfer:	Dipl. Med. Sys. Wiss. Simon Spielmann

Götz, Kai

Virtual Production

Possibilities and Limitations of Virtual Production Environments

Optimization through Implementation of Innovative Interfaces

Electronic Media Master

Media University Stuttgart

Eidesstattliche Erklärung

Hiermit versichere ich, Kai Götz, an Eides Statt, dass ich die vorliegende Masterarbeit mit dem Titel: „Virtual Production – Possibilities and Limitations of Virtual Production Environments, Optimization through Implementation of Innovative Interfaces“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der eidesstattlichen Versicherung und die prüfungsrechtlichen Folgen (§ 19 Abs. 2 Master-SPO der HdM) sowie die strafrechtlichen Folgen (gem. § 156 StGB) einer unrichtigen oder unvollständigen eidesstattlichen Versicherung zur Kenntnis genommen.

Stuttgart, 26. Februar 2015 ______________________________

Abstract

Over the last decades, the process of filmmaking has always been subject to a constant virtualization, resulting in empty stages that leave the entire on-set crew uninformed as real props are more and more replaced with virtual elements. With the development of the virtual production workflow, solutions have been introduced which enable the decision-makers to take a look into an augmented reality, where computer-generated characters are previewed in combination with live action footage and virtual set extension elements. This paper provides an overview of well-established technologies applied in virtual production environments and exposes advantages, challenges and shortcomings related to these approaches. When working on a virtual production set, it is often necessary to modify characters and assets in real-time. However, the existing tools show serious deficits in terms of intuitivity, especially when used by filmmakers without detailed knowledge of computer graphics and 3D software. Therefore a novel 3D user interface is developed, consisting of an Oculus Rift head-mounted display and a Leap Motion gesture-recognition controller. The resulting prototype aims at a more intuitive human-computer interaction and has already been applied and evaluated in an experimental production at Filmakademie Baden-Württemberg.

Keywords: virtual production, visual effects, 3D user interface, real-time

German Summary

Die Filmindustrie ist seit Jahren einer stetigen Umwälzung unterworfen, insbesondere durch die zunehmende Verwendung digitaler Aufnahme- und Postproduktionsverfahren. Visuelle Effekte werden dabei längst nicht mehr ausschließlich für jene komplexen Einstellungen herangezogen, die mit analogen Mitteln entweder schlichtweg unmöglich oder nur mit einem allzu hohen finanziellen Aufwand umzusetzen wären, sondern treten immer häufiger auch anstelle vormals traditionell produzierter Inhalte. Wird jedoch auf Schauspieler, Ausstattungsobjekte und Kulissen weitgehend verzichtet, um die Szenerien erst in der Postproduktion mit virtuellen Charakteren und Objekten zu versehen, bleiben die Filmemacher am Set ohne reale Anhaltpunkte zurück, ein Zustand, der jeglichem kreativen Arbeiten zuwiderläuft. Virtuelle Produktionsumgebungen stellen einen vielversprechenden Versuch dar, die computergenerierten Bildelemente schon am Filmset für den gesamten Produktionsstab, insbesondere aber für den Regisseur und den Kameramann, sichtbar zu machen. Die vorliegende Arbeit beleuchtet die Notwendigkeit virtueller Produktionsmethoden und stellt verschiedenartige Technologien vor, ohne die eine solche Filmproduktion nicht denkbar wäre. Weiterhin werden Vorteile, Herausforderungen und Limitationen bereits bestehender Ansätze aufgezeigt. In erster Linie fehlt es nach wie vor an intuitiven Lösungen um die Fülle an virtuellen Elementen direkt am Filmset nach Bedarf anpassen und optimieren zu können, ohne sich zuvor spezialisierte Kenntnisse in der Bedienung von 3D Software aneignen zu müssen. Aus diesem Grund befasst sich die Arbeit nicht nur mit den rein theoretischen Implikationen des virtuellen Filmschaffens, sondern widmet sich auch der Entwicklung einer eigenen prototypischen Produktionsumgebung, die neben mittlerweile recht etablierter Techniken auch neue Geräte und Schnittstellen für das Editieren von virtuellen Setelementen in Echtzeit umfasst. Das resultierende System wurde unter realen Bedingungen in einem dreitägigen Filmdreh an der Filmakademie Baden-Württemberg eingehend getestet, wodurch in einem abschließenden Teil der Arbeit fundierte Aussagen über den Mehrwert und die Defizite einer solcher Produktionsumgebung getroffen werden können.

Schlagwörter: virtuelle Produktion, visuelle Effekte, 3D User Interface, Echtzeit

List of Abbreviations

VES	Visual Effects Society
FAAI	Institute of Animation at Filmakademie Baden-Württemberg
VP	virtual production
VFX	visual effects
HdM	Media University Stuttgart
VCS	virtual camera system
DOP	director of photography
CG	computer graphics
CGI	computer generated imagery
mocap	motion capturing
DOF	degrees of freedom
GUI	graphical user interface
HMD	head mounted display
DCC	digital content creation
fps	frames per second
ToF	time of flight
GPU	graphics processing unit
CPU	central processing unit
previs	previsualization
techvis	technical previsualization
DMX	digital multiplex

Outline

Chapter 1

Introduction

According to Marshall McLuhan (1964, p.254) “the movie is not only a supreme expression of mechanism but paradoxically it offers as product the most magical of consumer commodities, namely dreams.” Thus film as medium does not only constitute a fusion of technology and art but, and much more important, enables a spectator to get immersed in a fascinating dream world beyond the constraints of everyday life. Since the invention of the cinématographe by the Lumière brothers in 1895, film has not lost any of its magic until this day while the sources from which the images derive got multiplied by now. Modern techniques allow the filmmakers to create even more perfect and impressive illusions while permanently establishing new methods and approaches.

Despite the ongoing popularity of cinema, the film industry is about to change dramatically. Over the past few years, the business has been dominated by visual effects, a term which relates to “any imagery created, altered or enhanced for a film, or other moving media, that cannot be accomplished during live action shooting” (Fink 2014, p.1). Today, VFX facilities all over the world are stuck in a crisis, mainly because an ever-increasing extent of computer-generated content forces them to dabble in low paid mass-production. To break free, the departments are in search of alternative methods, which enable them to create stunning effects in an even higher quality, while making procedures also faster and less expensive. Virtual production is widely presumed to be universal remedy, as it covers “the process of capturing live images and compositing them [...] into a scene in real-time” (Fink 2014, p.1), implying that much of the art of visual effects does not take place in postproduction anymore but is done live on set. However, the consequences of such a radical paradigm shift have not yet been adequately analysed.

1.1. Intention and Ambition

Although virtual production is a rapidly growing approach to filmmaking and introduces novel technologies and pipelines, which affect all stages of production, there is still a shortage of studies or surveys that provide a complete overview of related procedures and implications. This paper constitutes an attempt to shed some light on the peculiarities of virtual production tools, including the advantages, challenges and current limitations coming along. A detailed introduction further discusses the origins of this novel filmmaking approach and outlines those artistic and economic shortcomings, associated with more traditional VFX pipelines that led to the present crisis and induced the industry to look for alternative solutions. Moreover, the paper highlights the possibilities to advance the current virtual production workflow by integrating innovative input and output devices for allowing the filmmakers to modify 3D objects, virtual lights and animation interactively on set. In order to get an idea of the actual benefit such an interface might provide to a real production, a functioning system is to be designed, developed and even applied in a scenic project. A small plot is created to mimic the circumstances of a live action shooting, also including the preparation of concept arts, 3D models and storyboards. Like this, it is not only possible to evaluate the novel devices but also to answer the question of whether an entire virtual production or at least certain components can be applied in an advantageous way in a low-budget student project. Despite all attempts to carry out the experiments as close to reality as possible, the interface is not meant to show its potential as a market-ready solution as it constitutes just an initial prototype, which helps to arrive at conclusions that will serve as a basis for further developments within the scope of Project Dreamspace.

1.2. General Framework of Thesis

The present paper has been supported by supervisors from both Media University Stuttgart and Filmakademie Baden-Württemberg, while the development of the interface prototype as well as the related experimental production has been carried into execution partly in line with Project Dreamspace.

Project Dreamspace is initialized and funded by the Seventh Framework Programme of the European Union and brings together seven renowned institutions in order to research, develop and demonstrate tools that enable creative professionals to combine live performance, video and computer-generated imagery in real-time. The consortium is composed of business partners like Foundry[1], NCam[2] and Stargate Germany[3], research institutes such as Filmakademie[4], iMinds[5] and University of Saarland[6] and the performance art group CreW[7] (Dreamspace 2015). In several small experiments and papers the project partners will research on certain topics related to virtual production and develop prototypes to gain important insights that will later be taken into consideration when developing the final demonstrator. This paper is part of the second work package, which is aimed at the development of prototypical virtual production editing tools.

1 https://www.thefoundry.co.uk
2 http://www.ncam-tech.com
3 http://www.stargatestudios.de
4 http://www.filmakademie.de
5 http://www.iminds.be
6 http://www.intel-vci.uni-saarland.de
7 http://www.crewonline.org

Chapter 2

Virtual Production Principles

As virtual production techniques have already been in use for some years now, it is possible to identify common characteristics, technological and artistic, related to the principles of this filmmaking procedure.

2.1. Keynote

The term virtual production or virtual filmmaking basically refers to the application of real-time techniques within the scope of filmmaking (Dunlop 2014, p.304). Besides this vague transcription, the creative professionals have not yet decided on a valid definition. Sebastian Sylwan, former chief technology officer at Weta Digital and member of the VES Founding Board states, that "there's no checklist of things you can go through and say, if you have all of these, this is virtual production." Further he explains that the industry is still in a process of learning and exploration. Consequently "[…] it's hard to come up with a definition that serves all the possible aspects of virtual production […]" (quoted by Thacker 2012). Nevertheless the Virtual Production Committee, a joint initiative assembled by the American Society of Cinematographers, the Art Directors Guild, the Visual Effects Society, the Previsualization Society, and the Producers Guild of America, published a broad definition in 2012, which describes virtual production as "a collaborative and interactive digital filmmaking process which begins with virtual design and digital asset development and continues in an interactive, nonlinear process throughout the production" (Sargeant, Morin, Scheele 2014, p.444). This idea of an overall virtual production space has been originally suggested by Alex McDowell in 2007 (Beck 2014 a, p.74). Nevertheless this definition may still appear somehow elusive.

According to Wayne Stables, VFX supervisor at Weta Digital, virtual production is "about taking all the lessons we've learned about film production and applying them to […] [the] virtual world" (quoted by Thacker 2012). Commonly virtual production is associated with the method of transferring the motion captured movement of an actor to some kind of virtual equivalent, which is then previewed in real-time in combination with live action footage and virtual set extension elements. Furthermore the position and orientation of the camera is measured and assigned to a virtual camera which allows a correctly framed view into the CG environment. Instead of an almost empty soundstage, the creative professionals on set are able to take a look into a virtually augmented reality and explore a world beyond the physical constraints of our everyday surrounding. Alex McDowell (2012 a), award-winning production designer and founder of the 5D Institute, therefore describes the virtual production as “hybrid filmmaking”, bringing both worlds, the real and the virtual, finally together.

This kind of transcription seems to be quite popular, probably due to the success of James Cameron’s ‘Avatar’. Yet it oversimplifies the principles and does not reflect the full complexity and diversity of this ground-breaking approach. A virtual production covers a wide range of disciplines, including previsualization, motion capturing, live action and postvisualization (Dunlop 2014, p.287). In addition to that, virtual production techniques are not only applied when interweaving real footage and CG assets – be it an insert of real actors or objects into the digital world or an integration of digital characters or models into the real world – but appear most beneficial to completely 3D animated films as well, especially when using a motion captured virtual camera to frame the action.

Integration

As the wide definition from the “Virtual Production Committee” already indicates, virtual production does not only cover the work on set, but also all previous and successive production stages, from the first idea to delivery. Considering an ideal pipeline the virtual production methodologies already join in during preproduction, when designers create the visual aesthetics, while both director and DOP decide on a common cinematographic language, forming their vision of the screenplay. Establishing a central asset management system the virtual production ensures a dynamic workflow which relays the progress of each production stage to dependent departments. Stage designs are now scanned, previs assets recycled and lighting setups captured. The virtual production therefore completes the digitalization of the pipeline and, as David Morin (2012), chair of the VES Joint Technology Subcommittee on Virtual Production, states, constitutes the way of filmmaking in the digital age.

Iteration

Whereas the traditional pipeline describes a linear workflow, a virtual production represents a cyclic iterative approach. Filmmaking is a creative process and therefore depends on the freedom to try out and explore things, including the risk to fail miserably. In every production, certain elements remain questionable before they are reviewed and have to prove their value in practice, sometimes requiring improvement or revision until they finally meet the expectations. A virtual production pipeline enables the different departments to consult each other as early as possible, sharing previsualizations, test results and experiences, going through iterations together and bringing every detail to perfection before inducing an expensive realization.

Interactivity

Formerly, the positions and behaviours of virtual assets or characters were planned and set up in preproduction, refusing the possibility to optimize features afterwards on set, besides spontaneous redesigns which corrupted in most cases the entire time schedule for good. Novel virtual production systems are capable of changing attributes of virtual elements on the fly, using real-time techniques and experiences made in the game industry.

Intuitivity

With the technology getting more and more digital and complex the work on a film set appears increasingly alienating to artists with a traditional background. One would definitely not expect Martin Scorsese to control a virtual camera in Maya using mouse and keyboard. A virtual production can provide tools and interfaces which enable the creative professionals to work with familiar behaving devices in a confident and pleasant way, allowing them to concentrate on issues they should be concerned with instead of struggling with technical insufficiencies.

Collaboration

Glenn Derry (2011), VP supervisor of 'Avatar', calls the virtual production a "director-centric" workflow, bringing the power of decision back to the director. However and even more important, the virtualized pipeline also allows a synergetic collaborative cooperation of all departments including parallelized and therefore flexible and fast batch-processed tasks, without which every decision of the director would lack the necessary expertise. Therefore the virtual production constitutes a democratization of the filmmaking process.

Fig.1: Virtual production keywords

2.2. Formation Conditions

Virtual production principles are definitely not a phenomenon that appears from nowhere as they constitute an attempt to approach the striking problems imposed by the virtualization of filmmaking during the past decades.

2.2.1. Shifts in Film Industry

To expose the circumstances in which the development of virtual production approaches seemed necessary, it is important to briefly retrace the evolution of filmmaking in the last decades and understand the leading role of visual effects for the transformation of an entire industry. Filmmaking is a constantly evolving procedure (Patel 2009, p.2). As technologies progress, new devices and techniques are pushed onto the market introducing completely new approaches and allowing undreamed-of creative possibilities.

In the early days of filmmaking effects and creatures were shot directly on set, allowing the team to work simultaneously while factoring the contribution of each other into the further proceeding. The director was able to see the outcome and join the creative process instantly. However this method was limited to special effects which could be accomplished and afforded in the real world.

Rise of the Visual Effects Industry

In the late 1950s the techniques of Ray Harryhausen made it possible to shoot actors and stop-motion creatures in separate passes before merging them afterwards. Suddenly the directors were excluded from the image composition, partly losing control over the look of the final effect (Bennett, Carter 2014, p.3). With the development of non-linear editing systems in the 1980s the digitalization of filmmaking started and radically changed the traditional workflow. Editors were now capable of experimenting with the shots and claimed artistic freedom, while computer based colour grading allowed a belated revision of the overall film aesthetics. At the same time the first full CG animated films came up. Since the entire live action footage was now anyway scanned as digital intermediate the combination of virtual and real images was more than obvious (Patel 2009, p.2). Carl Rosendahl (2012), founder of PDI Dreamworks and pioneer of virtual filmmaking, describes the development of CGI and postproduction as "[…] a fundamental shift in the way films are made […]." Further he explains that the enthusiasts of these days "[…] did this with a passion for inventing, for building and wrestling with the technology until [they] could conform it to [their] will." All of a sudden, literally everything was possible – worlds beyond the laws of physics were ready to be explored. However, due to performance constraints, the process of filmmaking could no longer be accomplished in real-time.

From the moment of first application the amount of shots which required more or less complex computer generated extensions in postproduction increased steadily, while the percentage of virtual elements within on frame continued to rise as well. Since visual effects represented such an important part of most films the weight of production shifted from set to postproduction and from directors to VFX artists, resulting in a minimized participation and responsibility of traditional departments. Especially the director suffered the loss of influence and found himself or herself reduced to a reviewer, having almost no possibility to request changes afterwards because the fabrication of visual effects was just too expensive at that time (Patel 2009, p.3). With computers performing increasingly well and visual effects production getting routine, the decision-makers called for even more VFX shots. Today the majority of films in the box office top ranking counts on large-scale visual effects, including full body CG characters and entirely virtual worlds. Marc Weigert from Uncharted Territory predicts that this trend will go on. “We will have less and less set building and more virtual sets on all of our movies” (quoted by Knop 2014, p. 82). Carl Rosendahl (2012) admits that he is astonished by the fact “[…] that people keep going to the movies at all. But they do because the experience keeps getting bigger and more extreme – which it must do to compete with all our other activities. Movies have to give us magic we can’t experience in any other way. And that’s what visual effects do, and why every major film depends so heavily on what we do. They have to. Go big or go home.” By now the departments on set have got accustomed to almost empty soundstages and pervasive green screens, while previsualization, though primarily devised as kludge for reducing costs, offers the director the opportunity to regain some of his former competences while planning shots and stagings in advance. Nonetheless the on-set production remains isolated from the creation of digital content.

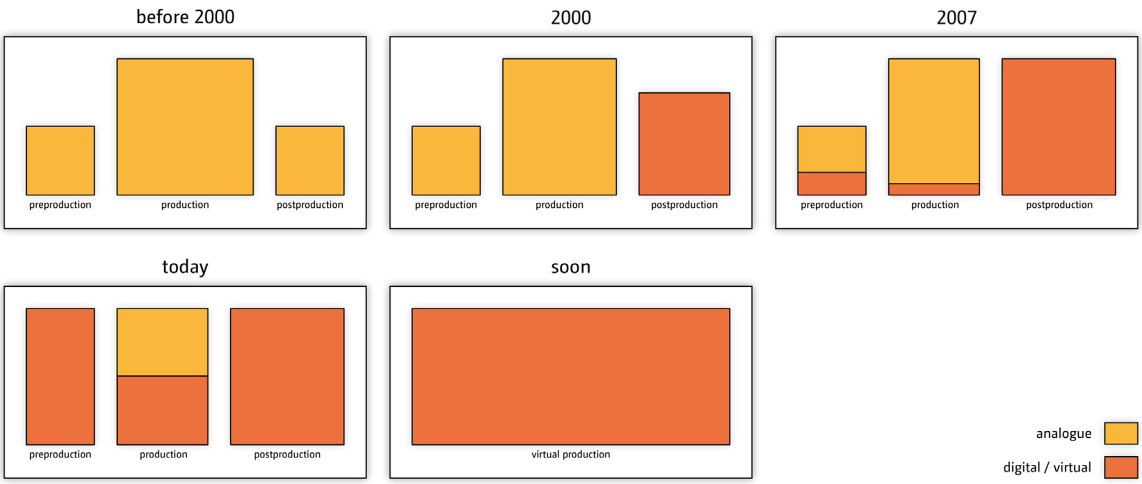


Fig.2: Virtualization of filmmaking

VFX Crisis

With the excessive application of virtual effects the costs rise accordingly. A total budget of 200 million dollars is nowadays in fact not unusual while the expenses for postproduction even outrun those for on-set shooting (Kilkenny 2012). Nevertheless the budgets for visual effects have not continued to rise significantly and cannot do justice to the great effort the postproduction departments make day after day.

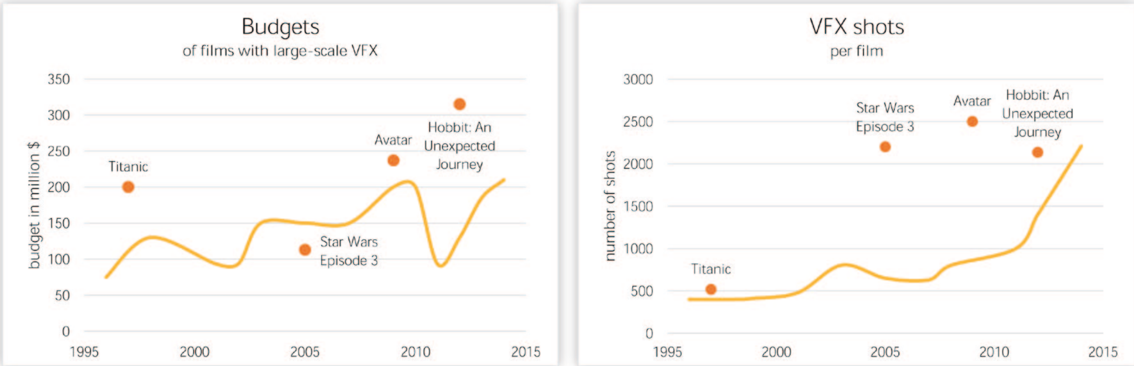


Fig.3: Development of film budgets (left), increasing number of VFX shots (right)

More and more actors and stage elements are now replaced by virtual counterparts, a tendency which does not seem to be artistically motivated at all but only driven by monetary interests. In fact it is much cheaper to create a virtual character than clothing a real performer, sometimes to the disadvantage of authenticity and design. Since the elaboration of high quality visual effects represents the most expensive part of modern filmmaking, the producers try to cut the prices, forcing the studios to operate without standardized contracts and viable bidding practices (Kaufman 2013).

In 2012 the system collapsed. The bankruptcy of Rhythm & Hues is just the most recent and popular case “in a long string of once-illustrious VFX houses that have closed their doors. […] How can Oscar-winning work from a much-admired, long-standing company not protect it and its employees from bankruptcy?” (Kaufman 2013). The circumstances which finally caused the fall of Rhythm & Hues are complex but in no way unique, as the implications of this bankruptcy shattered facilities all over the world in a similar manner.

Debt can be seen as the most serious problem. The facilities are always paid flat-rate regardless of the working hours they might need to finish their work. This fixed bid pricing constitutes an anomaly of the visual effects industry and traces back to an outdated business model born in the 1970s, when some departments, which were formerly associated to big studios, started to build up independent VFX companies. Today declining payments erode the savings of formerly consolidated companies and forces them to operate on thin and unpredictable margins. In addition to that, the directors tend to ask for revisions in a very late stage of postproduction, which appears comprehensible since they have until then nothing to access for reviewing besides green screen footage. The implementation of these change requests, though necessary for the final film, is unpaid. Together with tight schedules and failures or delays introduced by preceding production stages the fixed bid pricing makes it impossible to plan or invest (Williams, Rosendahl et al. 2013, pp.8). The artists and VFX producers do not know what to expect until they are swamped with work.

Moreover, the VFX industry has been built up by great artists and technicians but not by businessmen. Jeffrey Okun, chair of the VES, describes a widely spread way of thinking which incites companies to engage in losing deals: “If my monthly expenses are $100,000 and we have no work lined up, that’s a $100,000 loss. If a job comes along for $50,000, the thinking is that you only lose $50,000 and that’s a win” (quoted by Kaufman 2013). Like this the facilities live from job to job, barely realizing profit or even losing money. The resulting shortage is shifted on to the employees – overages in work are normally not paid.

Not long ago, devices and technologies were so expensive that only the major film studios could afford them in order to engineer customized systems. Today there is no need for developing proprietary solutions anymore since some basic workstation with an adequate render performance totally suffices for getting started, thus limited funds are no longer a show-stopper. By now an innumerable amount of small companies stands by to handle the majority of work while the big VFX facilities address the especially challenging projects which require research and innovation. Jeffrey Okun calls this phenomenon the “commoditization of visual effects” (quoted by Kaufman 2013). The VFX business now suffers from capitalism and globalization as expected. Companies try to maintain their margins and are therefore induced by tax incentives and lower wages to move abroad or found subsidiaries (Kilkenny 2012). Special economic areas, like Vancouver or London, developed to centres of the virtual effects industry, resulting in a highly concentrated cluster which boosts competition und price war. In addition to that the big companies accidentally train their competitors of tomorrow, especially when settling in low-wage regions. With the technology getting less expensive, countries which have formerly been underdeveloped in the field of media production will catch up and overstock the market with highly skilled workers. Already now the imbalance between supply and demand seems striking – prospects are not promising. However the facilities will have to adapt to changing conditions. As globalization is here to stay, the industry has to deal with it (Rosendahl 2012). Since the crash in 2012 the visual effects industry has changed considerably, often not turning to good account. As an effect of the crisis several studios closed or downsized. It is alarming that despite all negative examples the general conditions are still more or less unimproved while the companies stick to low wages and unprofitable bidding. Also Carl Rosendahl (2012) seems to be sceptical about the future of VFX. “[…] It’s disheartening to me to see the industry I love so much behave like it’s dying, when in fact it is just being born.”

Needs for Change

The industry now realizes that the crisis will not be relieved so soon and calls for unions and guilds to be established. Specified market guidelines and agreed wages can constitute however only part of the solution since the production methods also have to change. It is about time to make benefit of the innovative technologies that have emerged in the last few years in order to meet the challenges of a changing market. Since postproduction has been transformed to an uncontrollable money-burning machine the film industry tries to shift tasks back to live action shooting and seeks for novel techniques to combine real and virtual elements faster and less expensive, nonetheless retaining a quality level which satisfies the expectations of cinematic feature films. (Dreamspace DOW 2014, p.3) Beside the opportunity to preview visual effects in real-time on set, a virtual production environment allows the creative professionals to work in smaller teams at lower costs and can therefore comply with the struggle to reduce the overall volume of postproduction.

2.2.2. Downsides of Traditional VFX Production Methods

In a traditional filmmaking pipeline the visual effects creation is settled in the last third of the entire production process, namely in postproduction. Live action footage and plates arrive gradually from set and are instantaneously fed into the processing machinery. 2D artists start to match-move the camera motion, ideally accessing some well-placed tracking markers, while their colleagues set out to remove the green screen by chroma keying the unwanted image areas, followed by manual masking and rotoscoping. If not yet done modellers create 3D assets and characters according to the designs and artworks delivered by the postproduction department. Often the majority of real stage elements has to be rebuilt as CG versions as well to gain the flexibility for later modification and guarantee a seamless transition between real scenery and virtual set extension. After the geometry has been modelled and retopologized it is passed on to the texturing department for painting and further sculpting refinement. Shading and lighting TDs elaborate the look of the virtual scene, defining surface properties and illumination. In parallel, characters are rigged, including setups for facial expression, and subsequently provided with motion capture data from set or keyframe animation. After rendering the 2D artists take over, assuring that the computer generated visuals perfectly mingle with the live action footage.

This workflow seems straight-forward but has unfortunately little in common with reality. Due to their strong dependence on preceding departments and production stages VFX artists have to cope with erroneous data and delays while often not knowing what the decision makers actually expect them to deliver. “It’s like getting in the car to drive somewhere and you don’t know where you’re going” (Fulmer quoted by Leberecht, Storm 2014). Until now, the film industry rarely deploys novel technologies and procedures even if new procedures might simplify the work in postproduction. Often the decision makers are convinced that the traditional production patterns have proved their worth sufficiently and will work in the future as well. New techniques which do not fit the current methods are rapidly dismissed although the established approaches are hardly suited for the challenges of digital filmmaking and fail with increasing frequency in practice while leading to astronomical costs (Knop 2014, p.13).

While the technical drawbacks of traditional VFX methods are often only annoying and time-consuming but do not really affect the overall quality of the final result – there are some incredible good films out there which have been produced the old way – there are some serious creative constraints which actually may compromise the aesthetics of a film.

Creative Constraints

Especially in the field of visual effects a constant supervision would be of utmost importance as the creation of digital images constitutes a process of successive modification and improvement and should be guided by well-versed decision makers. However, in a traditional workflow the creative departments, which are responsible for cinematography, lighting and direction are widely excluded from the production of virtual content. Only a rather small part of the film is visible live on set. Instead of amazing landscapes and sceneries there is a “vacuous green void” (Workman 2014 b), instead of creatures and characters the director sees people in ridiculous mocap suits talking to sticks. It is especially difficult to evaluate the narrative quality when not seeing the outcome of the actors’ physical performances (Mazalek, Nitsche 2007, p.2). Though simplifying the keying process in postproduction, empty monochromatic soundstages leave the film crew disoriented, neither giving a clue about the image composition nor allowing an estimation of light and its effect on the scenery (Knop 2014, p.47). Moreover compositing and asset creation remain in turn isolated from the creative work on set (Dreamspace DOW 2014, p.60). Thus, two production stages that actually depend on each other, namely on-set production and postproduction, do not have the means to communicate sufficiently. It often takes several weeks to prepare the first material which allows an estimation of whether the results from shooting might work or not. By then it is sometimes even too late for a reshoot (Dreamspace DOW 2014, p.45). Furthermore the postproduction departments need some time to perform the revisions in order to comply with change requests. Modern filmmaking has degenerated into an agonizingly slow circuit of evaluation, modification and waiting – hence a process from which creativity suffers seriously.

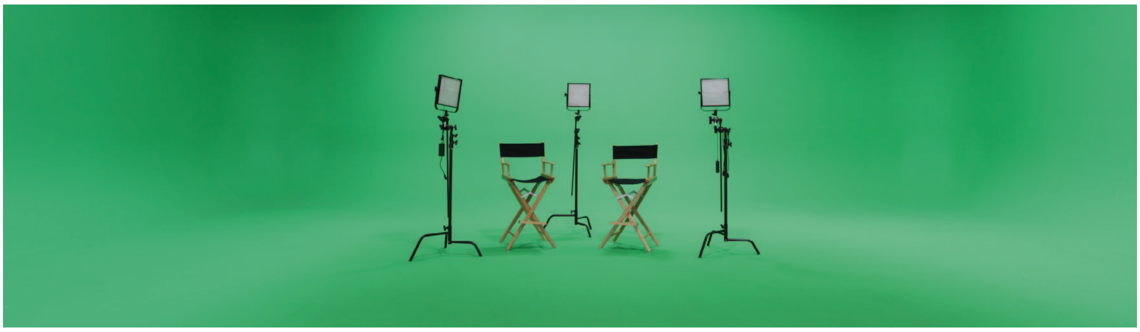


Fig.4: The vacuous green void

In addition to that, all too frequent demands for modification and improvement result in frustration among the CG artists. Nobody would dare to ask the production designers to tear down a stage and build up a completely new version while this procedure seems to be of daily occurrence when working in the visual effects industry (Hughes quoted by Leberecht, Storm 2014). Such a process of iteration might appear acceptable for screenplays or animatics but is completely unreasonable for large-scale virtual elements. However, as soon as some piece of work is digital, the decision makers think that it is also easily alterable (Kilkenny 2012). If mistakes have been made on set, one should agree on a reshoot, which usually costs between 1000 and 2000 dollars per minute, but often the producers tend to shift the responsibility on to postproduction without having to bear additional costs, since the VFX facilities get paid a pre-decided flat rate for each finished shot, no matter how the workload develops during production (Dunlop 2014, p.304). The artists then have to find a satisfying solution – a time-consuming and thankless task. This could be avoided by giving the decision-makers the opportunity to see already on set what they are going to get in postproduction. When sticking to the recent fix-it-in-post attitude, the artists lack time for elaborating designs but become occupied by an uncreative mass-production.

Pipeline Imbalance

Since the beginning of filmmaking the workflow has not changed considerably, still consisting of the traditional tripartition of preproduction, on-set shooting and postproduction. Scripts, designs, storyboards and previsualizations are done in the first stage, trying to work out the details of the director's vision as far as possible before entering the set. From the very first day of shooting the majority of artists and creative professionals which have been involved in preproduction basically loses influence on the project, while director and DOP carry on with staging shot after shot, sticking as close as possible to the preassigned schedule and planning. During shooting or shortly after the material is edited and finally handed over to the postproduction departments which start to elaborate 2D and 3D assets, visual effects and compositions. Again the decision-makers of precedent stages are barely involved.

This rigid production pipeline appears cumbersome, restrictive and somehow daunting, at least from a VFX artist's point of view (Patel 2009, p.50). The traditional communication channels are far too slow throughout the entire production while discussions about important issues often lack a common knowledge background, especially when the director is not a visionary and versed in the fields of visual effects. Furthermore the existing workflow cannot meet the requirements of a modern economically optimized production anymore and forces the filmmakers to "start shooting without really knowing what act three is going to be" (Hughes quoted by Leberecht, Storm 2014). If then visions or settings vary, the VFX departments cannot react on change requests appropriately, facing frustrating revisions and higher workload. Even if the artists are able to create the animation of cameras and characters on the basis of a thoroughly fleshed out previsualization, which has hopefully been approved by the director in advance, they will not escape being forced to modify their work over and over again. Perhaps the decision maker is experienced enough to suspect the need for improvement before hours and days have been invested to render the various image components, but usually the mistakes and flaws get first apparent when reviewing a slap-composed combination of all render passes. Walt Jones, former CG supervisor at Rhythm and Hues describes this frustrating procedure: "we do that work [...] and present it again. [...] They give their feedback. Then it goes back through the chain [...] and we find out whether or not we are done or have to keep going" (quoted by Leberecht, Storm 2014). This imposition gets even more burdening when simulation is involved. VFX artist Dave Rand remembers an absurd situation during the production of Life of Pi, when a shot got "shown to the client who [said] something like why is it even raining in this shot. It's not supposed to be raining [...]" (quoted by Leberecht, Storm 2014). In case the departments can fortunately access a previsualization, the footage from the shooting might nevertheless be suboptimal or unsuitable at worst. In fact a previs might result in an even more unproductive workflow on set when the director assumes that everything will turn out as it was planned in advance without having prepared an alternative solution to react on altering conditions rapidly (Kilkenny 2012).

Furthermore the recycling of assets from preproduction is not scheduled at all which might not surprise as the pipeline lacks standardization anyway. Hand-drawn storyboards and crafted sceneries always seem to be a relic of a quite analogue and hermetic filmmaking process but in fact all departments remain isolated to some degree. Like this 3D models and animations created during previsualization are usually not handed over to production or postproduction, resulting in unnecessarily repeated work.

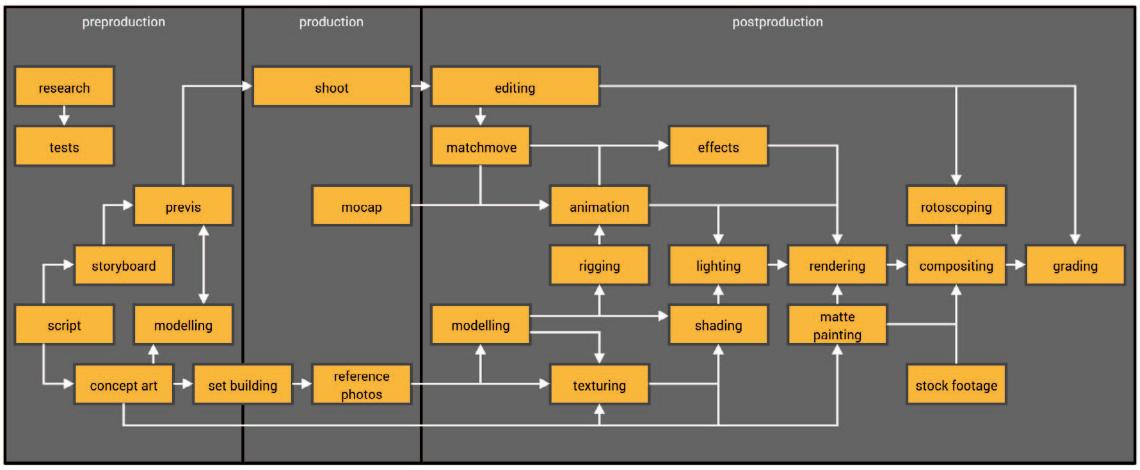


Fig.5: Traditional pipeline

2.2.3. Origins of Real-Time Filmmaking

Since video games and films are both modern instruments of storytelling, it is not a surprise that the procedures for creating such advanced visuals have influenced each other. By now, games do not only provide amazingly realistic 3D environments but have also developed a highly cinematic language which sometimes even matches the quality of feature films. In return, filmmakers and broadcasters have started to apply methods that have been originally created in the games industry.

Machinimas

Already in 1996, some players diverted the ego-shooter ‘Quake’ from its intended use by not playing the game to win but rather to create a short film (Mazalek, Nitsche 2007, p.156). The term machinima was born, which derives from the words machine and cinema and describes a computer-based approach for filmmaking within a real-time virtual 3D environment. By definition, this technique is related to an animation technique which enables “game enthusiasts” (Mazalek, Nitsche 2007, p.155) to create their own cinematics by accessing some kind of commercially available game engine. Such software has been optimized for real-time raytracing and therefore provides immediate feedback without the need to invest in high performance and expensive hardware. Nevertheless this cannot save the user of traditional DCC tools from waiting for hours until a final image is rendered for a first review. What you see inside the game engine in real-time is what you get as a result instantly.

In the early days of machinima the filmmakers had only available those characters and sceneries that were already included in the game – a constraint that limited the possibilities of storytelling to the narrative space of the original setting (Mazalek, Nitsche 2007, p.156). Today, several big companies provide specialized editors that have been designed for enabling the user to create high quality cinematics by staging either prefabricated entities or even loading self-made models and animation. Thus machinimas have opened up a field for professionals or amateurs to develop and accomplish filmic productions in an inexpensive and creative way (Johnson, Pettit 2012, p.21).

The film industry has acknowledged machinimas as a serious mean of virtual filmmaking and also adopted techniques that have been first introduced by machinimatographers. ILM for instance used the Unreal Engine as a tool for previsualization (Johnson, Pettit 2012, p.23). In the future virtual production environments will further establish game engines as an essential part of any live action shooting, enhancing the workflow on set with interactive real-time control while blurring the borders between game and film production.

Virtual Studios

In contrast to traditional chroma-keying techniques, the term virtual studio was introduced to describe a collection of various technologies that are required not only to combine live action foreground elements with any desired virtual background in real-time but also to create a virtual camera that synchronizes its parameters with the studio camera permanently (Thomas 2007, p.5). Virtual studios advance the setup known from weather forecasts or news shows, where a map or some kind of infographic is added to the space behind the television presenter, and provide new solutions for TV serial production, entertainment shows and children's broadcasting (Novy 2014, p.933). In addition to that, the technologies designed for virtual studios are applied in the film industry as well and appear of utmost importance for virtual production environments in particular. Consequently the three most relevant components of visual studios are to be briefly introduced below.

To capture foregrounds and surrounding elements separately, the background has to be extracted in real-time, using either chroma-keying, retroreflective cycloramas, ToF cameras or pulsating illumination (Thomas 2007, pp.7).

Furthermore the virtual elements have to be re-rendered 50 times a second from the point of view of a virtual camera that matches the attributes of the real studio camera. Consequently a solid tracking system is needed. Latest developments introduce methods for detecting the position of a camera by interpreting naturally occurring features. The company NCam provides a system that allows broadcasters to augment images of sport events with floating 3D graphics in real-time, even if shooting in open spaces or large stadiums where traditional marker-based tracking solutions would definitely fail.

A real-time mixing of virtual and real content appears especially useful whenever a certain interaction between a person inside the studio and a computer generated element is necessary or when the virtual element constitutes a significant point of interest. In order to guarantee a believable interaction, the presenter needs a visual feedback – a projection which is not visible to the cameras but provides an indication of the positions and movements of correspondent virtual elements. The BBC has developed a technique that interrupts the projection of the feedback synchronously with the oppositely phase-shifted shutter of the cameras (Thomas 2007, p.8).

Last but not least, the visual studio technology constitutes a way of creating background environments that are much bigger than the set buildings a broadcast production can actually afford (Novy 2014, p.933). Considering the increasing extent and quality of TV serials, such cost-saving solutions are in great demand. The Dreamspace partner Stargate has for instance excelled in the capturing and distribution of high resolution video footage for fully virtual studio backlots.

2.2.4. Birthplace of Virtual Production

Virtual production technology had already been applied in TV shows or features films for several years before James Cameron went for producing his legendary 'Avatar'. A virtual camera system based on fiducial ceiling markers was for instance used in 'A.I. – Artificial Intelligence', a science-fiction film written, produced and directed by Steven Spielberg in 2001 (Thomas 2007, p.9). However 'Avatar' is often called the "birthplace of pure virtual production" (Kilkenny quoted by Thacker 2012), mainly because the film brought together all those techniques that are today associated with the novel workflow. 'Avatar' thus highlighted a new way of digital filmmaking.

When Cameron worked on the very first treatment of a project called 'Avatar' in 1995, motion capture had just been developed to the extent that it allowed the measurement of an actor's body movement in a reasonable quality while it was completely impossible to record subtle facial expressions. While the ground-breaking 'Jurassic Park' got by with only 55 CG shots, it was obvious that a film like 'Avatar' would require high quality computer animation throughout the entire film, exceeding the capabilities of the hardware at that time. Hence, the project was put on hold. However, Cameron is not known for being easily discouraged by technological challenges. After the success of 'Titanic' and the breakthrough achieved by Weta Digital while creating the virtual character of Gollum for 'The Lord of the Rings', the work was resumed and the team started with research and development in 2005 (Duncan, Fitzpatrick 2010, pp. 15). John Kilkenny describes that 'Avatar' was actually made "guerrilla style" (Kilkenny 2012), trying out devices and approaches on the fly. In fact the production served as an experimental environment which allowed engineers and creative professionals to put novel solutions into practice.

The most apparent technological advancement of 'Avatar' consists in the development of a camera system which is capable of displaying not only the real world but also the virtual elements that are meant to be integrated later on in postproduction. The simulcam system has actually been designed for a concurrent preview of various video inputs whereas the team among Cameron diverted it from its original use by feeding in completely virtual footage as well. Turning the soundstage into a capture volume and the physical camera into a virtual camera, it was possible to display a composition of live action footage and computer generated assets live on set (Billington 2008). For this purpose an upgraded version of Autodesk's Motion Builder was employed as real-time engine, gathering all image and motion data before rendering the virtual scenes together with the live action footage. Cameron was fully committed to the system: "We have people in flying vehicles, and I can see what is outside the window, fed in, in real time" (quoted by Billington 2008). The camera system which had been developed during the production of 'Avatar' was also applied in 'The Adventures of Tintin', albeit considerably refined. Originally Peter Jackson and Steven Spielberg planned to realize their project as live action film but finally got convinced of the possibilities of virtual cinematography when James Cameron invited them to try the approaches on their own (Giardina 2011).

Since two-thirds of ‘Avatar’ are not built up of any live action but merely consist of virtual elements, an additional interface was needed which provided the required control elements without having the operator to carry a full camera body. Thus a customized system was set up, which used a portable tablet screen to display a view into the virtual world, while the intrinsic parameters of the virtual camera could be adjusted with tangible sticks and buttons. Since all virtual elements had been prepared in advance, the setup enabled Cameron to literally discover the planet of Pandora by walking through a virtual jungle, allowing him to get immersed by a world that could be perceived visually although it did not exist in reality. The virtual scouting became an essential part of Cameron’s workflow on set, as he let himself be inspired by the virtual environment and elaborated cinematographic ideas right away, often without accessing storyboards or previsualizations (Derry 2012).

In addition to that, a layer-based motion capture approach was introduced which enabled the team to record body performances and facial expressions successively. If Cameron was for example unsatisfied with the capture of a dialogue only the facial data needed to be replaced instead of asking the actors to put the suits on again to recapture the entire performance. Afterwards, when reviewing the dailies, only the favoured takes were assigned to the characters to produce a perfect overall result.

Now collaborating closely with Autodesk and Weta Digital, Lightstorm Entertainment announced more virtual production techniques to be introduced with the ‘Avatar’ sequels. Cameron points out that they “[…] have used the knowledge gained from this first experience to clearly define the ideal process and then develop the technology needed to streamline […] [the] workflow” (Giardina 2012).

Fig.6: Scenes from ‘Avatar’

2.3. Status Quo – Recent State

With the development of virtual production environments, various novel technologies have been introduced, enabling the decision-makers to explore a virtually augmented reality on set. However, the related workflows, which have been developed so far, do not only benefit the artistic and creative work, but also come along with challenges and limitations, future approaches will have to meet.

2.3.1. Technology

Most of the innovative approaches and paradigm shifts which led to the style of digital filmmaking now known as virtual production have been essentially driven by technological progress. With hardware getting faster and software working within performance boundaries in an increasingly sophisticated manner, the time for virtual production has now come, just because formerly merely impossible processes are now viable in real-time. Technical wise, the virtual production bases on a combination of motion capturing, compositing, real-time ray tracing and advanced input- and output-devices. Despite each of these procedures not being a newcomer at all, their concerted usage constitutes a potential far beyond the qualities of the particular device. In fact, while the virtual production appears as an entirely new concept, its technologies have already been heavily tested. They proved their reliability in practice and are currently prepared to benefit a new field of application. Even now, the impact of this union cannot be fully estimated.

Even though a virtual camera, which has been created solely within some DCC tool and looks into a virtual world when framing an animated film, can already be considered as a part of virtual production, the idea of displaying a real-time preview of CG-elements on a traditional film-set appears especially appealing, whereas the complexity of such application may vary considerably. Thus, a motion-captured camera filming a green screen background, which is instantly replaced by pre-recorded omnidirectional stock footage, yet represents the most simplified version of a virtual production. Libraries providing these virtual backlots are commonly used in TV-series production, not only for live-preview or daily editing but for final result. Even this basic attempt to show virtual elements on a live action shooting already requires a solid setup of some motion capture system for camera-tracking.

Motion Capturing

Motion capturing is specified as the process of recording the position and movement of objects, cameras or creatures (Knop 2014, p. 30). The resulting tracking data can be applied either directly in real-time to virtual assets and characters or after optimization and purposeful manipulation. When capturing characters, the position of every single joint has to be gathered before the occurring rotations between the limbs are calculated, while for rigid body objects and cameras a sole measurement of both position and orientation suffices. Performance capturing refers to an extended approach, where body movements and facial expressions are recorded concurrently (Root 2014, p.385). Peter Jackson, director of the 'Lord of the Rings' trilogy and mastermind behind 'The Hobbit', describes motion capturing as "[...] an interesting term that I think people misunderstand a lot. They regard it as being quite a mechanical operation, which it's not" (quoted by Giardina 2011). Jackson's films pioneered motion capturing in a quality adequate for recording body and facial performances, giving actors the opportunity to work within a creative environment and make CG characters come alive. An ever increasing quality of feature detection creates even more believable and living characters, one day finally bridging the uncanny valley. Modern filmmaking without motion capturing is hardly imaginable. In virtual production this technique is the central entity of control, delivering real-time capture data of camera motion and character movement, transforming an empty green screen soundstage into appealing scenery. All further steps depend on the quality of tracking.

As motion capturing is now successfully applied for quite a while, various approaches have been developed, each of them with different advantages and limitations, serving specialized requirements.

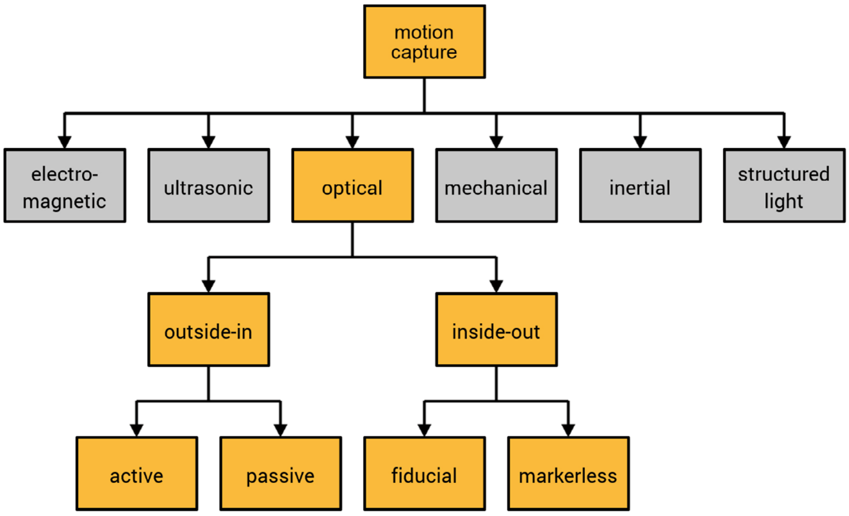


Fig.7: Mocap approaches

Electro-mechanical exoskeletons are the oldest but at the same time most futuristic looking devices for body capturing. A hard-surface structure splints the torso and the limbs of the actor and captures the relative motion in real-time using gyroscopic sensors (Bennett, Carter 2014, p.5). Although the idea of a performer looking like a sci-fi cyborg might initially sound appealing, the skeletons are heavy, fatiguing and too restrictive in terms of freedom of movement. Nevertheless mechanical capture systems are still used in medical research, as there seems to be no need for banalities like wearing comfort.

Inertial motion capture systems constitute an improvement upon exoskeletons, also working with accelerometers and gyroscopes to translate the movement into computable data. With smaller sensors the heavy armature developed to an affordable close-fitting suit, which can even be worn under a costume. Since there is no further equipment, like cameras or transmitters, necessary, the inertial system is easy to operate and deployable in large scale scenes or outdoor shootings. Even if the sensibility of sensors has been increased considerably, there still occurs some noise, especially when capturing subtle movements with small acceleration values. Furthermore the devices deliver inaccurate outcomes as soon as they reach their construction limits, resulting in non-linear and therefore useless data (McSherry, Root, Fischer 2014, pp.391).

Fig.8: Inertial mocap suit from XSens

When using electromagnetic motion capture systems the position of a sensor is identified by its distance to a static transmitter. By analysing the magnetic flux lines of the generated magnetic field, it is possible to draw distinct conclusions on the location of a sensor. Theoretically a set of measuring devices could be placed at each joint of an actor to apply this capturing method for recording an entire performance. Whereas the overall quality seems reasonable, the system suffers from its liability to electric interference from other electric devices or metal obstacles (Bennett, Carter 2014, p.5). Since optical or inertial capturing approaches are much more advanced, electromagnetic capturing only appears applicable with rigid body objects or for tracking cantilevered hand movements. The MIT presented a different system detecting the 3D motion of a body by emitting a Wi-Fi radio signal and interpreting its returning reflection. Still in development, the setup already localizes the centre of a human body with an accuracy of 10 to 20 cm (Adib, Kabelac, Katabi, Miller 2014, p.1).

Another technique uses ultrasonic sound instead of electromagnetic radiation to determine the position of an object. These acoustic capture systems transmit two audio signals with different frequency, while several receivers measure the time of flight and provide an estimation of distances within 1 mm deviation (Sato, Nakamura, Terabayashi, Sugimoto, Hashizume 2010, pp.445). Although the setup is cost-effective and works precisely, the range of capturing falls far short of optical motion capture systems. Furthermore the frequencies of ultrasonic sound are much lower than those of electromagnetic waves and, thus, only allow relatively low capture rates (Jud, Michel 2011, p.15).

A totally different approach consists in scanning the environment with infrared structured light. A certain pattern is projected onto the actor or object. Recording the distortion of the pattern caused by the interfering body, the data can be processed into a 3D model of the scene. This geometry is then evaluated to deduce a proper skeleton. However the method seriously lacks accuracy, while pattern evaluation and skeleton derivation appear computationally intensive and lead to some delay. Nevertheless structured light systems like the Microsoft Kinect may draw some interest, as they represent an inexpensive and compact ready-made alternative to high-end capture devices.

An optical motion capture device basically banks on cameras which detect marker points or image features. Two completely different methods have to be distinguished. Using outside-in motion capture systems, several infrared cameras are built up in the soundstage, pointing inside and covering a certain area, the capture volume. Retroreflective markers on solid objects or performers are optically tracked to triangulate the position (Bennett, Carter 2014, p.5). The quality and steadiness of the capture result depends on the number of cameras and their resolution. When set up correctly the optical motion tracking system delivers high quality capture data and a convenient and reliable workflow. The technique is flexible and can be used for capturing both performances and camera movements. It is able to adapt to extraordinary requests, for example when recording animals. In contrast to reflective markers, self-luminous active markers like infrared LEDs allow an even larger volume, while being brighter and less sensitive to reflections or atmospherics like smoke or fog. Consequently active markers are appropriate for outside shootings. However, the system is expensive. Numerous cameras are necessary to equip a practically sized volume and achieve a satisfying quality. Moreover the infrared system gets quickly compromised by reflecting surfaces like metal tripods, stage elements, liquids or actor's jewellery and fails completely when markers are hidden by occlusion. The preparation in advance including hardware setup, calibration and test measuring requires several hours and trained staff. If the system broke down during the shooting, the entire on-set crew would be forced to wait – an irrecoverable loss of time and money. Inside-out systems are a subcategory of optical motion capture devices and only applicable for camera tracking. Thus they will be examined separately in the virtual camera section below.

Last but not least, facial motion capturing complies with the everlasting ambition of completely recording an actor's performance. While several techniques have been developed, including laser scanners, structured light systems, lights stages and stereo reconstruction technologies, only the optical marker-based capturing is commonly used for real-time application in virtual production environments. In this approach a head mounted camera tracks small reflective or colourful markers on the actor's face, resulting in high accurate data for further rigging and animation (Root, Edwards, Alexander 2014, pp.426). Even if the resolution may not be high enough to recreate a dense point cloud for a 3D mesh of a certain mimic, the feature points depict how facial areas move and stretch, delivering information which can later be adopted to a premade facial rig.

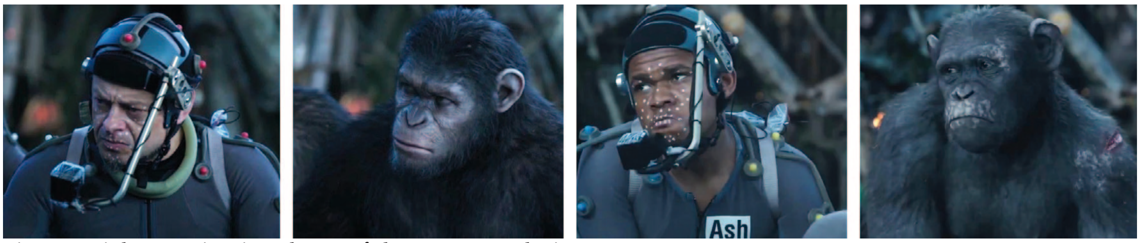


Fig.9: Facial capturing in 'Planet of the Apes Revolution'

As each approach comes along with different advantages, a combination of various techniques sounds appealing. Recent productions at "Institute of Animation" confirm an exclusive application of optical outside-in capture systems to be rather erroneous due to occlusion and erroneous marker detection, whereas a combination of optical or inertial capturing for performances and markerless inside-out tracking for camera movements delivers very satisfying results.

Virtual Camera Systems

Cinematography and camera work are considered as most crucial creative tools in filmmaking. Images convey the story and compose the overall aesthetic. Thus, the process of designing them should be planned and executed carefully. Since sceneries and props are pervasively replaced by virtual assets, at least in large-scale VFX productions, there is almost nothing left to see on a film set anymore, besides omnipresent green screens or grayscale stand-ins. Hence both director and DOP would profit enormously from a concurrent real-time preview of both computer-generated elements and life action footage. Anyway, the traditional VFX workflow still struggles with marker based camera tracking in postproduction even though motion capture or sensor based positioning have already shown their potential in various fields of application. Every standard smart phone or tablet is today capable of measuring its orientation and position in space in real-time with an acceptable precision. Should a million dollar business disclaim the advantages related to these techniques and build on error-prone procedures? In contrast to offline match-moving which constitutes an attempt to reconstruct a real camera by 2D post-processing, a setup is needed which immediately generates a perfectly aligned mimic.

Obviously a virtual camera exists only inside the virtual world. Thus, all parameters of the real camera have to be digitally reconstructed and properly mapped to the attributes of the virtual equivalent, allowing a correctly framed rendering, which can be combined with live action footage in real-time. The virtual camera looks into the virtual world and provides footage for virtually augmented images which are then displayed directly on set. The combination of several image streams, no matter if recorded simultaneously or successively, is often referred to a simulcam. This term was coined in 1997 by the Swiss Federal Institute of Technology in Lausanne and describes a system which synchronizes several video sequences in real-time by applying temporal and spatial alignment techniques (Dartfish, p.2).


Fig.10: Jackson and Spielberg working with a virtual camera system (left), Cameron (right)

In order to measure the location and orientation of a real camera, different approaches have been developed. Traditionally only robotic motion control cranes or dollies offered reproducible and therefore virtually applicable paths of motion. With motion capture devices working more and more reliably, cameras can be located practically at every position within a soundstage, allowing more natural movements freed from setup constraints. Like this even handheld camera work is possible. As already exposed, outside-in optical capturing systems experience some problems with occlusion and misleading reflection. These deficits become even more serious when applied to a virtual camera because even small inaccuracies or jitters emerge highly visible. In contrast, the inside-out method appears most promising. This technique assesses the spatial camera parameters by analysing images from either the camera itself or a second camera mounted on top of the main camera. The latter method, using an additional upward-looking camera, builds on patterns or indexed fiducial markers at the soundstage ceiling. As the installation has been well calibrated the position of the markers in space is known, providing precise information for real-time computing. Logically this setup only works in a specially prepared environment, like a film or TV studio (Thomas 2007, p.6). The latest and definitely most sophisticated attempt uses no markers at all but a real-time detection of natural features within the recorded video footage. Since a complete capture of camera data does not only cover the exact recording of the extrinsic parameters, meaning position and orientation, but also the measurement of all intrinsic parameters – aperture, focus, focal length and sensor size – the emulation of an on-set camera by an operable virtual camera proves to be much more challenging. The British company NCam brought such a system to market, building on sophisticated algorithms for image-based feature tracking and applying sensors which originate from the aerospace industry. The heart of the setup consists in two witness cameras, which are mounted underneath the main camera, capturing greyscale videos at 100 fps. The processing unit detects the naturally appearing features and reconstructs the position of the camera in three-dimensional space. Since only stationary points can be tracked, moving image content is automatically identified and thus discarded in calculation. The attitude sensors optimize the results and helps out in case the image-based tracking fails, for example when facing a too big monochromatic surface. Moreover the delay between the arrival of image and lens data is automatically used for smoothing the tracking results. Besides localisation, the intrinsic parameters need to be measured in order to be later correctly taken into consideration during computing. Therefore every single lens of the principal camera is rectified using dot pattern charts and dedicated software (Dreamspace DOW 2014, p.54). A calibration procedure may take hours for lenses with a dynamic focal length but appears essential, as the position of the virtual camera would otherwise be heavily affected as soon as the lens zooms (Thomas 2007, p.6). Fortunately the entire fine-tuning can be completed before arriving on set. If correctly calibrated, all intrinsic parameters can be captured instantaneously in real-time using camera control systems, like the FI + Z from Preston Cinema Systems[8]. Although this advanced hybrid system seems very compact and delivers robust data, it does not prove to be fully immune to altering contrast and lighting situations, what appears to be a bit of a throwback, since light especially on live action sets is a subject of constant modification. Furthermore the tracking quality depends on the amount of image noise tagging along with the witness cameras. The NCam system is not yet capable of delivering a final track but seems to be totally sufficient for previewing and previsualizing or when serving as a base for further match-moving in postproduction.

8 http://www.prestoncinema.com

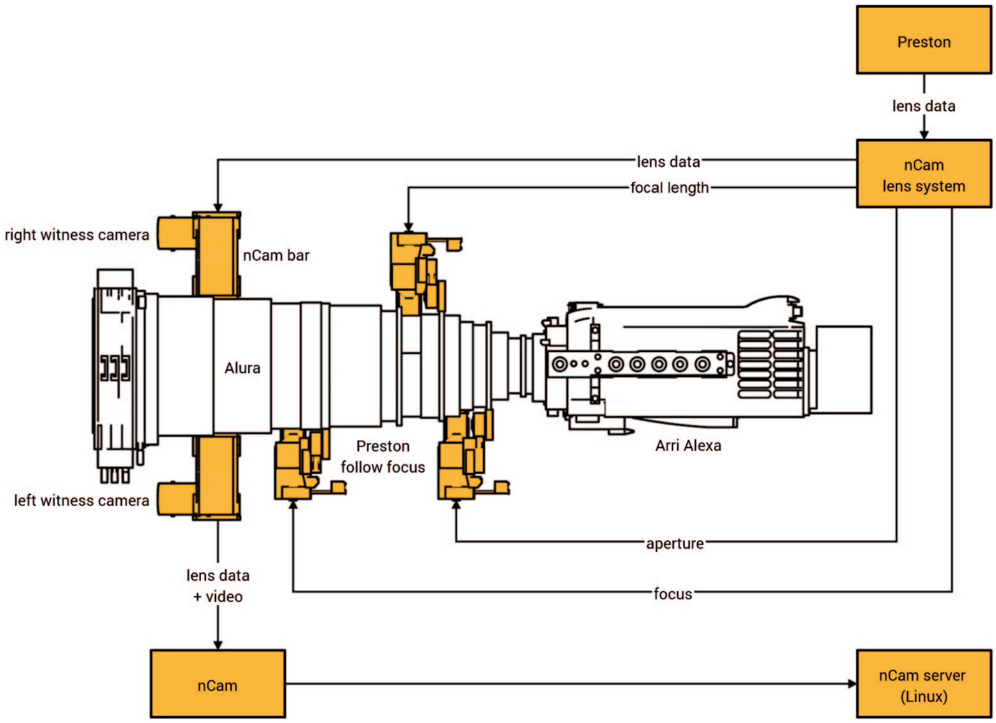


Fig.11: NCam system

After gathering the real camera parameters, all information is transferred to a 3D software package, thus a DCC tools or a game engine. Like that, it is now possible to create a virtual camera perfectly imitating the original camera on set.

A motion captured virtual camera can also be used when there is no live action footage at all but only CG environments and characters. Basically this setup consists of some kind of physical interface device serving as a camera controller, ranging from a simple stick or box with markers to tracked handheld devices or specialized virtual camera systems (Beck 2014 a, p.70). The camera operator can move the virtual camera much like a physical one while adjusting the intrinsic parameters on the fly via some kind of input device, like a standard computer peripheral, a tablet touchscreen interface or a piece of dedicated hardware. Clearly arranged graphical user interfaces allow further customization of camera features, matching the working habits of the operator (Patel 2009, p.8). Since the camera is completely virtual, physical boundaries no longer limit the creative options. Playing around with scaling, the virtual camera may cover a huge distance, while the controller only moves some centimetres within the soundstage. Like this the devices can be used practically anywhere, no matter if the team affords a big soundstage or a small office.

Virtual Lighting

When merging real set elements with virtual assets a homogeneous lighting has to be accomplished in order to bring both worlds in line. Consequently real lamps are expected to influence the CG elements while virtual light sources should affect the actors and props on stage. While the latter still constitutes a rather exotic procedure hardly carried into execution, with just some prototypes allowing a first glimpse on its potential, like the light stage used in 'Gravity', several techniques for capturing and digitalizing real lights are under discussion.

In line with the Dreamspace Project the University of Saarland recently examines a technique for evaluating light setups by shooting photographic light probes. At various spots within the light volume a panoramic vision of the entire surrounding is captured, using a position-tracked DSLR camera. The higher the density of capture positions the better the later quality of the gathered light field. All images are then computed to generate a digital copy of the real light characteristics. Compared to the conventional method with chrome-balls, this novel approach does not only include the position and colour of the light source but also its specific convergence and decline – values, which have to be equally factored when creating the virtual counterparts. The process of capturing light probes can be accelerated significantly when using omnidirectional cameras, as these devices look in each direction simultaneously.

Keying

Keying describes a commonly used process for compositing multiple images on top of each other while removing certain components of the occluding area in order to display the element beneath. First and foremost, this technique is used to replace backgrounds with virtual elements. While keying can traditionally be considered a part of postproduction, besides its usage in broadcasting, virtual production demands high quality keying on set in real-time.

In case the virtual object rests on top of the live action footage, an alpha pass within the CG element is required, which is the type these assets are normally provided anyway. If the digital parts are to appear behind the real images, the relevant region of the real footage is made transparent (Thomas 2007, p.7). Several different techniques are used or still in development.

Chroma keying refers to the most established approach commonly used in postproduction. A particular colour range is picked to get assigned an alpha value representing its opacity. Usually green or blue backdrops are used as the contrast most clearly in hue from human skin (Knop 2014, p.25). The results often fail to live up to expectations, especially when the background is insufficiently or unevenly illuminated, resulting in additional rotoscoping workload for the postproduction team. Considering live chroma keying, the masks cannot be manually optimized and therefore, the green- or blue-screens have to be properly set up. Dealing with limitations of real-time computing, certain vendors provide software solutions which support real-time chroma keying on a standard workstation. Even some game engines allow the user to feed in video footage and apply chroma keying. During a recent production at the Institute of Animation a small segment of a roman bath was actually built up in the green screen soundstage and, while shooting, enhanced with a real-time preview of the virtual set extension. Therefore a basic video mixer put the live action footage on top of the virtual set, applying a chroma key. Even if this method seems somehow outdated, it allowed a first impression of the offline postproduction that followed.

Several renowned companies are right now involved with the development of depth capturing procedures. The idea behind that technology, which is sometimes called the holy grail of VFX industry, is the measuring of depth data from camera perspective, which can be achieved either via multi-camera reconstruction or time of flight evaluation. The former approach builds on the disparity of images, which were taken from slightly different positions whereas the latter method registers the phase-shift of an infrared wave as it returns to its emitter, indicating the distance of the object, which has reflected the signal. The resulting depth maps can be used for various applications but appear especially valuable when employed for depth keying. Instead of being bound to a perfectly set up green screen, all foreground or background elements can now be replaced as long as they differ in depth values and are therefore distinguishable. Until now the resolution of the computed depth maps are far too low while the quality underlies heavy noise artefacts. But as the development of sensors proceeds, one can expect to get usable data within a few years.

Real-Time Compositing

In an ideal virtual production environment almost near to final images are displayed simultaneously on set. As modern VFX productions heavily depend on compositing, this important production phase also has to be somehow integrated into the real-time pipeline. Sophisticated compositing covers various steps – keying, merging, rotoscoping, retiming, colour grading or multipass-tuning to name just a few. Most of these operations are rather hard to compute and yet fully impossible in real-time.

Recent approaches suggest outsourcing computational load from CPU to GPU, parallelising complex calculation and therefore getting a step closer to real-time. The leading compositing platforms, After Effects and Nuke, both provide an environment for real-time compositing. While After Effects builds on plugin integration, Nuke's standard 'blink kernel' node offers a basic C++ scripting editor for rapid image processing (Introduction to Blink Kernels). The self-made kernels are then executed on the GPU. Nevertheless only basic image manipulation, like chroma keying or image repositioning, is feasible until now. The stack of operations to be worked off is limited to a few compositing steps.

Real-Time Engines and Rendering

In order to get a first impression of the far ahead postproduction output, basic models and rudimentary textures totally suffice. However this simplification embezzles important parameters and subverts an immersive and convincing experience which appears crucial for any creative work. Concerning virtual lighting and set dressing, physically correct surface properties, shadows, reflections and refractions have to be accomplished in real-time (Dreamspace DOW 2014, p.59). The quality of decisions made on a virtual production set exclusively depends on the quality of the displayed images. Offline render-engines already achieve photorealistic outcomes but use techniques which are much too slow and not suitable for interaction and live feedback. Even today one single frame may render for hours or days on a farm consisting of thousands of processors (Patel 2009, p.11). In contrast to that, 3D game environments seem to be predestined for application in a virtual production framework.

The development of real-time techniques is mainly driven by the game industry, manufacturing overwhelming artificial and interactive worlds in ever increasing excellence. Real-time 3D game engines bridge the gap between render quality and performance, using sophisticated ray tracing methods to speed up the traversal of rays. For every single frame the entire hierarchy of objects has to be updated and rendered, ensuring an extremely high refreshing frequency, as modern 3D games rely on dynamic content and require especially high frame rates (Dreamspace DOW 2014, p.59).

When used in a virtual production environment, the real-time engine should be capable of handling the data derived from the shooting ideally without any delay. The motion captured movements, performances and positions of all actors and cameras have to be assigned to virtual rigs, while live action footage is keyed and rendered into each frame respectively. Facial capturing videos can either be computed and applied as relative transformation data to facial rigs or projected as basic video textures, so-called Kabukis, onto the lowpoly head geometry. Most of the existing real-time engines manage this kind of data streaming, synchronization and live processing only to a quite rudimentary extent, relying on third-party plugins or user-created solutions. Therefore, big studios build up highly customized environments which indeed base on some available engine, mostly Motion Builder, but offer a bunch of new useful features.

High-end game engines allow artists to work within a most familiar 3D environment with options similar to those of DCC tools (Dunlop 2014, p.305). Crytek for example has released a version of its Cry Engine specialized on virtual filmmaking. This Cinebox features tools for animation and rendering and enables the user to carry out a full-CG film with a single software solution. Vendors of traditional 3D packages, like Autodesk, also extended their series by stand-alone real-time tools.

Even if the engines are now able to handle millions of polygons, high-resolution textures, photorealistic materials, particle simulations and tons of animation layers, they are still optimized for providing joyful gaming experiences and lack important features necessary for virtual productions. In particular, standardized input-output procedures, cross-system propagated genlocks and constant frame rates with minimal delay are recent matters of concern. Furthermore, data from motion capture sessions has to be streamed, applied and rendered in no time. Since creative work always includes playing around with things, pushing the boundaries, the engine must also be able to adapt to changing requirements and thus, provide features for customization (Patel 2009, p.12). Fortunately industry believes in progress. With game engines developing and virtual production meeting with universal approval, the pipeline will continuously upgrade to fit the needs of the new filmmaking methodology.

Display Devices

After combining virtual images and real footage, the composite has to be somehow screened to the creative professionals on set. A reference monitor, ideally properly calibrated, represents the most primitive display device, providing a real-time preview to the director and some selected staff. When shooting stereoscopic one should obviously access a stereoscopic display device. To include more team members a projector may screen the results on some blank surface within the soundstage. While this approach helps supplying additional reference to actors or motion capture performers, it usually does not appear practicable as it obstructs the work of gaffer or grip department. A virtual camera system often constitutes an input and output device at once, providing hardware interfaces for interactive user control and a digital viewfinder or display for immediate visual feedback.

However all these systems are either limited to a small group of spectators or rather immobile. An integration of novel display devices, like smart phones, tablets or head mounted displays, has not yet been achieved.

Technology in Progress

In the future, improved technologies will be pushed onto the market, expanding the possibilities of virtual production and changing the way of filmmaking for good. Right now companies and institutions all over the world put effort into research on new devices and methods, while some ideas are already near completion.

In postproduction digital copies of actors or real set elements are often needed to allow interactions between the real and the virtual world while guaranteeing a coherent integration. Instead of manually modelling or sculpting these replicas, they can simply be scanned using photogrammetry or LIDAR scanners. Like this, the work from the production design department seamlessly blends into the postproduction pipeline, avoiding additional expenses and last but not least cherishing the effort of the stage designers. Moreover, depth capturing is finally possible with ToF or light field cameras, enabling the compositing artists to apply a depth key in order to extract spatially layered image areas. High resolution omnidirectional cameras deliver HDR images for 360° set extensions or virtual light reconstructions, while real-time compositing and ray tracing systems evolve into reliable techniques, even affordable to small budget productions. One day not only directors or DOPs will get to see virtual creatures in real-time when starring at a preview monitor, but even actors will literally face holographic projections of virtual characters on set. Furthermore nowadays apparently everybody carries around a smart phone, so why not use them as service displays? By taking a look at these handheld devices all team members involved in the filmmaking process see the current framing, enhanced with special information dependent on the department they work in. A gaffer can review the lighting from camera perspective without climbing down the platform while a make-up artist may check an actor's face from backstage. When inspecting the virtual location, a head mounted display allows an immersive insight, comparable to a real scouting on a live action set.

2.3.2. Virtual Production Pipeline

Even if a virtual production comes along with a lot of innovative technologies, which alter the way of filmmaking effectively, the workflow remains similar to a traditional pipeline. Most of the departments basically stick to their tasks or face only slightly extended responsibilities (McDowell 2012 b). A virtual production builds on traditional filmmaking techniques, terminology and disciplines – virtual production is still film production (Dunlop 2014, p.287). A detailed description of the well-established production pipeline will thus not be delivered below, while highlighting the noteworthy differences and characteristics instead.

A virtual production constitutes a joint working environment for a broad variety of disciplines, (Dreamspace DOW 2014, p.49) covering all production stages from the very beginning until completion, merging into the entire creative process. Instead of relying on independent and hardly interconnected departments, often assembled by different companies, a virtual production requires "[…] close collaboration between all departments throughout the preproduction and production phases of the film" (Knop 2014, p.77). The traditional linear filmmaking pipeline has been finally overcome, establishing a more cyclic and flexible workflow, blurring the boundaries between departments and production stages (Morin 2012). In addition to that, completely new jobs are created. On film set, specialized technicians build up and calibrate mocap systems and VCS rigs while 3D artists oversee the virtual scenery and software engineers check programmes for proper operation. During pre- and postproduction database developers take care of a user-friendly asset-sharing system, which guarantees a smooth production flow. Whenever new professionals join the team and traditional pipelines change, the existing departments have to realign and accustom to different and possibly unfamiliar habits, accepting the fact that sometimes former unnecessary procedures have to be accomplished before they can resume their work.

Even if virtual production technologies may only be suitable for parts of a certain film production, the new pipeline principles influence all phases. Reacting on creative and technical demands from the very beginning, all changes and created assets have to be migrated and propagated downstream in a most efficient and comfortable manner (Dunlop 2014, p.288). All departments need access to all technical and creative work from every involved collaborator.

Although, as already mentioned, the boundaries between the production stages have been blurred, it appears reasonable to fall back to these subdivisions for managing and clarifying responsibilities and dependencies. However the single phases do not represent any temporal sequence but describe a structure of logical succession.

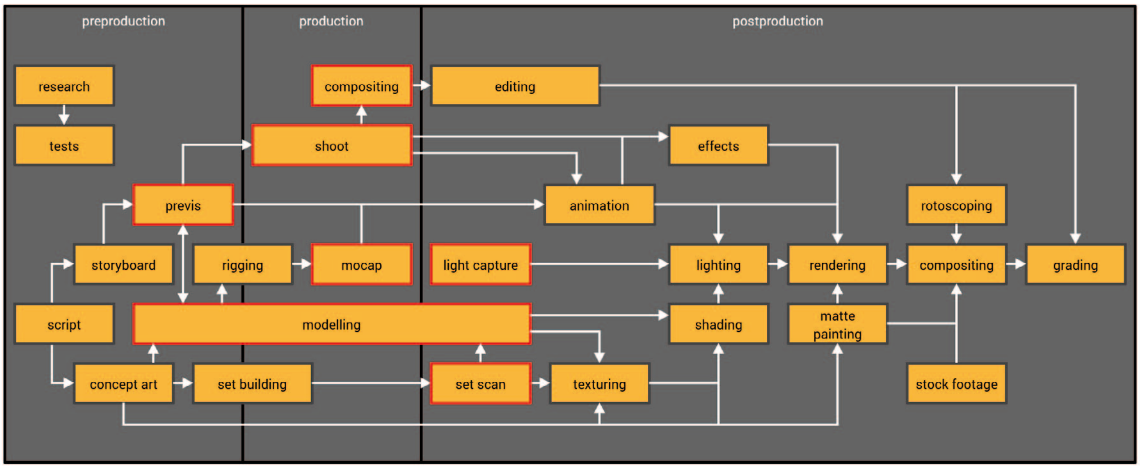


Fig.12: Simplified virtual production pipeline

Preproduction

Similar to the traditional workflow, ideas, screenplays, artworks and key assets are developed in preproduction. However, world building is now fully embedded into the overall pipeline. When creating designs and models, the artists have to keep in mind that their elements will be reused later on by different departments.

Like this, the assets from previsualization find their way into the soundstage. The modelling artists create characters, sceneries and props, which do not only comply with the quality of modern previsualizations but meet the specifications of a later real-time usage, such as polycount, texture resolution and lighting. In an ideal pipeline the assets are immediately handed over to postproduction, where the VFX artists can use them as starting points for high-quality versions or at least as references. In addition to that, set designers profit from the digital asset creation and plan their physical sceneries inside the virtual world while deciding together with the VFX supervisor whether parts should be real or computer-generated. On set the hand-crafted stage is scanned to extract material references and 3D models. In order to guarantee this kind of flexible asset recycling and smooth workflow throughout the entire production, all elements have to be named and tagged according to an overall convention before storing them accessibly (Dunlop 2014, p.290).

In contrast to the traditional pipeline, where the art departments are more or less completely dissolved as soon as the project enters production stage, the virtual production relies on the continuous contribution of these specialized professionals (McDowell 2012 b). Concurrently some VFX artists are involved early in preproduction, mentoring the asset-creators and preventing possible dead ends. Besides delivering breath-taking visuals in postproduction, the VFX team now elaborates strategies and effects which are optimized for an application within the performance constraints of real-time 3D engines and live compositing systems. Likewise they work on preview comps for testing formats and workflows, ensuring that timecodes, meta- and tracking data or camera parameters provide reasonable values and can be correctly interpreted after the shooting.

As soon as the director decided on a cast, character concepts and mocap rigs are adapted to the actors' physiognomy. By consulting a virtual production supervisor, previs and storyboard are broken down to sequences, scenes and shots, taking into consideration the requirements of the virtual production environment and identifying the systems and techniques, which are necessary but not yet developed.

Previsualization

Although mainly settled in preproduction the previsualization spans the entire filmmaking process, while being subject to constant improvement and change. A previs represents a rapid prototype and preliminary version of the film and constitutes an essential part of the virtual production pipeline.

Not long ago, a previs used to be a rough and quickly built slap-composition, primarily used for working out technical details or breaking down certain shots into single plates, which afterwards could be captured separately. This techvis was keyframe-animated, contained only lowpoly models and offered just very basic visuals which were not really convenient for advancing creative decisions. However it was possible to recognize and sort out potential problems early enough. Today a previs is a creative storytelling tool delivering valuable and highly elaborate outcomes while still allowing fast experiments and tests. Often the entire film is nicely processed down to the last cent, anticipating animation, timing and framing. Motion capturing and virtual cameras systems enable the creative professionals to choose an interactive and nearly playful approach, which makes the previs a "really great tool for fleshing out scenes" (Derry 2011). Furthermore it collaboratively associates the departments by visualizing the variety of viewpoints and promoting a mutual understanding (Beck 2014 b, p.46), reveals challenges and complexities which otherwise might have emerged unexpectedly and therefore helps to schedule and manage a virtual production.

When regularly updated throughout the production, a previsualization becomes an iterative process which approximates the final film. While a storyboard, even if well done, always represents a simplification, a previs offers valuable clues about timing and movements, minimalizing the need for unfunded guessing or interpreting. Instead of being limited to a forced perspective the creative professionals work with real camera characteristics and have to get specific much earlier in the production process. Previsualizations approach real filmmaking by offering both director and DOP an existing, albeit virtual, world to wander around, experiment and explore (Knop 2014, p.21).

The visual quality of previsualizations rises continuously, while more and more techniques are therefore transferred from the film set to preproduction. So where does previs end and virtual filmmaking begin? Especially when working on a full CG film or a project that hardly involves real shooting, one may ask whether "previs [is] still previs, or is it making the movie" (Beck 2014 a, p.73). In fact the boundaries fade as both approaches are part of an entangled workflow and run like a common thread through the entire production. Virtual filmmaking extends the previsualization process, bringing preproduction and production together to form a seamless virtual production environment (Patel 2009, p.16).

Production

After months of testing, planning and previsualizing one should expect to arrive on set perfectly prepared. However asset creation and shot optimization have not been brought to a close. Besides being staged and shot, the virtual and physical elements are constantly revised, adapted and repositioned. The changes have to be tracked, recorded and fed into the pipeline, respecting nomenclatures and storing conventions (Dunlop 2014, p.291). If necessary, 3D artists build even entirely new models from scratch right on set, also ensuring that these assets fit properly into the database and appear beneficial to later users.

While 3D modellers and VFX artists now face an enlarged field of activity, the tasks of the DOP also changed considerably. The cinematographer has to participate extensively in preproduction and previsualization in order to sustain his position as image designer (McDowell 2012 b). On a virtual production set the DOP needs to be "[…] well versed in previs, real world photography, editing and colour correction, visual effects and CG lighting" (Workman 2014 b), which might overburden more traditionally trained personnel and necessitates a different type of expertise. Thus camera operators with a competent knowledge of virtual production procedures have to assist the DOP. Even the director has to delegate some of his competences to an assistant director, if either not familiar with new technologies or totally occupied with the actors.

When working within a solely virtual environment some departments fall into disuse. Like this gaffers, best boys, dolly grip, production designers, set dressers, hair stylists, make-up and costume designers, special effects technicians and boom operators are not required anymore while virtual lighting supervisors, software engineers, virtual camera technicians, editors and real-time compositors stand in. The opportunity to work layer based and record camera and performance data separately means a consequential workflow shift to the production pipeline. Now the director is able to work within an almost empty soundstage, concentrating on the acting without being disturbed by the set crew. The performance data and the reference footage of the witness cameras are evaluated and edited, combining the most appealing facial expressions or body captures in a daily master, which is then restaged and framed in a second pass, sometimes long after the actual shooting. Consequently the on-set production is divided into phases and may range over a long period, sometimes even overlapping with postproduction. At the same time this procedure engages a smaller amount of people and simplifies the preceding organization.

Postproduction

After shooting, an enormous amount of data is passed on to the postproduction department. Perhaps this flood of information seems confusing, complex and potentially frightening at first view, but a well prepared and experienced virtual production team is able to keep track of the overall pipeline and guides the filmmakers and artists safely through the entire process. When planned and set up properly, a virtual production environment allows a reasoned and straightforward approach from which all creative professionals involved in the postproduction may profit (Dunlop 2014, p.295).

The footage arrives as preselected files inside a collated database. Mocap performances, rigs, facial captures, set and character scans, reference videos and audio files are already correctly edited and transferred, including EDLs and meta data, and can seamlessly blend into the standard visual effects and composition pipelines. The VFX departments therefore are enabled to work focused on their shots instead of spending time on working out solutions for problems caused in preproduction or production stage. Furthermore the artists do not run the risk of having the director surprisingly change his mind.

The motion capture data has to be heavily optimized and cleaned before being applied to virtual characters or objects, while set scans require a sophisticated retopology and light probes or change logs need to be correctly evaluated. Thus the overall approach hardly pays off in terms of time saving, whereas the later quality certainly compensates for that. The chronic overload, which most VFX companies face due to enormous amounts of shots and ever-growing expectations, cannot be completely defused by applying virtual production techniques, but the new methodologies lead to some relief at least, hopefully not resulting in an even more tightened time schedule in the future.

2.3.3. Creative and Artistic Benefits

Although various new technologies are needed to accomplish a working virtual production environment, resulting in an even more complex and perhaps daunting digital filmmaking process, the additional effort is without any doubt totally justified when considering the creative and artistic benefits coming along. The overall quality of the final product profits from the opportunity to check out different approaches quickly. Since for instance a previsualization, which constitutes the main playground for trying out things, is neither time-consuming nor expensive when purposefully integrated into a VP pipeline, the creative professionals can experiment and work iteratively again, regardless of any financial constraints. They regain the freedom to improvise and maybe come up with spontaneous but brilliant ideas. Alex McDowell points out that "[…] you can work with your fellow filmmakers in a very descriptive, data-rich, virtual representation of the film before you even start making it" (quoted by Faber 2009).

Moreover the virtual production appears deliberating for people with a traditional knowledge background (Derry 2011). According to James Cameron "[…] the aesthetics of physical production and the aesthetics of virtual production are […] pretty much […] identical," (quoted by Billington 2008) as novel techniques recreate a most familiar working environment for traditional set departments, allowing experts to contribute their essential skills and experiences. Instead of relying on postproduction and some "abstract guy sitting in a room" (Derry 2011), the director intervenes in the filmmaking process and delegates tasks to specialized staff.

Eliot Mack, CEO of "Lightcraft Technology", the leading vendor of virtual camera systems and real-time techniques, sees the advantage of traditional on-set productions in the fact "that 100 people can work simultaneously on the project, and everyone on the stage can instantly see the current results." Even if more and more elements on set are now replaced by virtual assets, the collaboration has to be sustained. As progress and results are shared, whether in preproduction or on a live action set, the virtual production enables all artists and creative professionals to work together, from the very first idea until finalization (Dreamspace DOW 2014, p.48), preventing "creative islands" (Thacker 2012) where the different departments work on their own without communicating. A common knowledge base appears most crucial for any constructive criticism or discussion and helps to prevent pointless debates. Consequently the team is able to face risks and uncertainties early enough, focusing on a goal-oriented solution.

Direction

Considering traditional filmmaking, the director is able to explore the stage while getting inspired by beautiful hand-made sceneries and set buildings. Visions and ideas are initiated visually. Since filmmaking has developed into a more and more virtualized process, being now mainly delegated to the postproduction department, the role of the director has been reduced to approving and confirming quasi-ready shots, especially when having limited visual effects experience or limited knowledge about the look of the intended outcome (Clavadetscher 2014, p.194). The virtual production constitutes an attempt to recreate the former responsibilities of directors working inside the digital world.

A real-time preview provides an immediate response, which helps to get an impression of the later composite in real-time without being forced to wait for months. Referring to Marc Weigert, former CEO of Uncharted Territory, the director has "the ability to see the virtual backgrounds or set extensions live on set while shooting, enabling […] [him] to make educated decisions about framing [and] lighting […]" (quoted by Knop 2014, p.7). Any work on a virtual set developed from a blind guessing to a visually supported decision making process. As the feedback is delivered instantaneous, the director is capable of interfering with the creative process, arriving at focused conclusions which are then propagated to all departments. After shooting, when handing over fine-tuned and approved shots to postproduction, most of the creative decisions are already made (Clavadetscher 2014, p.194).

When not following a precise frame-accurate previs, new ideas may emerge spontaneously. In a traditional linear pipeline this off-the-cut decisions might affect the postproduction disastrously (Derry 2012). In a virtual production, supervisors from all production stages support the director in advance or directly on set and prepare their departments for sudden changes appropriately. Nevertheless a detailed previsualization during preproduction turns to account. As all important technical and organizational issues have been clarified there, the director arrives on set and knows already how to stage a certain shot. Like this, it is possible to mainly concentrate on the actors and their performance, assuring a faster and smoother workflow while saving time and money. The director can do what he or she is supposed to do – to direct.

In addition to the mere translation of traditional filmmaking methods to VFX projects, the virtual production also introduces some novel techniques to the director. Since the motion-captured performances of every single actor can be recorded and replayed individually, a layer based approach is obvious. Instead of insisting on a perfect take, in which every department performs at its best, maybe waiting to no avail, the director commissions a combination and rearrangement of the recorded footage. Only the actions which require improvement are then restaged until one attempt meets up to the director's expectation. When an actor has performed an appealing movement, which has been applied to some kind of virtual creature and previewed in real-time, the director and the actor himself can focus in a successive recording entirely on the facial expression. The different components may thus be gathered temporally or even spatially separated, finally assembling a faultless master take. Some month before the release of ‘Avatar’, James Cameron stated that he is “still doing a lot virtual camera work on the film… on stuff that was shot six months ago” (Billington 2008).

Cinematography

Cinematography refers to the process of designing images, including framing, pace and timing, and is therefore dependent on the actions happening in front of the camera lens. But what should a DOP take into consideration when there is literally nothing left to see? When the virtual world is filled with CG elements while the soundstage remains empty, “how would you know where to aim?” (Glenn Derry 2012). Virtual cameras and simulcams constitute an indispensable technology for taking a look into the digital scenery as it was possible before the virtualization of the entire filmmaking process.

A virtual camera system essentially consists of a standard body enhanced with specialized equipment, like witness cameras, tracking markers or lens controls systems, depending on the applied method. They are designed to emulate the features of a real camera, even if not bearing any resemblance to them when for instance deviated from a tablet. Instead of having a 3D artist keyframe a virtual camera in some DCC tool, both devices allow camera work to be performed by camera experts, namely DOPs (Patel 2009, p.4). Years of experience enable them to comply with artistic and creative demands.

Since the CG set has been built already for previsualization, the DOP may take the opportunity to explore the virtual world in preproduction, trying out different camera lenses and sensor types, testing positions and movements and estimating the setup of dolly tracks and cranes (Knop 2014, p.11). During virtual scouting, director and DOP use the virtual camera system to examine the live action set together, looking for interesting spots and appealing framings, no matter if the depicted objects or characters are real or virtual. While shooting, a simulcam provides extended functionality beyond the physical limits of the real device. The DOP may increase or shrink the size of both virtual world and camera, covering an almost unlimited area for manoeuvring. When for example filming a bird’s eye view of a city, imitating the perspective of a flying helicopter, the city can be scaled down to a miniature, allowing mile long flights through the street canyons by taking just some small steps in the soundstage. In addition to that the camera may also be scaled up to support a faster exploration of the virtual scene. As soon as an appropriate location has been found, the camera is scaled back to actual size for shooting (Patel 2009, p.8). When working on a stereoscopic production the simulcam enables the DOP to adjust stereo-parameters, like intraocular separation, convergence and parallax, in real-time by providing immediate feedback and intuitive controls, improving the quality of the resulting stereo effect (Patel 2009, p.17).

As a preview composite is continuously taken into account when deciding on creative issues, it is not necessary anymore to shoot a vast number of different alternatives to make sure that at least one version is suitable for postproduction. Consequently the overall amount of takes is diminished, resulting in a more efficient and cost-saving workflow, while making the lives of director, DOP and actors easier. As soon as the VFX departments take over the DOP is usually released, losing control over further cinematographic decisions. In a virtual production the DOP stays for the entire filmmaking process alongside the director, supervising the virtual cinematography as well (cf. Workman 2014 b).

Acting

Performance capturing developed to a popular practice used in almost every contemporary film with large-scale VFX content, enabling actors to embody any creature, regardless of race, gender, age, anatomy or size (Patel 2009, p.10). After having finished the work on 'The Adventures of Tintin', Peter Jackson states that it was largely irrelevant how the actors looked like. "To some degree, Jamie [Bell] looks a little like Tintin, but he was cast for his screen presence and his acting ability" (quoted by Giardina 2011). While Hergé's character is still fairly human and therefore easy to approach, Andy Serkis created incredibly empathic and compelling interpretations of more animalistic beings, like Gollum and King Kong.

Fig.13: Scenes from 'The Adventures of Tintin'

By delivering a preview of the virtual creatures on set, the actors may evaluate their own performances after each take, continuing with a more specific idea of the effect their acting has on the CG character. Furthermore it is now possible to consider important pre-animated objects or obstacles, which are supposed to evoke some kind of interaction or correctly timed reaction. Nevertheless the ridiculous monster-sticks (Squires 2014, pp.183), which indicate the position of a virtual creature and allow a solid eye contact, will not disappear that quickly, since virtual elements still lack any physical representation in the real world.

Visual Effects

As virtual production techniques were mainly developed to redistribute some tasks from postproduction to previous stages in order to relieve the seriously overstrained VFX departments from their work overload, it appears only fair that animation and compositing artists or technical directors profit as well from the innovations coming along. First of all, the VFX studios are involved much earlier in the production process. From the very beginning they stand by for consultation and oversee the entire production from an artistic and technical point of view. After some modelling sessions for evaluating concepts and artworks the artists continue with the previsualization, create tons of medium-quality 3D models and puzzle out movements and timings in accordance with director and cinematographer. This inclusion into the early creative design process illustrates the transformation of VFX companies from suppliers of mere handcraft to art departments, finally honouring their work and constituting a paradigm shift within the entire industry.

The content received from previsualization is already elaborate to such an extent that it constitutes an ideal starting point for an efficient and framing-based asset creation, even before the actual shooting starts. Instead of building an entire virtual world not knowing which part might later find its way into the film, the 3D artists focus only on those virtual set extensions that are necessary concerning a certain framing. During production novel techniques enable the on-set crew, ideally reinforced by some VFX supervisors, to match digital and real image components precisely. Already existing live action plates can be lined up with virtual assets in real-time, resulting in a more consistent look and preventing erroneous or even useless outcomes, for which the VFX artists would have to pay later on (Patel 2009, p.17). Moreover, CG assets or animations can be replaced or optimized in no time even while shooting. Thus, only correct data is provided to the postproduction team, which then does not need to spend days for fixing problems or producing dozens of iterations but can focus on quality. As almost all elements have been approved by the director right on set, the decisions can be considered as final and are executed without any further delay.

After shooting, the pre-composited material contains all important components, previewing an informative combination of real and virtual elements. This footage constitutes an ideal initial point for a postvis, in which the subsequent steps in postproduction are visually anticipated. When animating characters or cleaning up motion captured animation it is extremely helpful to consider also the illumination, as the lighting influences the appearance of gestures and especially mimics to an extent that should not be underestimated. A sophisticated virtual production provides measured light probes or at least chrome-ball images which allow a reconstruction of a real lighting situation and therefore offer a beneficial feedback to the animators (Neufeldt, Baneham 2011). Alongside the main camera dozens of witness cameras observe the actors' performances from a static perspective and deliver useful reference material of mimics and gestures.

Other Departments

Besides direction, cinematography, acting and postproduction other departments profit as well from innovations introduced by a virtual production pipeline, since everybody has the possibility to see the virtual components in real-time on set or as composite afterwards (Trumbull 2012).

Similar to the visualization of architecture, sceneries and stage elements can be visualized and literally explored using a simulcam before a single piece of set has been crafted. Like this, the production design team is able to develop various concepts without spending enormous amounts of money. As the director approved the concept with regard to its compatibility to the previsualization the risk of an unavoidable demolition of already built elements is limited. In addition to that, more and more parts are replaced by virtual set-extensions, avoiding gigantic buildings but not lacking their optical presence on set (Legato 2012).

The displayed animations serve as cue for lighting or special effects. When for example real flames, smoke and blazing light are intended to be triggered in the moment a fire-breathing dragon attacks a group of knights, both pyro-technician and gaffer consult the preview of the virtual dragon on a monitor, awaiting the creature to reach a defined point before taking the initiative. In general, lighting technicians should take the virtual elements into account when planning their setup, thus guaranteeing an impressive atmosphere throughout the film, synchronizing virtual and real light persuasively.

The composites can be roughly trimmed and connected directly on set, providing a preview editorial for the director to decide on the further proceeding. After shooting, the editor can access descriptive footage, which already contains all important elements. Compared to inscrutable green screen material the footage derived from a virtual production enables the editor to see what he is supposed to see, identifying continuities and orders easily.

Last but not least, producers, advertising agencies or broadcasters get an impression of the final results already in preproduction or right on set. “With a virtual production the client is able to see changes on the fly” (Knop 2014, p.11), recovering the opportunity to ask for changes or give suggestions.

2.3.4. Challenges

Considering all the artistic and creative benefits and pipeline improvements coming along with the adoption of a virtual production environment, one might wonder, why most of the creative professionals still seem to elide this apparently beneficial procedure. Even if most of the technologies are well-tested and work quite reliable, the integration of all these different devices into a concerted toolset is not yet completed while workflows have to be further conciliated. In order to prepare virtual production systems for a widely-spread application in the future, some challenges are still to be met.

Creative Challenges

The virtual elements have to blend as seamlessly as possible with the live action footage. The closer the CG assets resemble the real world the more significant the creative decisions turn out (Patel 2009, p.12). Thus, an almost photorealistic visualization, including realistic lights and shadows, surface textures or even particles and fluid simulations, allows an immersive experience which gets close to the work on a real film set. While logically depending on the capabilities of the real-time engine, the overall look mainly relies on the quality of imported assets, textures, materials and animations and therefore requires skilled artists that have been trained for achieving exceptional results while respecting performance issues. Often, this balancing act demands creative approaches and solutions. On set some kind of physical representation of virtual elements is mandatory. Once an interaction with computer generated elements, like objects, surfaces or characters, is required, stand-ins help to feature realistic gestures and guarantee a convincible physical contact. When an actor is for example expected to climb a gentle slope it might be difficult to perform accordingly while standing on the flat floor of a soundstage without any tangible or visual feedback. For the shooting of 'Avatar' a modular terrain system allowed to build up a stylized landscape which totally sufficed for providing the needed orientation. Likewise a walk through high grass and fern had to be somehow reflected in the actor's movement and was therefore intentionally obstructed by plastic leaves and ropes (Derry 2012). Consequently the decision makers should consult the production design department early enough to go through the shots together and discuss the necessity of such stand-ins before shooting. Nevertheless, even if providing physical feedback, the actors remain blind to a large extend and can hardly benefit from the innovations the virtual production techniques introduce. They are not able to haptically feel the virtual world and thus cannot factor their perception into the performance (Derry 2012). Since one cannot hope to access large-scale holographic displays and whole-body suits for tactile feedback in the near future, the actors will have to live with this disadvantage for now and accept the challenge. Anyway, as Rob Legato points out, the actors never really saw the set as they usually performed towards the camera with the scenery behind them (Legato 2012).

Technical Challenges

Real-time visual effects and game engines come along with specific challenges and characteristics. First of all, the asset generation already requires a different workflow compared to the traditional pipeline, where images are rendered offline in postproduction. Now, for each virtual element a real-time capable interpretation is needed. As every single polygon has to be computed at least 25 times a second, the overall model and texture complexity has to be seriously limited, requesting artists and technical directors to leave behind their familiar approaches while getting used to a procedure more similar to video game production (Knop 2012, p.10). After creation, all CG models, scenes and animations have to be prepared and optimized for an appliance in the real-time engine. Textures are compressed, rigs and blend shapes simplified and illuminations baked. However the virtual elements will later be composited with live action footage and thus have to comply with the realism of a photographed image.

While shooting, the virtual production engineers have to ensure that all necessary motion data is correctly captured, stored and distributed. Therefore, the rotations of the actors' limbs, the facial performances and the variable attributes of the camera, namely position, orientation, focus, aperture and focal length, are measured constantly while inaccuracies like flipping joints or jitter have to be cleaned up right away. Though the existing systems deliver a convincing capture of motion, managing such a big amount of data constitutes not only a complex challenge when trying to keep track of all the different components but demands powerful and efficient hardware equipment (Dreamspace DOW 2014, p.45). Hence some formerly more or less unnoticed expertise or technology appears now essential.

The preparation and set up of a well-functioning virtual production environment is time consuming, exhausting and costly in terms of material, devices and money. Film sets may appear already complex enough, but now more computers, servers and reams of cables join in and complicate the workflow additionally. As already seen, virtual production is mainly technique-driven and goes along with an ever-growing technical complexity. New camera types, innovative interfaces or sensors and new strategies for data transfer have been established and must now be correctly applied (Knop 2012, p.71). For example the camera lenses have to be extensively calibrated in advance while every single recording step requires synchronized timecodes and genlocks. Generally the entire setup is still too laborious or error-prone and requires further optimization (Clavadetscher 2014, p.198). Moreover, the novel technologies of virtual production have not yet been completely integrated into the current structure and require improvisation of the responsible departments to some degree. Ideally the various parts of equipment, especially the camera tracking system, should be handy and compatible to existing hardware, grips and mounts, allowing a robust setup even for handheld, dolly, steadicam and crane shots (Knop 2012, p.71). However the new devices often lack universal applicability and need some tinkering.

When leaving the soundstage for an outdoor shooting, more complications arise. Beside the obligatory weather uncertainties, the overall dimension of the set challenges the virtual production procedures. With more actors involved and a larger area to cover, the size of the needed capture volume increases and soon reaches the constraints of optical motion capture systems. Moreover these setups often suffer from altering light situations and variable contrast ranges of cloudy skies.

Workflow Challenges

In order to prepare a smooth and harmonic integration of virtual elements into the real world, the virtual production pipeline demands advanced assets already in a quite early stage of production as well as sophisticated object databases, asset kits and strategies for working with instances. To some of the creative professionals this workflow may appear unfamiliar and inconvenient as it requires an especially concerted collaboration in which all involved departments, especially the CG artists and production designers, are synchronized, but it constitutes the only way to avoid unnecessary complications on set. Like this "the facility has to build, texture and even light assets before production, unlike normal schedules where this process can be started later." However the conversion to a more forward-looking production seems challenging and must not always lead to success. John Knoll, senior visual effects supervisor at ILM, further says, that "there is a danger of getting the worst of both possible worlds. That you build everything before, shoot and then have to build everything after because parties involved won't make a decision until the very last moment" (quoted by Montgomery 2014). When relying on teamwork and a seamless propagation of data to such an extent, a homogeneous workflow is indispensable. The pipeline therefore should support versioning, syncing and file conversion, while providing an asset history and metadata. In addition to that the departments will have to decide on a consistent colour space (Knop 2014, pp.79). As the standardization of formats and workflows is still not accomplished (Trumbull 2012) and most of the studios and facilities stick to their longstanding pipelines, the development of an efficient working environment remains challenging and relies on customized solutions. Virtual Production workflows totally prove their worth when exercised within the scope of previsualizations, experiments and tests in a secure soundstage, but seem not really prepared for the "heat of production where timing and reliability are of higher value than excitement of using new gear" (Knop 2014, p.78). The idea that an entire production team waits idly for some software engineer to fix for example the asset stream between two software packages appears fairly abstruse. Thus technicians and virtual production supervisors have to make an even bigger effort to achieve reliable and faultless solutions which blend in as unseen as possible into the traditional proceedings. However it seems factually impractical not to obstruct the staff on set when setting up a motion capture system on stage, thus in an area which has formerly been reserved for lighting and production design. In general, the virtual production methodologies are demanding on staff resources as well as hardware dependant as they come along with lots of monitors, workstations, mocap cameras, test grids, makers and tons of electronics. Such additional hardware might get smaller or less obtrusive, but will have to be accepted as an essential part of virtual filmmaking.

Revising Mentalities

The biggest challenge the virtual production environments may have to meet is the stubborn mentality of the creative professionals – the well-known 'we have always done it like that'. According to Eliot Mack from Lightcraft, "people have been making movies quite successfully for at least a century, with remarkably little technical innovation along the way. A lighting grip from the 1930's could walk onto a modern stage, spend a couple of weeks coming up to speed on the new types of lights and become a productive member of the work crew in 2014" (Knop 2014, p.13). As the filmmakers work within such an enclosed surrounding largely insulated from innovative ideas for years, it will be difficult to convince them to overcome ingrained habits. In addition to that most of the CG professionals have never been on a live action set. Therefore they will have to get accustomed to a more fluid and iterative workflow and accept the fact that last minute decisions are a daily occurrence (Knop 2014, p.77).

Moreover, the directors join in with their very own approaches, bringing along specialized knowledge and previous experiences. A technophile geek like James Cameron can contribute to the virtual production workflow, trying out innovative techniques on his own, while a visionary of the old school like Martin Scorsese may be more concerned with visual language, pace and performance. Both director-archetypes used virtual production in recent projects, Cameron in 'Avatar', Scorsese in 'Hugo Cabret', and were able to put their vision into practise by following completely different strategies. Cameron devoted himself to the technology and worked nearly autonomous, Scorsese relied on the competent guidance of a virtual production supervisor (Legato 2012).

Actors may be called a strange breed as well. As virtual production techniques massively affect their way of working, it should be hard to argue them into the advantages a camera operator or director might take. Steven Spielberg, director of 'Tintin', understands the reluctance. "They're wearing motion capture suits, they've got marks on their faces, they wear these helmets with a camera and light built in. It takes a while not to crack up while you're doing serious dialogue with a fellow actor who looks like a scuba diver" (quoted by Giardina 2011). Even if 3D printed facial tracking toolsets allow a personalized and therefore perfectly suiting setup, the mounts are still bulky and obstructive.

Actually every department, no matter if in preproduction, on set or in postproduction, will have to adapt to altering workflows. Since the introduction of new procedures and tasks can rarely be accomplished without running into any lag or troubles one must hope for the professionals to understand the benefits and chances going along with virtual production.

2.3.5. Limitations

The challenges striking the filmmakers when deciding on a virtual production do unfortunately not constitute the only barrier though seeming already annoying enough. While the creative professionals are at least able to handle these issues somehow, there are serious limitations which cannot be negotiated yet. Especially technical deficiencies diminish the artistic and creative capabilities of novel approaches so that the virtual production environments cannot fully show their potential. However future developments, innovations and inventions, like the interfaces presented below, aim at addressing these lacks.

Technical Limitations

When streaming data and signals a time lag cannot be completely avoided. Considering a real-time environment, the overall network delay should stay beneath five frames or 200ms when shooting with 25fps because otherwise the evaluation of motion becomes guesswork. Advanced systems guarantee 40ms when applied correctly. However, as a virtual production requires lots of different systems, ranging from virtual cameras to real-time keyers or motion capture systems, the delays add up. Furthermore game-engines are designed to deliver always the highest possible framerate instead of providing a constant but maybe lower refreshing frequency. This variance may constitute a problem when interconnecting real-time renderings and camera footage on traditional display devices on set.

Another critical constraint consists in the absence of depth information. Without a high-resolution depth map it is impossible to merge virtual elements with live action footage properly. The CG objects are either screened atop the real images or positioned behind a chroma-keyed area not previewing the single composition layers in the right order and thus losing occlusion and spatial staggering which would be necessary for a well-funded anticipation of the postproduction result. In fact several technologies for depth capturing are developing right now, but the real-time processing of gathered data is still too computationally intensive, not to mention the lack of quality.

Though being in use for some years, the mocap systems also suffer from some insufficiencies. The very robust optical tracking solutions are nevertheless vulnerable to reflections and limited in terms of range and height while their cameras cannot get rid of the image noise and might cause jittering. Inertial, electromagnetic or ultrasonic approaches fail at accuracy (Clavadetscher 2014, pp.195). In addition to that, build-up and calibration appear cumbersome and time-consuming, no matter which technology is applied. In a worst-case scenario the systems have to be rearranged and recalibrated before each shooting session (Jones 2012).

In virtual productions lighting is widely discounted, although the creative professionals should actually be aware of the importance of this cinematic tool. Physically correct shadows, reflections and refractions are to some extent possible within a real-time raytracer but all too often ignored when assembling the previews. The art director Andrew Jones (2012) states that “[…] lighting is not so developed as it might be and a lot of the lighting cues are difficult to achieve with any sophistication. There was on scene in ‘Tintin’ where you had to have a character receding into shadow and we just had to take guesses of how that would work.” Moreover there is still no way to measure the real illumination directly on set and apply this capture data to the virtual scenes.

Until now the virtual production systems lack visual feedback for evaluating facial expressions in real-time. As already mentioned Kabukis project the recorded video footage of the actor’s mimics onto the meshes and indeed allow a rudimentary preview. However, without a rig-controlled deformation of the facial mesh it is hardly possible to evaluate a full performance.

Lack of Dedicated Software

Game engines provide an ever-expanding visual quality but are logically optimized for creating video games and therefore lack intuitive tools and control mechanisms for filmmaking (Mazalek, Nitsche 2007, p.1). When designing real-time environments for games the developers have to consider the interactivity of an entirely accessible world and focus on the correct representation of physics, whereas on set options for frame-accurate timing and shot-based modification appear much more important. Moreover raytracing and rendering techniques have been prepared for visualizing prefabricated effects which are part of a sequence of scripted events and mostly triggered by the player, thus not providing the filmmakers with the necessary instruments of manipulation. In addition to that the engines are designed to save games and scores but often cannot record revisions of animations or parameters or export the changes to DCC and postproduction tools (Knop 2014, p.38). Consequently the modifications made on set are lost after shooting and have to be reperformed afterwards, which totally reduces the idea of a virtual production to absurdity.

Software packages and DCC tools are often limited in functionality and specialized on a certain aspect of media production. Therefore the artists have to access numerous frameworks during the production of a film and take the risk of later incompatibilities, especially when dealing with different file conventions or data streams. Until now there is no ideal solution for bringing preproduction assets to a film set and further to postproduction without running through different standards, formats and codecs (Knop 2014, p.78). There are some tools for allowing the artists to align their workflows, communicate changes and share creations but yet not all departments have direct access to these systems (Knop 2014, p.7).

Furthermore, an intuitive real-time compositing environment, in which 2D artist can work directly on set, is still sorely needed. Established tools like Nuke or After Effects base on stacks of more or less complex effects which are batch processed successively. As already seen, these operations fall short of real-time when computed with the CPU. Since the parallelized and therefore faster GPU processing is not yet ready for use, apart from basic pixel operations, the compositing on set is still only applied for offline previews, like in ‘Hugo’, where 2D artists put together a fast slap-comp of the recent shot in case the director requested a review. Though being an elementary part of the compositing workflow, advanced techniques like rotoscoping, mattes, depth of field or multipass-layouts are not supported while the overall pipeline remains widely unsuitable for passing through visual effects from set to postproduction. The latest attempts to accomplish some kind of real-time compositing, for example the ‘Frischluft Lensfeed’ plugin for After Effects, limit themselves to real-time keying and short stacks of simultaneous effects, but already appear useful for optimizing shots for composition and allow an outlook on the potential of future real-time compositing environments (Dreamspace DOW 2014, p.61).

Lack of Specialized Hardware Interfaces

In addition to the apparently inadequate software, also the hardware lacks optimization for virtual production systems. Cameras for example should be “[…] controlled by tactile means and not in a disembodied fly-through way as seen in games” (Nitsche, Kirschner 2013, p.304). Consequently tangible interfaces are necessary and simulcams or virtual camera systems already constitute definitely a quiet satisfying solution. Apart from that there is actually no dedicated equipment from which the filmmakers might benefit on set. Traditional game interfaces like controllers or joysticks could be diverted from their intended use by applying them as interaction device for shifting objects or changing animation, but they are nevertheless optimized for gameplay and cannot meet the whole range of requirements of filmic production when applied out of the box. As already seen, the iterative workflow might require a permanent adjustment of CG assets in order to put the creative decisions of the director or cinematographer into practise. During the shooting of ‘The Adventures of Tintin’ a 3D artist sat at a Maya workstation and waited for being asked to position or replace an object. This procedure did not prove to be very practicable since any change request had to be communicated to an animator or modeller before realization, including the risk of misunderstandings. The decision makers have to trust in second-hand workmanship that way, unable to influence or assist directly. Taking everything into consideration, the existing hardware is unintuitive and does not suit the collaborative character, at the same time undermining the main advantage of virtual production environments, namely the opportunity for experts to work within their familiar remit in a way they have been accustomed to. The only solution consists in the development of innovative and intuitive hardware interfaces, which are optimized for coping with the challenges of virtual productions.

Chapter 3

Virtual Production 2.0 – Innovative Interfaces on Set

Considering the advancements made in the fields of workflow and technology in the last few years, it is obvious that the film industry has finally become aware of the potential of virtual production. By now several devices have been developed which allow the user to navigate through a hybrid world, consisting of virtual elements and live action plates, while enabling the decision makers to modify assets in no time on set. Having said that, one must not disregard the fact that most of the shortcomings described in the previous chapter have still to be accepted as true, even for the latest approaches.

3.1. Interface Basics

Unlike a robot, that processes predefined patterns autonomously, all computing systems, no matter how advanced, imply the presence of a human operator (MacKenzie 1995, p.437). In the last decades a large variety of devices has been developed, ranging from punched cards to sophisticated user interfaces that are either all-purpose or customized for a certain field of application and aim for a convenient and intuitive human-computer interaction.

A general definition refers to a human-computer interface as an environment in which a person is able to communicate and interact with a software application or hardware device. Strictly speaking, a user interface does not only consist of input and output devices but also includes the interaction techniques and transfer functions that are required to translate the user input into commands that can be processed by a computer.

Input Devices

When designing interfaces, the selection of appropriate input devices proves to be of utmost importance, since a wide range of user interactions has to be captured adequately.

The most relevant property for classifying an input device is its degree of freedom. The DOF value gives an indication of the amount of movement types that can be measured simultaneously, thus specifies how complex a device actually is. Accordingly, every degree of freedom represents one particular axis of user interaction (Bowman et al. 2005, p.88). A tablet or a standard 2D mouse provides only two DOF, namely the freedom to move on a horizontal plane for instance from left to right or from back to front, while a 3D input device like a spacemouse offers six DOF, including not only all three spatial axes but also the equivalent rotations. Since the control space of an input device should at least correspond to the space the user needs to perform a certain task, a traditional mouse should be completely unsuitable for moving a CG object in three dimensions (Mazalek, Nitsche 2007, p.157). Nevertheless all DCC tools are handled that way, as they supply additional modifiers to change the mode of transformation. When pressing a key or a button while shifting the mouse, one and the same movement results in a different outcome inside the application.

Furthermore, one has to distinguish the types of input. An isotonic device is sensitive neither to pressure nor to force and capable of moving freely in space without resistance. In addition to that it is not automatically reset to origin after the user has performed an operation. Such device, be it a mouse or a slider, appears especially appropriate for controlling positions. Isometric devices cannot be moved at all as they remain fixed at their origin, offering a theoretically infinite resistance to user input. A spaceball serves as an example for this kind of gadget whereas a joystick belongs to the group of so-called elastic devices which are in turn reset to origin automatically while allowing minor movements of some millimetres or degrees. Both isometric and elastic devices are force-sensitive and therefore suitable for changing the velocity of modification instead of affecting the absolute position.

Finally, input devices can be described by specifying whether or not they rely on physical interaction. Without a user pressing a button or performing some kind of task, active input devices do nothing but wait, as they always require a human operator. As soon as one of their components is manually modified, these devices continue to transmit data. Passive input devices however operate continuously without any user involved. Like this, tracking systems deliver a measurement of both position and motion, no matter if an actor moves within the volume or the stage remains in fact empty (Bowman et al. 2005, p.89).

Interaction Techniques

Interaction techniques are those methods that are applied by a user in order to fulfil a particular task. Since the quality of human-computer interaction does not only depend on an appropriate utilization of input devices but is also fundamentally affected by transfer functions and output procedures, one has to consider all software and hardware components of an interface, when designing customized elements for one of the following interaction techniques.

In order to work within a virtual space, regardless of whether it is a 2D desktop or a 3D environment, one has to be able to travel from the current position to any other desired location. The interface thus provides methods for navigation, encompassing the associated sub-steps of searching, wayfinding, exploring, and manoeuvring. While designing the interface, it appears important to take into account the distance to be travelled and the accuracy to be achieved. Furthermore some methods may be inapplicable when the target is for instance not visible from the starting position or when the navigation has to be performed in parallel to another interaction technique, while the input device provides only a limited amount of DOF or modifiers (Bowman et al. 2005, pp.184). In practice the process of travelling can be carried out by relocating a graphical representation of the own position on a map or by manipulating the viewpoint of the virtual camera via some input device. The most natural way of moving through a virtual scene consists in a setup which enables the user to feed in a real forward motion, like a walk in a treadmill or a ride on a sensor-rigged bicycle.

Once the user has detected an object to modify and perhaps travelled to the correspondent target location to take a closer look, it is necessary to communicate to the computer which asset to pick. At first glance, a selection seems to be a rather trivial task as it constitutes a fundamental way of manual interaction in our everyday non-virtual world. However, when working with a computer the operator is not capable of using his or her hand to directly touch and grasp an object. Hence the interface has to feature sophisticated strategies for selecting virtual assets. While it appears fully sufficient for 2D applications to examine whether the position of a graphical overlay like a cursor matches the coordinates of the target, issues get much more complex when working in three-dimensional space, as objects may there have an additional offset in depth. Ray-casting constitutes the most common 3D selection technique and enables the user to point at a target with a single ray that is attached either to the virtual counterpart of the user's hand or to some kind of widget. Although this approach proves to be reliable and intuitive, it lacks accuracy when trying to select very small or far distant objects. In a slightly varied approach an adjustable cone, comparable to a flashlight, allows the user to select an object even without pointing exactly at it, taking the risk of several elements falling within the scope of the cone. In general, the process of selection is heavily affected by the size of an object, its distance to the virtual camera and the density of occluding objects. In order to complete a selection, the user has to confirm the target for example by pressing a button, ideally receiving a graphical or acoustic feedback which indicates the modified status (Bowman et al. 2005, pp.149).

As soon as an object has been selected, the user can continue with manipulation. This interaction technique covers the modification of position, orientation, size and any other case-related attribute. Manipulation is again quite easy to accomplish for 2D interfaces by attaching the target object to a cursor or integrating widgets, whereas 3D environments require advanced procedures and may fall back on solutions similar to those used for selection, namely ray-casting and virtual hands. Pure ray-casting is easy to handle, yet it cannot come up to the required accuracy. On the contrary, hand-centred object manipulation turns out to be much more precise and guarantees a natural and therefore intuitive control but might become fatiguing or even restrictive when sticking to an inappropriate mapping function. If assigning the real motion to the virtual world one-to-one, it would for instance be impossible to relocate an object in a position beyond the range of the user's arm – a serious constraint that can be avoided by applying a non-linear mapping function, which allows the virtual arm to exceed the length of the real one for far distant operations (Poupyrev et al. 1996, p.79). Alternative approaches propose downscaled virtual worlds in which the user is able to perform any manipulation within the natural area of reach.

It is most likely that a device does not provide enough degrees of freedom to enable an operator to adjust all parameters or access the entire range of interaction techniques at once. Consequently, an interface needs additional system controls to offer the user the possibility to switch between different functionality or system states. These modifiers can either be keys on a standard keyboard or base on more complex graphical user interfaces, including menus, icons and widgets. Graphical menus for 3D interfaces often recycle the techniques introduced by 2D desktop applications, as these approaches have proven their worth for several years and appear familiar to most of the users. One suitable adaption comprises traditional menus that are attached to the virtual counterpart of the user's hand or to a plane in three-dimensional space, while 1 DOF ring menus encircle the virtual hand and allow the operator to switch over to a different mode via a simple twist of the wrist (Bowman et al. 2005, pp.261). 3D Widgets are well-known to artists working in DCC tools and provide not only an indication of which interaction technique or manipulation method is currently selected but also make it possible to choose a certain way of modification by moving only the correspondent part of the widget. Since this graphical menu is rendered on top of the related object, its functionality is directly linked to the target (Conner et al. 1992, p.184).

Last but not least, an interface may provide symbolic input, including alphanumerical symbols like numbers or letters, which enable the user to obtain abstracted but explicit information quickly and concisely. This kind of interaction is needed for labelling or entering numeric values for exact parameter manipulation. Moreover annotations can be attached, that allow for instance designers to share and communicate information and thoughts. While numeric inputs can be considered as the backbone of every 2D application, they are nevertheless hardly applied for 3D user interfaces, mainly because the operators cannot simply access a standard keyboard. Alternative techniques like pen-based input, gesture-driven interaction or speech recognition have already been developed to a promising quality as they are now applied for example in cars for satnav or console handling, but still lack practice in 3D environments, mainly because they are difficult to implement (Bowman et al. 2005, pp.287).

Transfer Functions

A transfer function describes a mathematical method for transferring a user's interaction from the physical object to the virtual scene. This procedure is also called mapping, since every value measured by the device is mapped to a correspondent parameter value. When classifying transfer functions for 3D user interfaces, one has to distinguish between position, rate and acceleration control.

Zero order mapping or position control describes a straight transfer of position values from the device to the virtual object, what does not mean that the related function has to be linear since it may include some kind of coefficient which implicates for instance the speed of the input movement. In case the function transfers the data directly and contains neither nonlinear components nor multipliers for scaling the incoming values, the mapping is called isomorph. Accordingly, a system may provide functionality to switch for example a mouse mapping over to a non-isomorph zero order mapping which alters the absolute translation value of the cursor while factoring also the velocity of input into the function. In contrast to position control, rate control directly accesses the speed at which a certain object parameter is modified. When displacing the input device from zero state without performing further interaction, the object will continue to move or rotate at a constant rate. Acceleration control defines the increase of velocity. As long as the device remains steadily displaced from zero state the object speeds up with constant acceleration (Helzle, Spielmann 2014, p.26).

Output Devices

Whenever the user navigates through menus, travels across three dimensional spaces or selects and manipulates objects, a feedback is needed, be it visual, haptic or auditive, which provides all required information about the current state of the respective human-computer interaction. Thus an output device constitutes an indispensable component of any interface. When deciding on an appropriate output device, one has to take into consideration the field of application on an individual base. However display devices normally prove to be the most flexible choice as they cover a wide range of different techniques and use cases, ranging from small portable screens for displaying selected information to large multi viewer solutions (Helzle, Spielmann 2014, p.21).

3.2. Developing innovative Interfaces

Before developing an interface for a dedicated field of application, one must not only become aware of similar developments made by other professionals and companies but also keep in mind the particular tasks the novel system is meant to fulfil. Accordingly, appropriate devices have to be selected prior to the implementation.

3.2.1. Latest Developments

Identifying the shortcomings of current virtual production technology, several developers and companies go for improving both software and hardware systems or work on completely novel interfaces and workflows.

SmartVCS

When previsualizing shots or scouting virtual worlds a virtual camera system like the one applied in 'Avatar' requires expensive hardware and large mocap volumes in huge studio environments and appears therefore hardly affordable for small or independent projects. Girish Balakrishnan developed a hybrid interface consisting of ready-made devices such as game controllers and multitouch tablets (Workman 2014 a). By using only consumer-level technology and openly accessible game engines, the SmartVCS "[…] is designed for directors, both amateur and professional, who wish to embrace the notion of Virtual Production for films and game cinematics without a big studio budget" (Balakrishnan, Diefenbach 2013, p.1). Besides recording the camera attributes including position, orientation and intrinsic parameters, the system also provides a fully functional real-time engine to load 3D geometry immediately.

In detail, the SmartVCS utilizes an Apple iPad, a Playstation Move controller and the Unity[9] game engine. As the iPad mainly serves as a display device, the Move controller is applied for measuring both position and orientation in space. For this purpose, a fixed Playstation Eye camera tracks the position of the glowing ball on top of the Move controller while a sensor inside registers the orientation. Two Playstation remote input devices are attached to the tablet and enable the user to perform additional commands which can be mapped arbitrary to the virtual camera parameters by writing custom scripts. All gathered data is streamed wirelessly over the 'Move.Me' server application to the Unity game engine, where the virtual scene is rendered in real-time while supporting rudimentary asset editing via iPad touchscreen control (Workman 2014 a).

Balakrishnan states that his "[…] first goal in developing the system was to take all the functionality of traditional virtual camera systems such as freespace motion, camera and lens control […] and bring it all to a mobile platform" (quoted by Workman 2014 a). In addition to that, his system uses

9 http://unity3d.com

Unity's multiplayer networking structure to set up a collaborative working environment that breaks the physical barriers of virtual production by allowing people to contribute to the filmmaking process while not even being in the same studio, or city or part of the world.

Taking everything into consideration, Balakrishnan's SmartVCS system empowers professional directors and amateur filmmakers to elaborate a shot layout for both films and game cinematics with a most familiar and intuitive technology. Moreover, the touchscreen control for scene modification already appears very promising, although the tools for set editing do not tap the full potential of the Unity game engine.

After presenting the SmartVCS at several big conferences around the world, Girish Balakrishnan got offered a job as virtual production technical director at Digital Domain (Workman 2014 a).

Fig.14: Smart VCS system

Zeus Scout

Already in 2012, the visual effects studio Zoic developed a proprietary application that allowed a physically correct view into the virtual world by enabling the user to adjust the parameters of a virtual camera in a way that mimicked the features of a real camera. This early system was hardly anything but a basic framework for loading GUI elements and assets, while each of these components had to be loaded and compiled separately. Even if the workflow appeared sufficient for in-house productions in which the tools are only applied by the developers themselves, one could not reasonably expect to reach a broader group of customers. Thus Zoic set out to advance the application and released Zeus, the Zoic Environmental Unification System, which is now available as iPad app via iTunes for $9.99 (Altman).

The Scout version of Zeus constitutes a previs and scouting tool, as the name implies, and accesses only the standard components of the tablet device without requiring additional interfaces or sensors. Like this, objects can be modified or settings changed by tapping on the multitouch screen, while the position and orientation of the device is measured by retrieving the corresponding values from the gyroscopes and accelerometers inside the iPad. The Zeus Scout comes along with seven different operation modes. The view setting enables the user to walk around freely in search of appropriate camera positions, which can be saved and stored for blocking purposes. In measurement mode, it is possible to define the scale of the virtual scene by gathering numeric values in the real world. Textured 2D figures can be added in character mode while CG assets are positioned, scaled and rotated in prop mode. The previs service provides the required tools for previsualizing a shot whereas tracking mode enables the user to apply the Zeus Scout as virtual camera. Last but not least, live action footage captured by the iPad camera can be chroma-keyed and integrated into the virtual world when switching to video mode (Wolfe 2014).

Fig.15: Zeus Scout on tablet

RTFX

The real-time special effects tool RTFX constitutes at a broad-based solution for creating previsualizations. While most of the existing toolsets are designed to perform only one particular task, like data producing, editing or rendering, the RTFX application builds up a “chat-like client-server architecture” (Northam, Istead, Kaplan 2012, p.1) to combine the different features into one single system. As the resulting framework is generic, it is possible to address whatever DCC tool or game engine by writing specialized plug-ins.

The developers tested their RTFX tool in a virtual production use case, applying a Vicon motion capture system, the Houdini DCC package and the Unreal Engine. The resulting setup appeared to be most reliable for a unified previs of motion capture data, simulated effects and 3D scenery (Northam, Istead, Kaplan 2012, p.1). However, as this approach lacks video input, the RTFX tool is definitely more suitable for machinimas and full-CG previsualizations but cannot be used in a filmic live action production.

3.2.2. Requirements

Ideally a virtual production employs various hardware and software components to enable the creative professionals on set not only to access a flexible real-time environment to work in but also to modify characters and assets immediately. As seen above, SmartVCS and Zeus Scout introduced a couple of promising advancements in the fields of virtual scouting and camera control, additionally offering basic toolsets for different forms of interaction and manipulation. However there are still no innovative solutions when it comes to intuitive interfaces for a real-time modification of objects and sceneries. Moreover most of these environments are either proprietary, expensive or lack usability. Particularly when applied by filmmakers without a specialized knowledge of computer graphics and 3D software, the current tools show apparent deficits in terms of intuitivity.

Considering the progress induced by virtual and augmented reality, though the current hype might tempt to overestimate the actual potential of these techniques, any virtual production may profit from adapting novel interfaces for a mixed-reality application. Innovative display devices and advanced technologies for manipulation, navigation and visualization have been developed to a considerable quality and hence must not be ignored any longer when designing virtual production environments (Helzle, Spielmann 2014, p.19). The necessity for novel interfaces has to be mainly approached by sophistically implemented and interconnected devices, which enable an intuitive human-computer interaction. These hardware interfaces are furthermore complemented with clearly arranged GUIs.

Input Requirements

Since the main advantage of innovative virtual production environments can be seen in the ability to manipulate virtual objects, cameras and lights, particular devices for real-time animation, set and light editing seem especially important.

Set editing refers to the interactive modification of static virtual objects in real-time on a film set. Due to tight time schedules, which minimalize the amount of changes an artist can perform between two takes, it is obviously not possible to replicate the full functionality of an offline DCC tool. However basic object parameters like position, orientation and scale are meant to be altered in real-time. Besides an instant visual feedback, this modification requires dedicated control units and intuitive interfaces, conveniently customized for the various departments on set. When committing changes, annotations or notes should be attached in order to document history and progress for future recapping.

Light editing covers the harmonization of real and virtual lights. Since lighting on a real film set constitutes an essential creative tool, significantly influencing the overall look of the final result, the digital elements on a virtual production set are expected to represent the natural illumination as well to guarantee a convincing integration. Ideally the parameters of real light sources would be captured and automatically assigned to the virtual world. Since no reliable device for gathering light probes is accessible until now, the virtual lights have to be manually adjusted to match the real lights with respect to colour, intensity, position and orientation. Thus, an editing interface is required which offers the possibility to select and modify single attributes. In addition to that, changes of the virtual world might also be transferred via DMX control to real lights inside the soundstage.

Animation editing describes the subsequent modification of prefabricated backed animation without using a DCC Tool. With more and more 3D assets entering the soundstage, some form of interaction between virtual and physical elements appears desirable, thus demanding user controlled and perfectly timed animation, especially when trying to match digital objects with their real counterparts. Since animation does not only include the three spatial dimensions but also the instant of time an appropriate projection has to be found to map all necessary information to an interface and make the user perceive the virtual environment, including the possibility to jump back and forth in time. This can be achieved by accessing suitable types of input and display devices (Helzle, Spielmann 2014, pp.23).

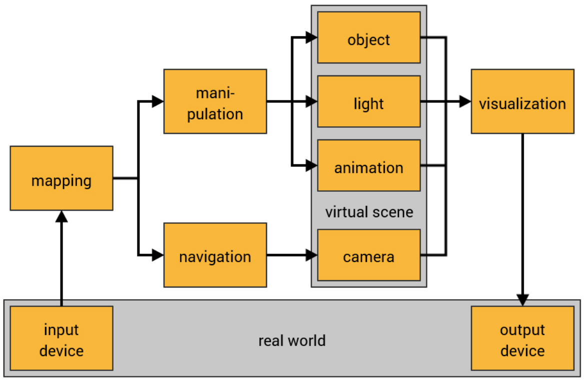


Fig.16: Generic editing workflow

Display Requirements

A traditional monitor provides visual information to a small group of people within a limited area and forces the decision makers to stick around it all the time. Moreover it constitutes a quite immobile and exclusively passive display device, not at all meeting the iterative and interactive character of a virtual production. In contrast to that head-mounted displays may allow a more natural and immersive exploration and help to find new visual aesthetics for specific procedures like omnidirectional or stereoscopic recording, while handheld display devices like tablets seem to comply with the attempt to enable everybody on set to examine the real and virtual stage elements from any position within the soundstage, ideally having all the creative professionals using their own smartphones and tablets. Moreover portable devices allow the user to move around the objects and therefore support a three dimensional observation of the environment, being most beneficial when relying on depth-perception during repositioning and animation process.

3.2.3. Selecting Devices

Since most of the transformation tasks within a virtual production environment require a concurrent access to all three spatial axes, only those devices come into consideration that have at least three degrees of freedom. Furthermore, when deciding on an appropriate input or output device, it is important to keep in mind that a lot of novel systems might be suitable for the tasks on a film set but can in fact not be applied at all without a sophisticated software solution. They are just hardware and constitute only a minor part of a completely operable interface.

Device Survey

Before choosing a suitable device one has to get an overview of the currently available systems. The company Sixense is specialized on novel input devices for virtual reality applications and has brought two advanced systems to the market by now. Intended primarily for innovative game control, the products may be of interest to virtual production environments as well. The Razer Hydra consists of two 6 DOF motion sensing controllers for tracking the movement of both hands concurrently. While the devices travel through a magnetic field generated by a stationary transmitter, a charge is induced that can be interpreted to determine the position and orientation of each single controller. This electromagnetic tracking technique guarantees an accurate measurement down to a millimetre but is also highly susceptible to interferences (Steinlechner 2011). A non-commercial project recently demonstrated the capabilities of the system by using the Razer Hydra as input device for the Source Filmmaker, including camera control and puppeteering for virtual characters (Cox 2014).

The second and indeed much more promising device from Sixense is the Kick-Starter funded STEM System. Similar to the Razer Hydra, several portable sensors are tracked inside a magnetic field. However this setup is not only capable of recording the hands with two controllers but supports up to three additional wireless motion tracking modules that can be attached to any part of the body, allowing a full body tracking (Krause 2014).

Fig.17: Razer Hydra (right), Sixense STEM system (left)

A tablet, not matter which brand, offers a camera-fed window into the real world and constitutes an ideal version of a virtual camera system packed into a most compact shape. The multi-touch display receives user input and screens graphical output at the same time, making this device an extremely flexible and intuitive piece of hardware. Since tablets come along with an integrated monoscopic camera, there is no need to feed in video footage from a supplementary system. When additionally using advanced real-time tracking procedures for detecting naturally occurring features it is even possible to localize the tablet without falling back on external hardware like optical outside-in systems that would require a bulky and expensive setup. Embedded gyroscopes measure the orientation in case the optical tracking fails. In addition to that, a tablet appears especially suitable for virtual productions because it is inexpensive, easy to handle and present anyway. If every department on set had a tablet at its disposal, it would be easy to share ideas and suggestions as well as to explore and modify the virtual world – a teamwork that complies perfectly with the idea of a collaborative environment.

The Leap Motion[10] controller is a tracking sensor which has been developed by the American company Leap Motion and consists of three infrared LEDs and a pair of monochromatic cameras, capturing up to 200 stereoscopic images per second. Using a proprietary motion-tracking software the recorded data is processed to measure the distance of the hands as well as to track their movement and gestures close to real-time with submillimetre accuracy (Bachmann, Weichert, Rinkenauer 2014, p.214). Michael Buckwald, CEO of Leap Motion, explains that his company has started working on this intuitive gesture recognition sensor because “too many apps today involve just one finger or the whole hand. […] The developer should be able to focus on creating a physical experience and not have to think about the finger tracking or the hand tracking” (quoted by Bell 2014).

Fig.18: Leap Motion controller

The current version enables the user to literally mirror his or her hands into the target application, allowing a natural and intuitive human-computer interaction. Even if the tracking volume appears comparatively small, the overall accuracy totally compensates for that. When for instance

10 https://www.leapmotion.com

performing simple pointing tasks, the error is even below the human hand tremor. However the system is not fully devoid from constraints. Since the sensor measures the position and posture of the fingers by illuminating the nearby region with infrared light before handing over the resulting differences in luminance to the software for further computing, the entire system is very sensible to disturbing glares and thus works best in dark environments.

The Oculus Rift[11] DK2 is exemplary for the wide range of head-mounted displays pushed onto the market by various companies right now. Starting out as a Kick-Starter funded project, the device is currently developed and distributed by the Facebook Company. The Oculus Rift consists of a closed case holding a full-HD resolution OLED screen that provides a field of view of about 90 degrees (Desai et al. 2014, p.176). Special plano-convex lenses placed above the screen create a stereoscopic view and allow the user to focus on the image plane, which would otherwise be too close for a sharp vision. Both orientation and movement of the head are tracked using the embedded gyroscopes and accelerometers, while an additional external camera can be applied for measuring the position of the Oculus Rift within a short range by tracking the infrared markers on the casing. The HMD is meant for use cases, which enable the user to explore an entirely virtual world, while augmented reality applications are by default not supported, as the setup does not provide any solution for combining digital and real images.

Fig.19: Oculus Rift DK2

11 https://www.oculus.com

Device Selection

After having analysed the characteristics, advantages and shortcomings of different input and output devices, one has finally to decide on suitable and well matching systems. By now, there are several tablet-based tools for virtual production environments available, which already include some of the features intended for the prototype to be developed within the scope of this thesis. Like this, the SmartVCS enables an operator to load and adjust objects, using the Playstation Move controller for tracking the tablet in space. The Zeus Scout iPad application offers similar functionality, even if not fully complying with the idea of set editing. Consequently, when working on a tablet system, the resulting prototype would definitely lack originality.

When aiming for something completely new, the Oculus Rift appears to be a bad choice as well. Right now, head-mounted displays arouse a lot of public interest, thus every small start-up or pseudo-innovative media agency seems to be willing to contribute to that hype. Hundreds of companies are developing virtual reality content or games while also the film industry begins to produce omnidirectional videos for this new type of display device. However a HMD has not been applied for advancing a virtual production so far. In reply to the question of whether virtual reality HMDs like Oculus will affect the virtual production field, Girish Balakrishnan, developer of the SmartVCS, states that he is “a believer of virtual reality and […] curious in exploring how these technologies could be integrated into the filmmaking process.” Further he explains that “VR is yet another tool that enables a director to literally walk and scout in the world they are developing, months before ever stepping on set” (quoted by Workman 2014 a).

Due to their limitation on planar tablet screens, multi-touch based approaches lack the necessary degrees of freedom for manipulating objects in three-dimensional space (Nitsche, Kirschner 2013, p.304), whereas 6 DOF devices like the Razer Hydra or the Leap Motion controller enable the user to access all three axes simultaneously. Since the novel prototype constitutes an attempt to empower DOPs or directors to work in virtually extended scenes without professional supervision, a gesture-driven interface based on a Leap Motion sensor appears especially suitable. If succeeding in designing an intuitive system, the decision makers on set will be able to make modifications in a most natural and familiar way, namely by tapping on objects and controlling them with a simple movement of their hands.

Taking everything into consideration, Oculus Rift and Leap Motion both constitute the tools of choice for contemporary virtual reality projects and 3D interfaces. An application beyond their original scopes may be of particular interest, finding out how these devices can benefit a virtual production environment.

3.2.4. Tablet Mockup

When integrating novel interfaces, an early outlook on the intended range of options may appear helpful. Mockups originally refer to scale models or replicas which are built for presenting machines and technologies without including any functionality of the later prototype. Considering software engineering, graphical mockups preview designs for web pages or applications and enable the reviewer to get a first picture of some elements of the GUI while not providing any interactivity. Hence, workflows or movements may either be promoted by videos or completely left out. Since only Leap Motion and Oculus Rift were meant to be fully integrated and tested in an exemplary virtual production, the tablet should at least prove its suitability when examined in a mockup. Several versions have been elaborated as video clips in After Effects, describing the different scopes of application. The graphics were proposed to look similar for all devices and could therefore be reused for the design of the final Oculus user interface. Furthermore, the analysis of the mockup helped to recognize which GUI elements worked and which not, offering some indication of problems connected to the layout in terms of user-friendliness and clarity. In the beginning some hand drawn scribbles helped to create an appropriate arrangement of GUI elements and could afterwards be made accessible to the team members for discussion. The mockup proposes a tablet GUI that tries to keep the screen space as unspoiled as possible. In idle mode, when no object is selected, the graphical elements are reduced to zero, guaranteeing an image uncompromised by occlusion. When tapping on a virtual entity a circular menu appears, providing various buttons, whereas the number of options depends on the type of object. All menus and buttons work without any written information but show only descriptive icons.

Icon Design

The majority of icons created for the mockup bases on commonly used graphics and pictograms, which can be discovered in most of the contemporary software packages and enable the creative professionals to start working without any further familiarization. Thus the usage of these well-established visuals seemed advisable. However the icons have been slightly modified to match both the overall concept of the prototypical virtual production environment and the corporate design of the fictitious company that shows up in the video clips. The main colour is a warmish yellow, framed by a thick black outline, whereas a dark grey adds to the colour palette for more complex icons with several overlapping elements. All graphics turn white when selected.

Fig.20: Icon design

Navigation

The introduced mockup assumes a position tracked tablet. Probably such a device would combine optical motion capture approaches with inertial sensor-driven techniques, deducing the position and orientation of the tablet in three-dimensional space from the gathered data. Anyway, the user is capable of manoeuvring through the scene just by walking and looking, additionally recovering the possibility to explore his virtually expanded surrounding in an independent and natural manner.

Selection

Considering a tablet, a basic tap on the display definitely constitutes the most obvious and intuitive practice to select 3D models or GUI elements. When approaching a nearby object a yellow outline accentuates the silhouette and indicates that the item is accessible at all. At the same time the user gets to understand that a further touch will lock the selection on the highlighted object. In case elements occlude each other, the object which is located closest to the camera will be favoured.

Set Editing

When a geometrical set element or a character has been picked with a finger of the right hand, a white outline illustrates the altered status of the object. The concurrently appearing menu offers five different buttons, namely translation, rotation, scale, animation and annotation. After selecting one of the transformation options, the menu disappears and a contextual 3D widget pops up, which indicates the chosen axis by highlighting. As the design of these widgets is well-known and follows conventions that have been approved and commonly applied across the range of DCC tools for years, there is no need for poor-quality alternatives. By sliding over the display, the virtual object can be repositioned, rotated and scaled. In order to access all three axes independently the three-dimensional widget has to be somehow addressed within the limits of the two dimensions of the touch-sensitive surface. One approach suggests adding three more buttons for discrete axis selection. A finger of the left hand then constantly locks an axis while the right hand moves horizontally to transform the object accordingly. However this method allows only one degree of freedom at a time and thus, forfeits the opportunity of a concurrent control in three dimensions. Another attempt to include the z-axis consists in the integration of the pinch gesture, which is natively supported by all current tablet types and mostly applied for zooming. When using this additional input specification, the objects are positioned, rotated and scaled on the x-axis when moving horizontally, on the y-axis when shifting vertically and on the z-axis when pinching. When an object has come to rest at the intended position, thus successfully transformed, a short tap on an empty area of the display resets the selection and induces the tablet to return to idle mode and a blank screen.

Animation Editing

Imported 3D objects may contain more or less complex animation, which requires spatial and temporal fine-tuning in real-time. When tapping on the animation button the application switches to animation mode and previews the trajectory of the selected object as red curve alongside a row of key frames. In order to manipulate a single key frame, a submenu offers options for removal, positioning and retiming. Furthermore the mockup suggests a button for inserting new key frames anywhere on the curve. A fourth option leads to a dope sheet view where the timing of animation data can be adjusted with frame-accurate precision. Once the positioning tool has been selected the keyframe is repositioned according to the object transformation method.

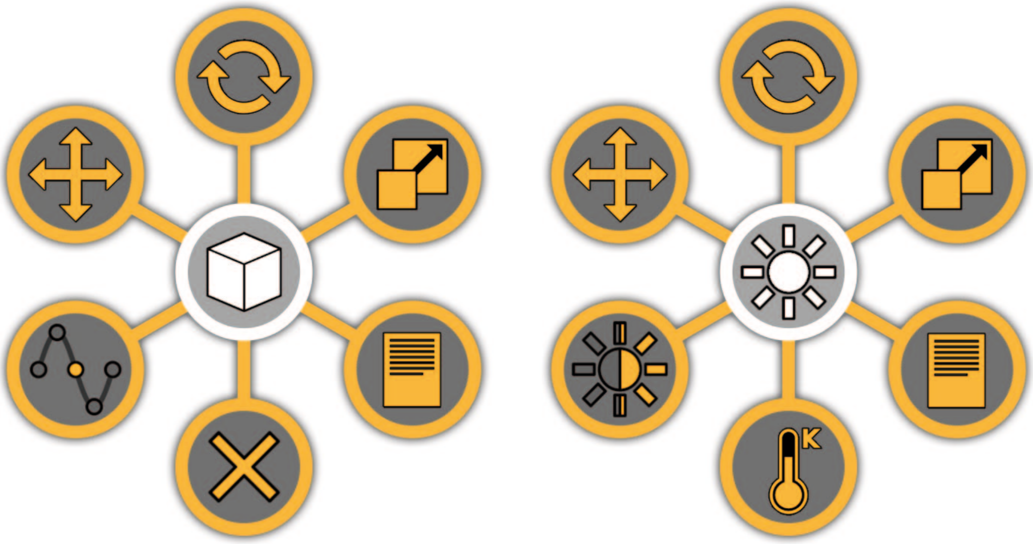


Fig.21: Object menu (left) and light menu (right) for mockup

Light Editing

In the mockup every light source is permanently previewed as a 3D gizmo, clearly summarizing the most important parameters, like bulb centre position, orientation, range, cone angle and colour. Since the light may be transformed analogous to objects, while rotation realigns the cone of a spotlight and scaling alters the overall range in case of a linear fall-off, the three transformation buttons can stay the same. Also the annotation is kept. Additionally, the menu now provides two new intrinsic light options – light colour and intensity. When picking those with a finger of the right hand a context-dependent slider appears at the upper edge of the display. The colour slider indicates the current light temperature in Kelvin, ranging from red to blue, additionally providing the numerical values. By dragging horizontally the user is able to adjust the temperature. A switch to RGB colours is reasonable but not yet included in the mockup. The light intensity is manipulated the same way, whereas a greyscale slider delivers the necessary real-time feedback.

Annotations

Annotations may appear reasonable when different people work within one scene in succession and wish to document the changes and modifications they have dealt with. The mockup introduces an option for appending notes to single objects. When tapping on the correspondent button a separated frame appears displaying the current annotation. A pencil icon guides the user to an editable version of the text field, which comes along with the natively supported display keyboard. Naturally a save option stores the changes and registers the time the user has last accessed the annotation.

Slide-In Menu

When approaching the right edge of the screen, another menu slides in to reveal general options for loading objects, changing the perspective and switching the editor mode. Once the insert button has been touched a dropdown menu features various prefab elements, which can be imported into the scene right away, while the well-known folder icon signifies the possibility to access custom models somewhere on the hard drive. A loaded object remains attached to a finger and can be dragged across the screen until it is placed with a tap. The perspective button enables the user to choose between the live view of the tablet camera and an orthographic editor camera. Finally the editor mode switches to cine mode when using the third button, resulting in a simplified state where manipulation and transformation of objects and lights is impossible, whereas new buttons at the upper edge provide playback, rewind and pause functions for screening animation.

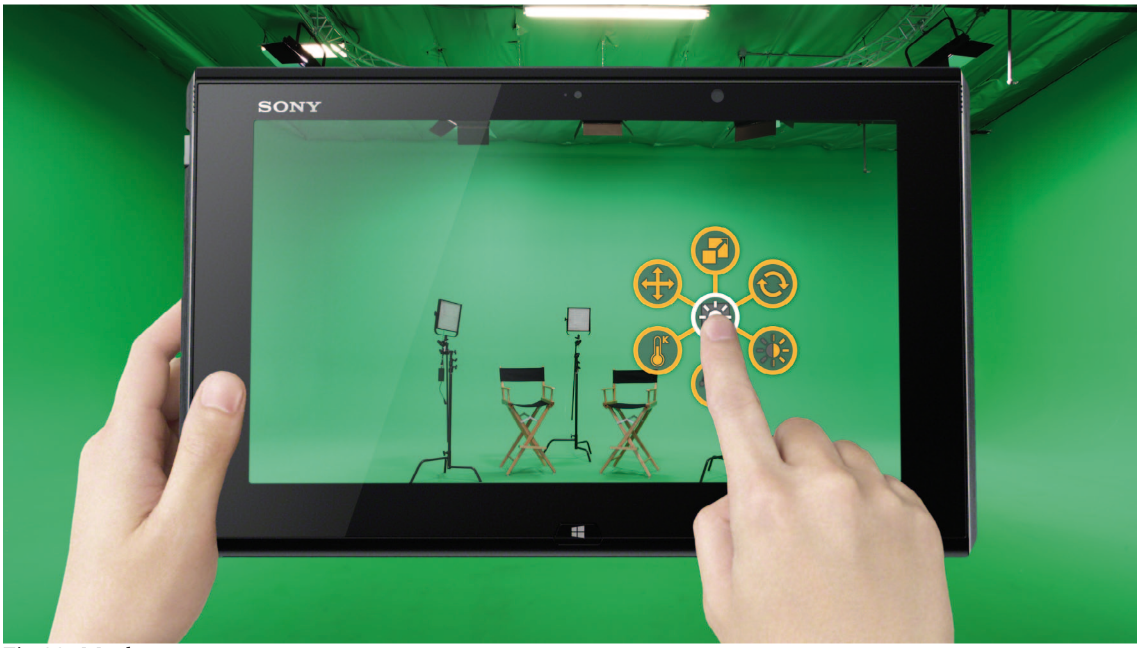


Fig.22: Mockup

3.2.5. Interface Prototype

When designing a prototypical interface for an extended virtual production environment, one has to keep in mind that the main objective still is to provide a set editing tool for people who are not digital natives and therefore unfamiliar with procedures and interaction techniques in three-dimensional space. The real-time editing tool for virtual productions, herein after referred to as RET-VP, consists of an augmented-reality version of the Oculus Rift head-mounted display and a Leap Motion gesture recognition sensor, thus enabling a user to transform objects by selecting and shifting them with movements of the hand while looking into a virtually augmented surrounding.

The HMD serves as an input and output device and has been upgraded in two different aspects. First of all a Microsoft Studio webcam[12] is mounted on top of the Oculus Rift body, expanding the idea of virtual reality by introducing a combination of real video footage with virtual elements from the user's point of view. Thus the device is actually used for creating an augmented reality, where images from the real world are furnished with digital content. Furthermore the setup is provided with four retroreflective markers which make it possible to localize the HMD continuously using the OptiTrack mocap system. Like this the user can navigate freely inside the mocap volume while not relying on the short-range tracker coming with the Oculus hardware.

Fig.23: RET-VP prototype: (left to right) Leap Motion controller, Oculus Rift with webcam, complete setup

The Leap Motion sensor has been fixed to an angled mount, which is then in turn tied to the user at waist-height, allowing a comfortable recognition of gestures while guaranteeing an unrestricted mobility. This setup constitutes a self-made variation of the original attachment to the Oculus Rift, enabling the user to turn his head without having the hands leave the Leap Motion tracking volume – a big advantage when applying the gesture recognition sensor as control tool separately from the HMD as navigation device. Both hands are required to operate the interface, either to navigate through the menus or to select and transform objects, while CG equivalents of the user's hands act as 3D cursors that constantly represent the current posture and provide a feedback of the

12 http://www.microsoft.com/hardware/de-de/p/lifecam-studio

tracking quality. Since the Leap Motion sensor does nothing but track the postures of both hands, the device theoretically has 54 DOF, 27 for each hand. However, in practice one has to take into account that the fingers can only be abducted or adducted to a quite limited extent, while also the wrist is definitely not capable of rotating freely (ElKoura, Singh 2003, p.112). Anyway, to make use of the full range of movement, some extremely sophisticated recognition would be required, whereas the current implementation into Unity only provides basic gestures. The interface thus applies the Leap Motion as a 6 DOF device, encompassing only all three translation axes and two modifying gestures for each hand. 3D entities are selected by pointing with an infinite ray that prolongs the virtual index finger of the right hand. When hitting the target for one to two seconds, depending on the framerate of the real-time engine, the selection is locked on the object and immediately indicated by a white edging. At the same time a button pops up which says 'edit'. When hovering above it with the left hand, a circular 2D menu appears, comprising options for translation, rotation, scaling, animation, annotations and escape, whereas only the object transformation has been fully integrated into the prototype until now.

	navigation		selection		transformation		
	exploring	manoeuvring	selection	deselection	translation	rotation	scaling
input	optical tracked head rotation	optical tracked body movement	ray-casting with right hand	right hand out of volume or escape button	movement of the right hand clenched to fist	movement of the right hand clenched to fist	movement of the right hand clenched to fist
output			infinite ray at index finger white edging	white edging disappears	3D translation widgets	3D rotation widgets	3D scaling widgets

Fig.24: RET-VP interaction techniques

Pushing an icon away from the centre of the menu, the desired interaction technique gets picked and a 3D widget shows up, representing the orientation of the object in space and indicating the current transformation mode. The user is now able to modify an asset according to the chosen setting by clenching the right fist while moving the left hand through the volume. Travelling from left to right results in a transformation along the x-axis, moving forth and back alters the z-value and lifting or lowering the hand addresses the y-axis. The modification is confirmed when the user either enters the submenu 'axis' with the left hand and picks the 'ok' button or takes the right hand completely out of the Leap Motion tracking volume. Furthermore the 'axis' menu contains options for disabling certain axes, increasing the transformation scale, switching over to local coordinates and undoing the latest changes. By using ray-casting for selection and virtual hands for manipulation, the RET-VP prototype complies for the most part with the hand-centred object manipulation extending ray-casting technique, which was first recommended by Bowman and Hodges (1997, p.38).

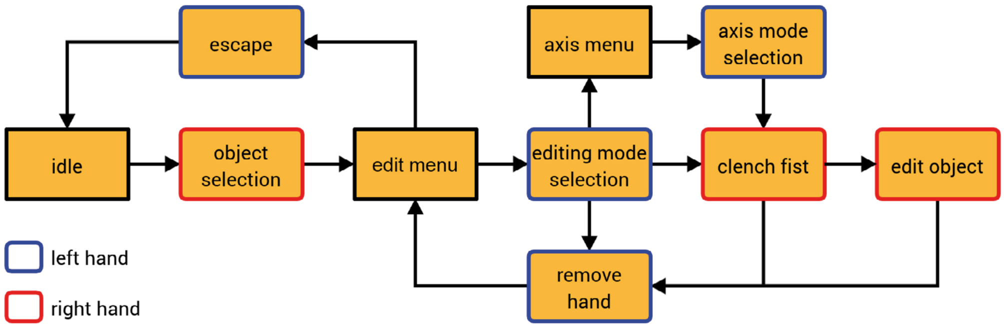


Fig.25: Interaction and menu scheme for set editing

Another important task was to decide on appropriate transfer functions, as the interface has to ensure a convenient interaction persistently. When navigating through the menus or selecting objects, the movement of the real hands is transferred to the correspondent virtual hands by applying a zero-order mapping. As soon as an asset has been picked, the transfer function switches over to a motion-dependent mode. Shifting gently the user is still able to control the position of the target directly, while an abrupt movement assigns a velocity that causes the object to move on in a straight line with constant speed until the transformation is stopped by unclenching the fist.

When making modifications to the scene, the affected parameters are updated in Unity in real-time. Already in the beginning it was obvious that the Unity game engine would neither provide state-of-the-art render results nor guarantee the smallest possible latencies. Nevertheless this environment appeared most suitable for designing a novel interface since it offers the possibility to access hundreds of ready-made plugins for integrating established hardware systems and thus obviates a time-consuming preparation in advance of the actual development. In addition to that, alternative real-time engines like Crytek's Cryengine[13], the Unreal Engine[14], Shark 3D[15] or Motion Builder[16] cannot come up with a comparably large and committed user base or such a big amount of examples and code snippets free of charge. Taking everything into consideration, Unity offers an ideal framework for developing rapid prototypes in which functionality stands above graphical quality. On set the interface is primarily applied for modifying virtual elements in an environment, which can be seen through the HMD. Consequently the outcomes of the real-time rendering are transmitted back to the display inside the Oculus Rift and provide the user with the necessary feedback. While shooting, there is usually no need any more to comply with change requests and the video signal can be forwarded to the viewfinder of the principal camera, allowing the DOP to scout the environment and frame the action by consulting the composite images that consist of real and virtual elements.

13 http://www.crytek.com/cryengine
14 https://www.unrealengine.com
15 http://www.spinor.com
16 http://www.autodesk.com/products/motionbuilder/overview

3.3. Experimental Production

In order to be able to evaluate the benefits and shortcomings of innovative devices and graphical user interfaces it is advisable to test them in a realistic but experimental scenario of a filmic production.

3.3.1. Test Scenario

Before forming a story concept and preparing the shooting one has to define what the experimental production is actually meant to reveal and how the single tests can be evaluated afterwards to derive the required information. Five different key aspects are to be examined: novel interfaces, motion capture workflows, NCam utilization, scalability of the virtual production methodology in common and its advantages for certain departments.

Novel Interfaces

Since the prototypical interfaces have been designed for simplifying the modification of objects, lights and motion paths in real-time on set, the test scenarios mainly cover use cases for set, light and animation editing. Professionals with different levels of experience and knowledge use the devices for performing various tasks, which they might also face on a real film set. Like this computer generated objects are selected and repositioned by moving them with gesture control through the virtual scene, at the same time rotating and scaling them to completely meet the given requirements. Such relocation may become necessary when the virtual scene has to adapt to altering movements of actors or real stage elements or when the decision makers decide that an object would just look better in a different position. Likewise virtual light sources are to be adjusted and pushed around until they sit perfectly on a real spotlight and match the equivalent colour temperature. During the transformation of assets, lights and keyframes, the head-mounted display allows a three dimensional perception and should speed up the navigation and spatial coordination in theory. Furthermore the tests reveal whether the pictographical user interface is as well-arranged and intuitive as expected. The quality of an interface can only be evaluated when rating precision, speed and intuitivity of navigation, selection and modification. While accuracy and speed constitute measurable indicators that can be quantified quite easily, the intuitivity only allows personally statable conclusions, which appears acceptable however, since the research focuses on a qualitative subsumption of user benefits. However even if acquiring measurement values, there is nothing to compare the data with. Thus the user group has to perform the tasks by first accessing traditional input and output devices, namely mouse, keyboard and standard desktop monitor, before getting to know the novel interfaces. For both test series, speed and precision of operation are measured, producing values that stand by afterwards for being compared and classified. Questionnaires enable the testers to reflect their impressions and verbalize their thoughts and suggestions, thus offering a more or less reliable feedback about how intuitive the handling actually is. Since the user group is rather small the results might lack significance. Moreover it is unfortunately impossible to narrow down the scope of analysis to get an idea of the quality of a single specific feature because the different components influence and affect each other. The evaluation of the graphical user interface is for instance always distorted by the user's opinion about the handling of the devices. If the Oculus Rift nauseates a tester or the gesture recognition fails, one cannot expect to get a positive response when talking about the GUI icons. Nevertheless the test results illustrate a representative tendency at least and allow an estimation of whether the present state of development heads for a promising future.

Mocap Workflows

Beside the evaluation of interfaces different motion capture approaches are to be examined in consideration of their respective advantages and domains. In the first run the actor and the camera are measured at the same time within one single tracking volume. Such a simultaneous mocap test requires sophisticated preparation and focused coordination on set and appears only reasonable when the camera records live action footage concurrently, since the sole advantage of this methods consists in the actor's ability to let the virtual creature interact with real objects or people. The CG character is rendered in real-time exactly atop the video footage, covering the mocap performer. Like this two actors can for example hand over a glass or bump into each other, while the set crew is able to preview the composed material with one participant replaced by the virtual counterpart. Simultaneous mocap can definitely be considered as the most challenging approach but has been included in the test scenarios, as interaction is just indispensable. The second method to be examined is parallelized mocap and describes another one-pass approach, in which both camera motion and performance are captured at once but in different locations. Thus the tracking volume remains spatially separated from the area the camera is shooting in, what may become crucial when there is not enough space for setting up a mocap system in the spot the virtual character is meant to be. If interaction is not required while the filmmakers have some time to spare, the third variation, serial mocap, is strongly recommended. Here the actor and the camera are captured both spatially and temporally separated, providing maximum flexibility and creative freedom because the prerecorded performance capture can be replayed as often as necessary until the camera operator achieved a prefect shot. During the experimental production the three different approaches are tested in realistic scenarios one after another to evidence the advantages and challenges in practise.

NCam

Furthermore, the NCam system, which has been purchased by the Institute of Animation quite recently, is applied and tested in detail. The hardware is thus brand-new and used for the first time within the scope of these experiments. During the tests the NCam bar provides preprocessed tracking data which is instantly fed into the Unity pipeline for emulating the parameters of the main Alexa camera. Afterwards both lens calibrating procedures and final tracks are evaluated and compared to established approaches.

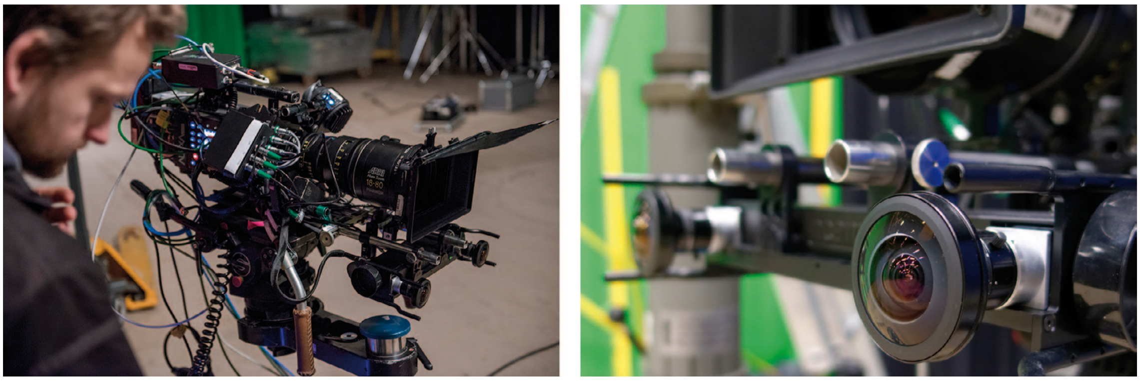


Fig.26: Arri Alexa with NCam system

Scalability

Until now virtual production methods are usually only applied for improving the pipeline of big blockbuster productions with large-scale visual effects, although the creation of less expensive independent films, television broadcasts or commercials may profit from an appliance of the innovative workflow as well. 'Avatar', 'Real Steal' and 'Tintin' have already demonstrated the potential of virtual production techniques on a grand scale, while it is still questionable if the approaches can be applied for small films, like student projects, at all. Thus not only the operational reliability of the interfaces is to be examined but also the benefits a virtual production environment might have to offer to a less narrow circle of filmmakers, finding out how far one can get with reasonably priced devices.

Advantages

Most of the advantages, especially those in the fields of direction and cinematography, have only been theoretically deduced or gathered from papers and books so far and must now prove true in the heat of production. The tests reveal to what extent the idea of an iterative and collaborative environment agrees with reality, again using the questionnaires for obtaining feedback and opinions but also considering own impressions.

3.3.2. Production Scenario

After having decided what to find out, one has to determine an appropriate framework. To gather the required information it would be totally justified to perform the tests without making further efforts to build up an entire production scenario. However virtual production is, though driven by technology, not a merely scientific field of research but a playground for design and artistic work, which has to be experienced to understand its essence. Thus one cannot come up to the actual idea behind the novel workflows when sticking to hypothetical derivations or tests which have little in common with real filmmaking.

Technical Framework

A scenic production including all stages from concept art to shooting and compositing helps to release the tests from theorization. Due to the tight schedule only parts of the filmmaking process can be carried into execution using virtual production principles while most of the tasks are unfortunately performed the traditional way. Like this there is no previsualization from which assets or animations could be handed over to the visual effects department. Furthermore the artists in postproduction cannot access the trajectories and positions of objects which have been manually modified in real-time on set, since the framework is not capable of storing user-driven manipulation until now and can thus provide only raw mocap data and camera footage. At least the work on set follows the virtual production guidelines to the greatest possible extent.

The shooting takes place in studio 1 at the Filmakademie Baden-Württemberg. 24 motion capture cameras from OptiTrack are mounted onto a hanging gantry and cover a square volume of approximately 5 by 5 meters, fortunately not requiring additional stands which might compromise the tracking result. This optical outside-in mocap system is here intended for measuring the actor's performance but can also be used for localizing the RET-VP editing device, thus the Leap Motion controller and the Oculus Rift HMD. Three ethernet switches ensure a bundled data transfer and connect the infrared mocap cameras with a central workstation, where the information is processed inside Motive, the proprietary control tool of OptiTrack[17].

Fig.27: Studio 1 at Filmakademie

17 http://www.optitrack.com

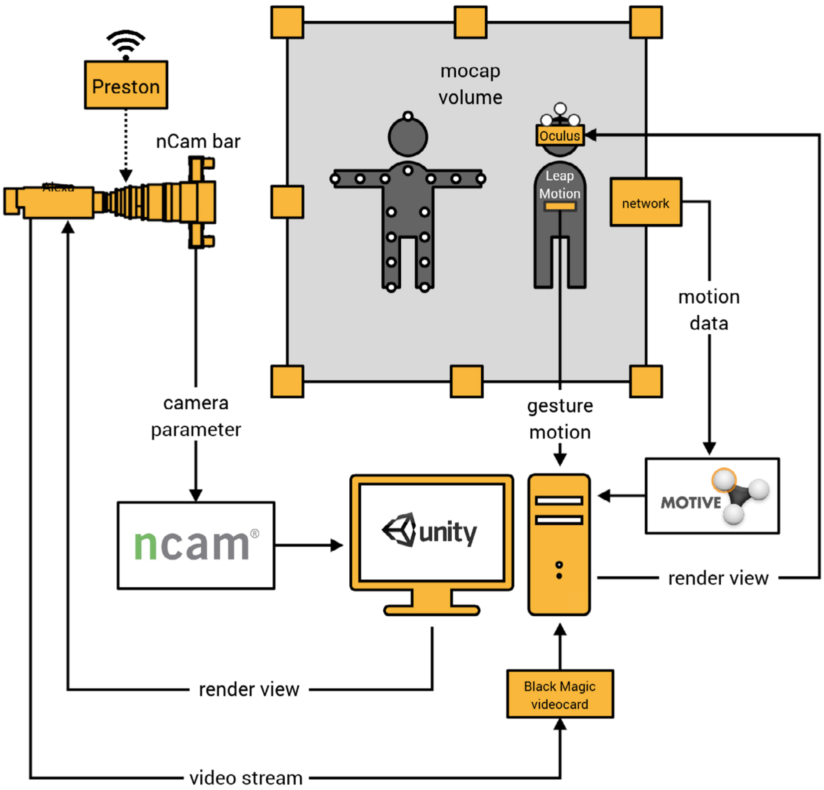


Fig.28: Hardware setup on set

The main camera, an Arri Alexa[18], is localized by the NCam system, which allows a markerless and flexible tracking. As this technique requires special software for computing, the camera parameters are preprocessed in the NCam server before being passed on to the real-time engine. All gathered motion data is then assigned to the correspondent assets in Unity, which means that the mocap performance has to be handed over to the character model while the measured camera parameters are applied to virtual cameras, representing the point of view of either the Arri Alexa or the webcam mounted on top of the HMD, depending on the use case. The video footage from both cameras is fed into a video card and applied to a plane that remains fixed to the virtual cameras. Furthermore the user input from the Leap Motion controller is translated into a correct transformation of virtual objects before the entire computer generated scene is rendered atop the video footage and sent back to the Oculus Rift display or the Alexa viewfinder. In order to minimize the delay and ensure a feedback close to real-time, the setup inside Unity, which definitely constitutes the bottle neck of the entire procedure, has to be reduced to a simplified version without complex reflective materials or additional video-streams for backplates. Nevertheless the simultaneous application of all these different techniques appears without any doubt challenging enough.

18 http://www.arri.com/camera/alexa

Narrative Framework

The experimental environment is supposed to include a background story, one protagonist and a basic plot. These fundamental components of every film do not only help to achieve a close approximation to real filmmaking conditions but also guarantee an outcome that lives up to very own ambitions. In several short clips the camera robot CamBot Mark 2, the latest invention of the company Camdroid, is presented in action on a film set. Despite the promised qualities the robot proves to be fully incompetent and fails miserably. In the beginning of each episode, a commercial promotes the CamBot as technical masterpiece and essential innovation, which is meant to change the way of filmmaking for good. Absurd infographics and pseudo-scientific formulas are to suggest expertise and trustworthiness. Suddenly an interference disrupts the video and reveals what is actually happening on set. The spectator finds himself inside a studio – spotlights have been set up and a camera stands by to observe the procedures. However the CamBot is missing. The hand-held camera pans around nervously, trying to detect the robot, and locates it during an attempt to leave the room while bumping into the wall over and over again, incapable of finding the door. In the second episode, a shot from the robot's point of view shows that the CamBot is about to open a suspiciously named email attachment. As expected, a virus gets installed immediately, thereby making the robot dance weirdly. Another point of view perspective in the third episode illustrates how the android permanently chooses a wrong framing, either cropping the actors or focusing on something completely different. After a while the spectator realizes that it is the catering table that attracts the robot's attention. Finally the CamBot walks there and starts flirting with a toaster. In the fourth episode the robot obeys and switches over to record mode subsequent to the director's obligatory 'action' command. However, the image freezes after a second and the well-known progress wheel indicates the installation of an update. In another fail situation the CamBot is again not where it is supposed to be but sits on the floor, folding its arms huffily. As soon as the robot becomes aware of the camera it jumps up and leaves the soundstage in rage, rudely bumping into a boom operator. Of course professional image effects are included in delivery, at least that is what the commercial promises. In reality the CamBot comes along with a rather disappointing package of cheesy or completely immoderate effects – a small compilation of these visual missteps can be seen in the sixth episode. Finally the camera observes the robot failing in the attempt to climb a single step of some staircase. As the point of view changes slightly the spectator realizes that the step is in fact some meters away from the spot where the helpless machine supposes it to be. Hence the depth perception and the GPS-based positioning of the CamBot require optimization urgently.

While coming up with entertaining slapstick comedy, the episodes touch a more serious subject as well by visualizing the risks of an excessive mechanization. They promote traditional skills and workmanship on a film set – no robot or machine will ever substitute human expertise. Every single episode thus concludes with the claim 'there is nothing like true professionals'. Superficially this statement appears contrary to the idea of virtual production, since the related workflows build on technology-driven optimization as well and might indeed supersede some formerly indispensable jobs. Nevertheless virtual production environments in fact require well-trained staff and help the decision makers and creative professionals to regain control.

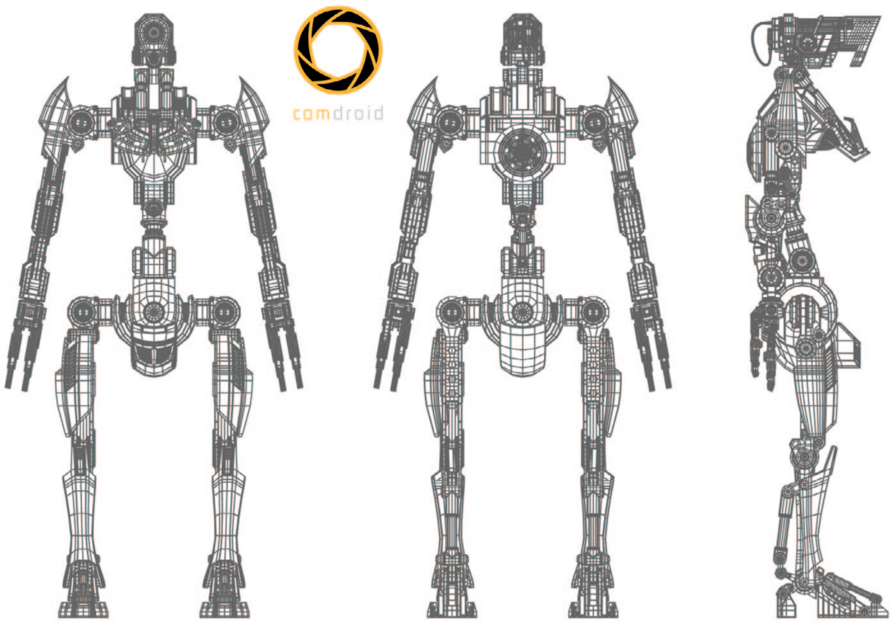


Fig.29: CamBot wireframe

3.3.3. Project Management

Since the workload appears extremely challenging throughout the entire production a sophisticated project management becomes indispensable. The core team consists of only two students and is therefore quite easy to oversee. However a vast number of totally different tasks have to be coordinated in an efficient way, allowing consecutive operations, ideally without any idling. Above all one must not disregard the fact that first and foremost two complete scientific theses are to be written at the same time, while the scenic production actually constitutes just a use case of the theoretical work, no more, no less.

Team

The project would not be possible without the close collaboration of students from completely different fields of media production. Stefan Seibert is a bachelor student of Computer Science and Media at HdM Stuttgart and contributes not only his experience in programming and software engineering but also scientific and mathematical knowledge, which allows him to approach problems analytically. Kai Götz studies Electronic Media in the master program at HdM Stuttgart and looks back on several years of practical work in film production, especially focusing on computer animation and visual effects, thus being aware of the challenges and possibilities of the creative filmmaking process. Temporarily, Sebastian Moreno, exchange student from Gobelins, supports the team with his skills in drawing, rigging and animation. The entire production is supervised by Simon Spielmann, research associate at FAAI, while Volker Helzle, head of research and development at FAAI, and Katja Schmid, professor for VFX and postproduction at HdM, provide advisory support as well.

When working on set, the small core team can certainly not perform all incoming tasks. Hence, additional assistants or specialized professionals are required. Although striving for an enjoyable scenic production the project still focuses on the experimental application and evaluation of interfaces, both device-related and graphical. Therefore, the staff on set is reduced to the positions, which appear most relevant for the tests, namely DOP, DIT and software engineer. The production even has to get along without a director, since priority is not given to an ambitious exhibition of skills but to a scientific analysis. Moreover it is not necessary that a real actor makes the robot come alive, even though it is obvious that the quality of motion capture depends on the performer's talent. However, a shooting with professional actors requires planning to an extent that does not meet the idea of an iterative test environment in which one can experiment freely and react on problems without time pressure.

Schedule

Right at the beginning of the project a schedule (see Appendix: Thesis and Production Schedule) has been set up, which visualizes the weekly tasks and highlights the distribution of work. Even though the various stages of production differ in terms of volume, complexity and scope, the plan tries to spread the workload as fair as possible. As the team goes for a complete scenic production, additional tasks come along, which have to be accomplished prior to the actual tests. Starting with brainstorming and plot definition already in August the goal actually was to finalize all seven episodes until the delivery of the written thesis by the end of February. However, due to a delayed shooting the entire postproduction had to be postponed until the month of March, what seemed to be a bit of a setback at first but later turned out to be quite beneficial since that way there is enough time for evaluating the experiments in detail.

Despite all planning and scheduling the application of innovative technology remains always partly unpredictable. Some of the techniques could not be tested sufficiently in advance while several devices and implementations had not been ready for use until shortly before the shooting. Consequently the experiments were prepared as well as possible to meet even totally unexpected challenges.

3.3.4. Preparation

In order to be able to carry out a complete albeit small on-set virtual production, an extensive preparation has to be taken into account. Thus, one has not only to create all kinds of visual and narrative content but also to prepare the innovative prototype as well as the shooting, while meeting dozens of challenges in all stages of production.

Tasks

After having decided on a story idea and a plot, moodboards and concept arts are elaborated to advance the design of the robot. While specifying the requirements of the novel interface, the first 3D assets are modelled, unwrapped and prepared for an early export to the real-time engine. Subsequent to that, one focuses on the prototype for real-time editing.

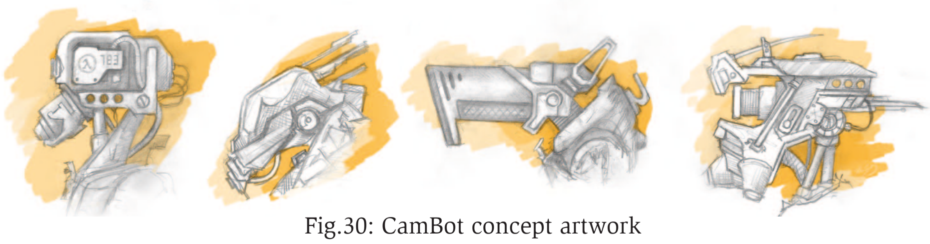


Fig.30: CamBot concept artwork

First and foremost, the features of the RET-VP are to be implemented and thoroughly tested. The period before the shooting has therefore been divided into four phases of development, each one describing an essential milestone in the preparation of the interface. In the first stage basic 3D editing tools are integrated by falling back on a simple ray-casting approach for both selection and manipulation. Furthermore one intends to accomplish the SDI video feed as well as the motion data stream from the NCam server to Unity. The second phase deals with the question of how an actor's mocap performance can be transferred to a virtual character. At the same time the elements of the graphical user interface are designed and afterwards visualized in a mockup, trying out various button arrangements and shaping descriptive icons. As soon as the rig of the robot corresponds to the standard setup inside Motive, including a mirrored coordinate system and a proper rotation order, it is possible to assign some prerecorded motion data to the geometry. In the third phase, the software engineer is primarily concerned with the gesture control and integrates the Leap Motion functionality into the Unity framework. After the GUI has been added to the software prototype in the fourth phase, a special augmented-reality version of the HMD is prepared by mounting a webcam to the Oculus Rift. With the construction of a waist belt that attaches the Leap Motion sensor to the user, the development of the RET-VP interface is finally completed. As soon as the development of the interface draws to a close, the shooting has to be prepared. For this purpose a special schedule summarizes the tasks, which must be accomplished before the beginning of production. The team elaborates a detailed shooting plan, compiles a list of equipment to rent, hires supporters, sets up a detailed evaluation plan (see Appendix: Evaluation Plan) and prepares questionnaires, to name just a few.

Challenges

When designing the robot, one tried to be as physically correct as possible. Instead of ball joints, mechanical cardan or three-part hinge joints have been modelled to give the CamBot the appearance of a modern industrial product. Consequently it is not possible anymore to simply pass on the orientation of a rig bone to the body parts. The rotation matrix has to be split up into single components that are then assigned to the respective elements of the mechanical joint, resulting in a correct outcome if the rotation order corresponds to the sequential arrangement of the geometry. Far too late, one became aware of the fact that Unity does not allow the user to change the rotation order, while any attempts to overwrite the standard ZXY order led to completely useless results. As time ran out, all mechanical joints were thus replaced by ball joints, what appears now acceptable for the initial tests and the on-set preview, but will certainly be retracted before importing the recorded motion data to Maya in postproduction. Furthermore, the overall robot had to be simplified to fulfil the requirements of a real-time engine. By inserting edge loops, the geometry was initially meant for being subdivided in a DCC tool, resulting in hundreds of thousands of polygons. When working in the lowpoly environment of a game engine, these additional vertices become needless and should be removed prior to the export of the model. Consequently the number of polygons has been reduced by 44 percent from unsubdivided 130.717 to 72.808.

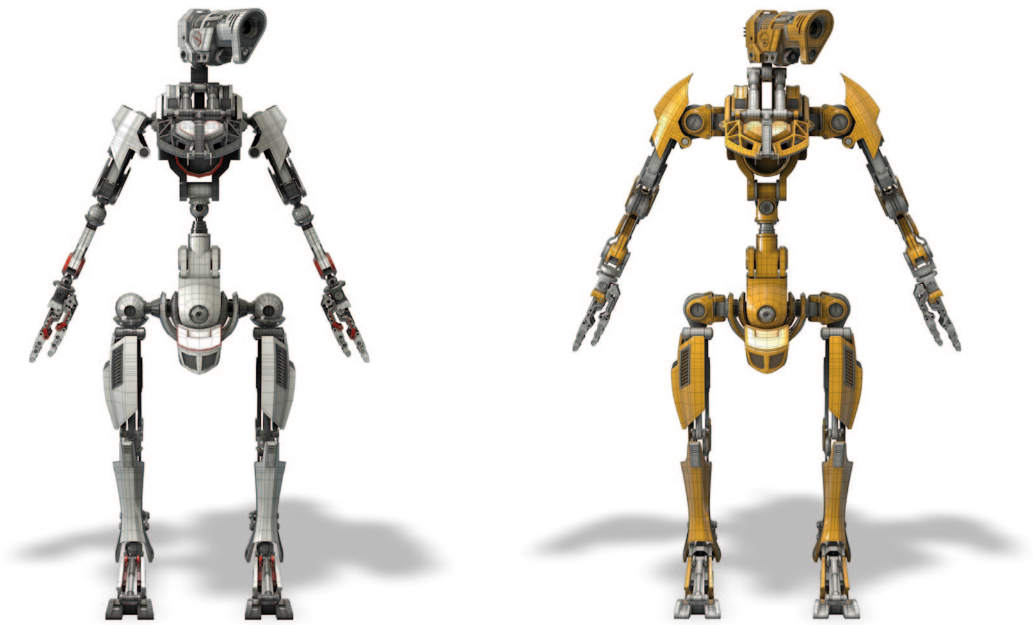


Fig.31: CamBot real-time version (left), subdivided high-poly CamBot (right)

Another challenge consisted in the necessity to guarantee a stable frame rate as well as a low latency. Due to a careful optimization of scripts and scenes the entire prototype, including all the incoming and outgoing data and video streams, now comes up to approximately 40 frames per second on a medium-performance workstation[19]. However the video footage from both the Alexa and the webcam arrives in Unity with a delay of around half a second – a value which is indeed totally insufficient for a real-time usage. What makes matters even worse is that the latency is not at all constant but differs from source to source. Since the motion data from Motive or NCam consists only of position and rotation vectors, it comes as no surprise that the related information is passed on more or less in real-time, while the videos lag behind. The Unity workstation can be considered the bottleneck of the entire pipeline as it is hardware-wise not designed for low latency video and data streaming. One tried to reduce the delay by decreasing the video resolution or disabling the VSync, but nevertheless the latency could hardly be improved.

By attaching a webcam to the Oculus Rift, a virtual camera is needed that receives not only the position and orientation of the HMD but also the intrinsic parameters of the camera. In order to be able to use this video input as a perspectively correct background on which the virtual assets can be rendered, the field of view of the camera has to match the specifications of the Oculus Rift perfectly. Thanks to the NCam system, such alignment is already implied when transferring the lens data of the principal Alexa camera, whereas the field of view of the virtual webcam counterpart has to be calculated manually.

19 Windows 7 64-bit, Intel Xeon CPU 3.1Ghz (2x16 cores), 32 GB RAM, Nvidia GTX 580

3.3.5. On-Set Production

On the first day of production, the soundstage was prepared for the shooting, by setting up the motion capture system and the video village. In the meantime, the DOP rigged the Alexa camera and the dolly. Four space lights were mounted to the studio ceiling, flooding the interior with a warmish and soft top light, which had to work as a principle illumination throughout the entire shooting as there would be neither the time nor the personnel to adjust the light on a shot-basis. After that, the NCam system was assembled and tested, including the calibration of two lenses, namely the Uniqoptics 35mm prime and the Alura zoom lens with a focal length ranging from 18 to 80mm. At the same time the software engineer worked on the RET-VP prototype and set out to correct those mistakes that had not been visible until all the components were first used in combination on set. In doing so, it appeared especially challenging to assign the tracked position and rotation data to the virtual objects, be it the robot or the virtual camera, so that the resulting motion in Unity corresponded to the movements of the real world.

Fig.32: Snapshots of the shooting

The second day was again mainly used for optimizing the entire setup and getting the different components up and running properly. The actual shooting and the experiments took place on the third and fourth day, starting with a simply shot in which the mocap performer is captured while dancing. Clothing the actor indeed took some time until the markers were place perfectly on the suit but then several takes could be recorded in rapid succession. The resulting motion data was replayed and framed by the DOP in a second pass. In the afternoon an interaction between the robot and a real person was staged by overlaying the mocap performer with the virtual character in real-time.

Fig.33: Mocap performer

The next day, the actor was asked to pick up a cardboard box, which was provided with tracking markers and directly replaced by a virtual toaster. Then, the last shot showed the robot while attempting to climb stairs to no avail with both the character and the staircase being virtual objects. In this test, all systems finally worked properly, illustrating the overall idea of a virtual production in a most impressive manner. During the shooting, there was always enough time for guiding visitors through the studio, while the users had the opportunity to test the RET-VP in practice and report their findings.

Fig.34: More snapshots of the shooting

A small making-of now portrays the work on set and allows insights into the handling of the novel devices, enabling external reviewers to understand the procedures even if they did not attend the shooting in person. Retrospectively, the production turned out to be less systematic than planned, mainly because it was still a shooting, despite the efforts to keep it as flexible as possible. Due to a small team, the thesis students had to be project managers, directors, software engineers, VFX supervisors and caterers at once. However, the experiments can be considered as successful and offer valuable clues about the virtual production environment in common and the interface prototype in particular.

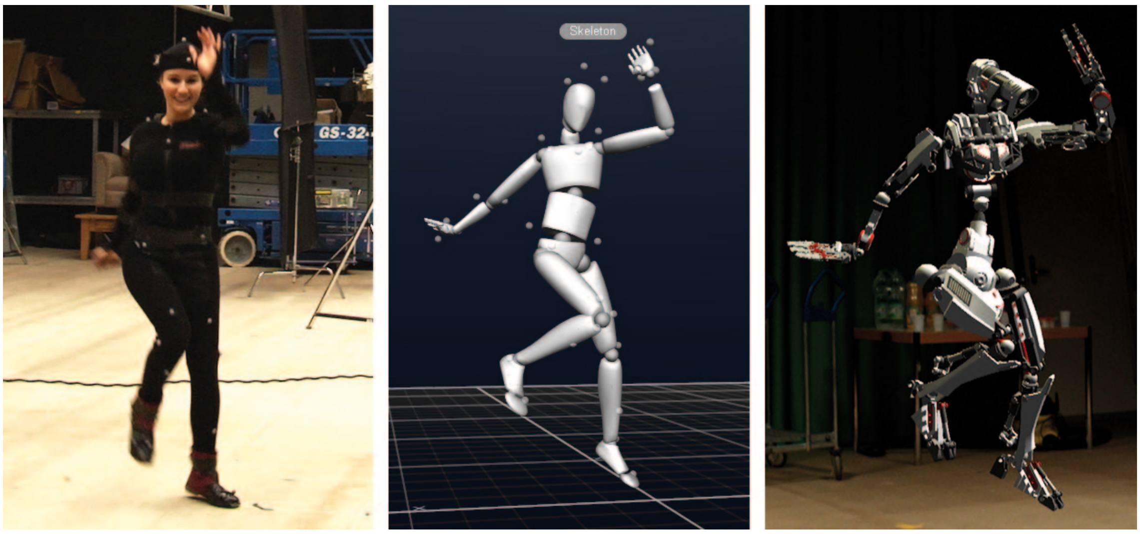


Fig.35: (left to right) Mocap performer, Motive skeleton, motion data assigned to CamBot model

3.4. Evaluation

Since the shooting has primarily been an experiment for applying and testing the interface prototype as well as for verifying or falsifying the theoretically examined benefits of a virtual production, a detailed evaluation is necessary. Directly on set, notes have been taken, summarizing the findings and conclusions derived from the observation. These results are presented in a processed and concentrated form below.

3.4.1. Personal Review

Just reading about the characteristics of a virtual production cannot supersede a real hands-on experience in which the procedures have to prove their worth in practice. While working on set, the advantages and problems of this new way of filmmaking became apparent. Furthermore the RET-VP interface could be evaluated by watching the users perform exemplary editing tasks.

Mocap Procedures

Principally, the entire motion capture workflow is well established and operates reliable without serious complications. Like this, the hardware was set up and initialized in about two hours, including the time for mounting the 24 infrared cameras onto the gantry and connecting them to the Motive workstation. Only calibrated once every morning, the system delivered high quality motion data throughout day. Furthermore, when assigning the motion data to the virtual character, the nature of the mocap performer became visible to a large extent, although the robot model is not capable of showing facial expressions.

The parallelized mocap approach, where the actor and the principal camera are tracked concurrently in separated spaces, needs some especially careful preparation, since the position of all obstacles has to be indicated inside the mocap volume by marking the shapes with tape. In this particular case, the mocap performer was asked to walk to a table and pick up a toaster, while an armchair and a stand prevented him from moving in a straight line. The props were positioned far outside the volume, not providing any clue to the actor about how to manoeuvre without having the virtual character collide with the furniture. Marking the possible path was time-consuming and cumbersome, not saving the team from readjusting the setup in case the scenery changed. Consequently parallelized motion capturing can only be recommended for empty stages without interaction or obstacles.

When capturing simultaneously the camera is tracked while looking at the mocap performer inside the volume. Once the scene is properly prepared, this procedure works reliable and allows the filmmakers to stage any kind of interaction between real and virtual characters in a most convenient manner. However, if the character is not designed to completely cover the underlying actor, complex rotoscoping will be necessary in postproduction. In the experimental scenario, the mocap performer rushed through the studio and bumped into a boom operator. Despite the latency between video and motion, the resulting collision of the virtual character with the real person appeared believable, while the preview provided the entire team on set with an immediate feedback. Even if not accessing the monitored composite, the interaction was comprehensible. Nevertheless, some disadvantages became apparent as well. By applying the simultaneous mocap technique, the DOP was forced to react on the actor in real-time, not having the possibility to restage one and the same take over and over again to try out different variations. Moreover the camera had to be constantly aimed at the mocap volume. In order to guarantee a capturing untainted with occluding objects, the entire space covered by the infrared cameras remained empty, what did in fact not constitute a problem for the Camdroid shooting, but appears not practicable for productions in which the studio environment is not by chance the intended setting. Finally one must not forget that the overall quality of motion capturing suffers from an increasing amount of actors within the volume, especially when performing interactions which require body contact and cause occlusions. As an additional experiment the described scenario was therefore also recorded applying the serial mocap procedure. For this purpose the stored motion data of the robot was replayed while the boom operator pretended to be pushed over on cue. However the resulting interaction was not at all credible.

On the first day of shooting, the mocap performer danced and jumped, portraying the CamBot at the moment of a virus attack. After having captured several takes, the team decided on a favourite version and framed the recorded motion in a second pass. As expected, this serial motion capturing allowed the decision-makers to concentrate exclusively on the performance before the DOP went on restaging the action untroubled by the hectic rush of a normal production. Furthermore, when shooting the take, it was possible to shift the moving robot to areas of the soundstage, where the scenery seemed more suitable for the particular narration. Despite all advantages some challenges became obvious as well. The coordinate systems of all incoming data sets had to be aligned in advance in order to display all virtual elements at a precise position in relation to the viewpoint of the camera. This process was time-consuming and relied heavily on visual judgement. Beyond that, it was almost impossible for the DOP to frame the action properly because the entire video signal returned to the viewfinder with delay. The video footage fell further behind the virtual elements as the underlying data streams had been derived from completely different sources without any preceding synchronization. Consequently the operator always panned too far and too late. The serial motion capturing setup appears in this conjunction especially susceptible to latencies since the DOP cannot factor the live performance of the actor into the camera work.

Conclusively, all applied motion capture techniques come along with their very own advantages and problems and defy a general recommendation. When deciding on an appropriate procedure one has to reconsider the requirements on a shot basis.

RET-VP Prototype

The main goal of the RET-VP interface, namely to enable the DOP or the director to manipulate virtual assets on set in real-time, could unfortunately not be attained. Aside from the delays, the gesture recognition was apparently not intuitive enough to provide a serious alternative to mouse and keyboard.

When editing 3D objects using the Leap Motion controller, one may expect that this way of manual modification constitutes the most natural and therefore also most intuitive input method. However gesture control is actually hard to master, especially when navigating through complex menus that have to cover the entire range of functions needed for a transformation of objects in three-dimensional space. If a creative professional is accustomed to work in a DCC tool with mouse and keyboard, the gesture-driven selection of menu options or CG assets will always lag behind a selection with conventional devices in terms of accuracy and speed. When it comes to the manipulation of objects, both approaches come up to similar results. While the sole process of set editing proves to be more intuitively to accomplish using mouse and keyboard, it is almost impossible to estimate the z-position when only accessing a flat desktop monitor. In contrast to that, the RET-VP prototype is more difficult to handle but allows a depth perception as it provides a stereoscopic vision and enables the user to move freely within the mocap volume.

The fact that mouse and keyboard were the tools of choice for performing pending set editing tasks during the experimental production does not speak in favour of the novel interface. However, this was obviously due to the problem that the system was a subject to constant modification and improvement and did not stand by for fast changes. Although the team was intent on paying closest attention to the tests, the routine of shooting slowly began to shift the weight of production to a mere completion of shots. To be able to test the interface at all, the RET-VP was applied in an alternative scenario, in which the 3D model of a space station had to be localized, selected and repositioned. The process of navigating through the virtual world can be considered as convenient, as it was possible to look around and move in a natural and unrestricted way. Selecting the target object with a casted ray also worked well, as long as the object to manipulate was not too far away from the point of view. However, while trying to transform the object, several problems occurred. First of all one needed some time to get accustomed to the related menus, since the task of controlling buttons requires accurate movements and is therefore rather difficult to perform with gestures, especially when a head-mounted display makes it impossible to check whether the hands are still properly placed above the Leap Motion sensor. Furthermore, gesture-driven control lacks any kind of haptic feedback, thus forces the user to fidget with untouchable floating menus. At least, the graphics proved to be understandable and well-arranged. As soon as the desired transformation mode had finally been picked, the modification of objects could easily be accomplished by moving the right clenched fist. In doing so, one had to pay attention not to lose sight of the selected objects.

With respect to all the occurring imperfections, one must not disregard the fact that the interface was nevertheless useful for pointing out some well working features, on which future developments may build on. At first, the RET-VP showed clearly what a fully operable mobile set editing device is most likely to offer to a virtual production environment. The prototype does not only enable the decision-makers to explore the virtual scenery by means comparable to a virtual camera system but offers tools for an immediate transformation of CG objects. Consequently the director is not forced to communicate the change requests to a 3D artist but can directly intervene, standing amid the modified scene, watching the commands take shape. For that purpose, it does not even matter if the interface consists of a multi-touch tablet or a combined solution including a head-mounted display and a gesture sensor. In addition to that, the general idea of editing objects by simply moving the hands turned out to be quite promising, assuming that the tracking quality will be improved considerably in the future. So far, the Leap Motion system comes along with just a small catalogue of supported gestures, while only the clenched fist is recognized reliably.

The RET-VP prototype points out how virtual productions may benefit from intuitive interfaces for real-time set editing. However, compared to traditional input metaphors, the current system does not really constitute an advancement in terms of intuitivity and lacks some important features. Instead of the Oculus Rift, an actual see-through HMD, like the Meta Glasses, could be applied, augmenting the natural view with virtual images rendered onto a transparent pane in front of the eyes. While the gesture control via Leap Motion appears cumbersome, hand-held 6 DOF controllers might perform better. Alternatively one should also consider a multi-touch tablet solution, since most users are already familiar to these most versatile devices. When it comes to editing, the implementation of additional tools for changing animation data or modifying virtual lights would constitute another important step towards returning the creative power to the decision-makers.

NCam System

The NCam system had not been used under the conditions of a serious production before, thus came up with a couple of challenges, which had to be met by the responsible supervisor in advance. Like this, the handling of the proprietary software appeared quite alienating at first, primarily because the graphical interface is not as well-arranged as one might expect of a highly priced product. This deficit was even made worse as the including documentation seems incomplete. However, when finally understanding the functionality, the NCam system worked very reliable without any crashes.

The complete rig, consisting of the Alexa body, an Alura zoom lens, a Preston follow focus, the NCam camera bar and two additional NCam distribution boxes weighs approximately 25 kg, making the setup completely unsuitable for a handheld application. Since the Camdroid project demanded a dynamic and breathing look for narrative reasons, the DOP used a bungee rig to mimic the appearance of a shoulder mounted camera.

As soon as both the Alexa camera and the lens had been registered and calibrated, the daily preparation of the system took only about half an hour. Having an Alura available for the Camdroid shooting, it was not necessary to use multiple lenses and the system was only calibrated once every morning. If swapping the Alura for a Prime anyway, the setup was quickly recalibrated by realigning the lenses and homing the Preston encoder, provided that the new lens had been already stored in the database. In an additional test, the soundstage was filled with haze to test whether the NCam system gets impaired by low-contrast surroundings. As expected, the tracking results remained unaffected when panning, since the gyroscopes sustain a continuous measuring of the orientation, even if no video footage can be accessed at all. When changing the position of the camera, the NCam system should work more or less reliable as long as the witness cameras are able to detect some features.

A bug in the built version of the NCam editor prevented the camera parameters from being recorded during the shooting, resulting in a bunch of empty and hence useless FBX files. This malfunction had already been identified during the first day of shooting but could not be solved in time. Thus, the NCam system carried on tracking the camera only to stream the measured values to Unity, not storing any information that might have been essential for postproduction. Drastically spoken, a minor bug rendered an entire production useless.

Virtual Production Environment

Even though the experimental shooting made use of just a few techniques a professional virtual production environment has to offer, the advantages which have been theoretically explained above proved true to a surprisingly large extent.

Like this the DOP consulted the virtually augmented images for scouting the studio prior to the actual shooting, determining the most suitable position and trying out different focal lengths. Together with the decision-makers he was able to evaluate the framing in real-time, as the composite was displayed on screens as well as in the Alexa viewfinder. Furthermore the entire team was provided with descriptive images that helped to get an idea of the director's vision and made it possible to work on the project in an attentive and efficient way. The mocap actors as well accessed the recorded output to see their performance applied to the robot character, understanding how the movements should be improved to take full effect.

These advantages became especially apparent while shooting a shot in which the robot was supposed to fail at climbing some stairs without noticing that the first step was still several metres away. From the initial point of view, the camera really saw the character standing directly in front of the stairs, and so did the entire team, at least while looking at some monitor. As soon as the perspective changed, the optical illusion got destroyed, revealing the mistake of the robot. As all essential objects were virtual, such tricks would not have been possible without a real-time preview. Even if the composite was simplified and could definitely be improved by adding shadows or textures, the quality seemed totally sufficient for providing a benefit on-set. In general one tends to buy the displayed images, as soon as the perspectives harmonize, not matter what the virtual objects look like.

In contrast to that, the preview appeared hardly usable for capturing fast action, as the delay between the video footage and the virtual elements made it impossible to keep for instance a dancing robot in sight. One attempt to approach this problem was to replay the recorded mocap data at half speed, enabling the DOP to anticipate the actor's performance. However, if later returning to the original footage, the camera motion needs to be sped up, resulting in weirdly looking movements that breathe and jitter twice as fast as they should actually do. Furthermore it was not yet possible to preview occlusions in case the robot moves behind a real prop. This shortcoming resulted in composites in which the virtual elements were always rendered atop the video footage, even though they were supposed to be almost completely hidden behind stage elements or objects. Consequently it was difficult for the DOP to frame the action in a way he was accustomed to. After all, the fact that a character is partly masked up should definitely be considered when designing an image. The entire setup of a virtual production environment is time-consuming and error-prone, as various techniques have to be prepared and coordinated, that are in fact not meant to be used in combination. Like this, the Unity game engine has definitely not been designed for being the backbone of a complete real-time filmmaking environment but was nevertheless used for receiving, processing and streaming all kinds of data, while additionally serving as a video mixer. That being said, the framework was not even capable of synchronizing the outputs, simply because it was not possible to feed in a running timecode as it was propagated by the Alexa. Once the setup worked properly, apart from the delay, the prototypical virtual production environment could be applied in a comfortable manner. However one should not expect such a workflow to be faster or less expensive compared to a traditional shooting. Even if the team often consisted of only five people that kept the production running, an overlong preparation as well as additional expenses for assembling the equipment certainly outweighs the savings from a downsized staff. Moreover, the experimental production also pointed out that the workload is far from being equally distributed. On the contrary, director and DOP both suffer from new areas of responsibility, as most of the creative work has to be done directly on set instead of deciding on issues bit by bit during months of postproduction. Finally, the question whether a virtual production makes life easier for the VFX departments cannot be answered with absolute certainty if the data from set has not yet been used in postproduction. At least some high quality outcomes can be handed over to the VFX team, including motion capture data, composites and, theoretically, camera tracks. Furthermore the artists work on footage that has been approved in advance, as it was for instance the director in person who modified the virtual elements. Of course, even then the VFX departments are not beyond revising a shot in case the decision-makers change their mind.

3.4.2. User Group Opinion

During the experimental shooting, industry experts as well as colleagues, visitors and fellow students got to know a working virtual production environment and had the chance to apply the interface prototype in practice. Afterwards, the users filled in a short questionnaire, evaluating the RET-VP and sharing further ideas and suggestions.

In the end, 20 people participated in the survey. The user group consisted of media students and professionals with expertise in the fields of feature film, computer animation, visual effects, postproduction, research and development, interactive media, photography, media economy and education.

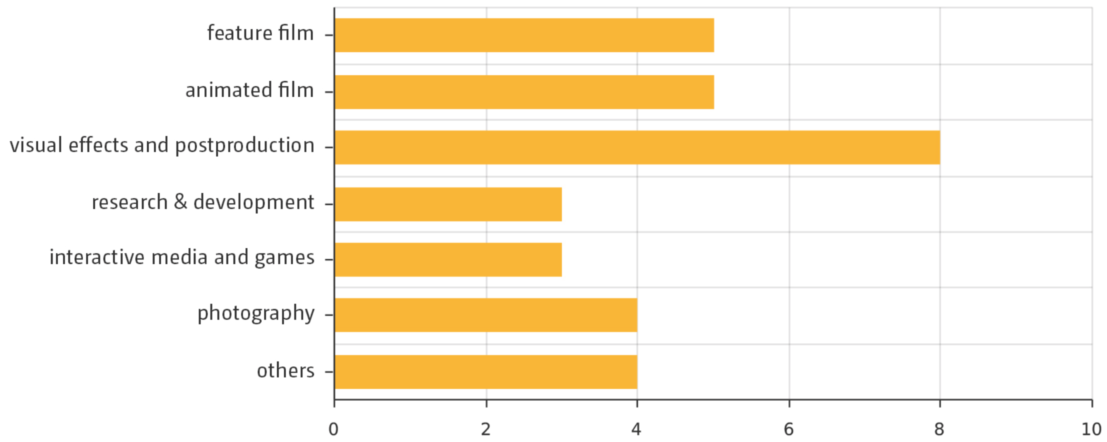


Fig.36: Fields of experience

In the questionnaire two-thirds of the surveyors describe themselves as young professionals while seven more participants look back on five or even ten years of experience. In general, most of the users are quite impressed by the opportunities a virtual production can offer, although only 60 percent of them contributed to a VFX production before.

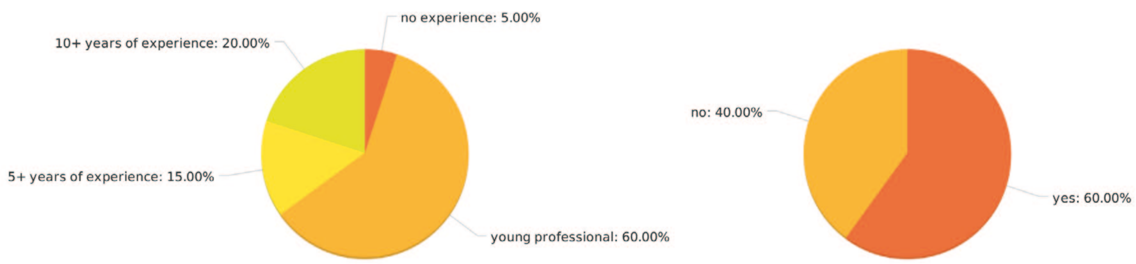


Fig.37: Years of experience (left), experience in the fields of visual effects (right)

Virtual Production Environment

The first part of the questionnaire is more generally concerned with the virtual production environments. When asked which part of the visual effects workflow is in most need of optimization, several people agree with the main concern of virtual production, namely the real-time preview and modification of virtual elements on a live action shooting. Furthermore a surveyor states that the industry requires a consistent flow of metadata from preproduction to shooting, postproduction and distribution in order to reduce the overhead work, which comes along with current trial and error approaches. Only if the industry succeeded in coordinating the different hardware and software components, such a smooth data transfer from on-set tools to VFX postproduction systems would be possible. Further statements expose the need for simplified keying and camera tracking procedures as well as for standardized ways to capture the data from set. Additionally common standards for sharing ideas and assets are considered necessary.

95 percent of the participants already know the term virtual production and are even able to describe their idea of this novel workflow. Like this, Thomas Knop, CEO of Stargate Studios, delivers a model definition by stating that a "virtual production is most commonly understood […] as a production method using virtual assets like computer generated objects […] together with practical elements like actors, props and environments to enhance reality being told […] in feature films and TV productions." Additionally, he explains that a virtual production also comprises a far more integrated pipeline across multiple disciplines, using the created data and material for a more efficient workflow. One definition relates to the process of bringing postproduction elements back to the shooting by introducing real-time technology, while another surveyor points out that a virtual production just requires virtual sets or characters filmed by a virtual camera and does not necessarily involve a real camera on location. These statements constitute only a rather small fraction of the more or less identical but nevertheless amazingly accurate answers given by the user group. To tell the truth, one did definitely not expect the majority of participants to be this familiar with the characteristics of a virtual production.

However, the answers above come as no real surprise, when realizing that one third of the surveyors have already applied virtual production techniques. The most important advantage of the overall workflow is seen in the fact that creative personnel and decision-makers, like DOPs, authors, directors or supervisors are enabled to spend more time on quality instead of being busy with delegating jobs. It addition to that, there is a broad consensus that a virtual production enables everybody on set to get a better feeling and understanding of the final outcome, avoiding guesswork when it comes to timing, especially in shots where virtual characters are involved. Live previews are regarded as a feature that can help the VFX artists to do a lot of their work directly on set, where they are perfectly able to react on the studio setup. Furthermore several participants are convinced that the workflow can help to reduce costs by optimizing the workflows on set and in postproduction as the amount of necessary iteration steps may decrease. One user calls attention to the benefit of an immediate feedback for DOP and performers and points out how the green screen acting might be advanced, if it was possible for the actors to see the virtual elements as well. When directly asked to pick from a list of advantages, the user group agreed that a virtual production constitutes a creative and integrative environment that comes up with a more efficient workflow.

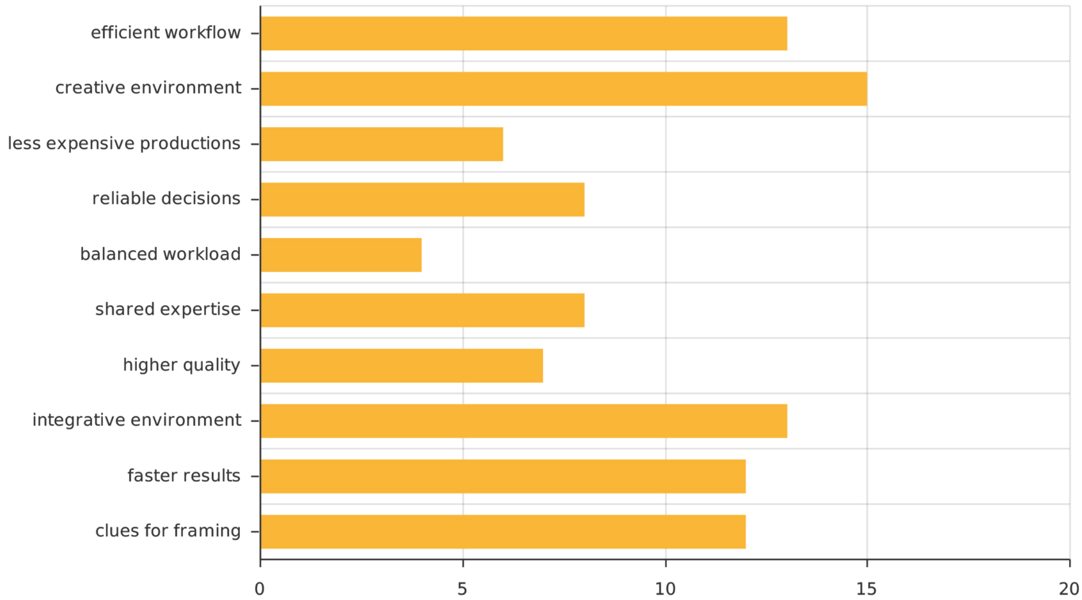


Fig.38: Assumed advantages of virtual production environments

Around 90 percent of the surveyors think that the advantages coming along with a virtual production environment are worth the effort, while 95 percent would even apply the related procedures in their own projects.

RET-VP Prototype

The second part of the questionnaire focuses on the evaluation of the set editing prototype. Again, the user group has already some experience with the applied techniques as three-quarters of the participants used a head-mounted display before, while 65 percent are familiar with gesture control.

Since a modification of virtual objects covers various subtasks, every single interaction technique has to be evaluated one by one to deliver significant results. Thus, each category is rated with grades, reaching from one to six, while the grade one represents an excellent interaction method whereas a six stands for an awfully functioning technique. Apparently, the surveyors are able to cope well with the Oculus Rift as an instrument for exploring the virtual world, as the orientation in three-dimensional space achieves a result of 1.80. As soon as some kind of gesture control is required, the grades get worse – a result which totally complies with the personal evaluation presented in the chapter above. Like this, the users submit a 2.50 for the navigation tasks performed with the virtual hands, while the selection of an object is rated with an average value of 3.30. The manipulation lags behind with a grade of 3.90. However, one has to point out that the standard deviation for all gesture-driven interaction techniques is between 0.89 and 1.17, revealing that the participants do not at all reach a common conclusion.

	excellent (1)		good (2)		not so good (3)		not so bad (4)		bad (5)		awful (6)			
	Σ	%	Σ	%	Σ	%	Σ	%	Σ	%	Σ	%	Ø	±
orientation in space	5x	25,00	14x	70,00	1x	5,00	-	-	-	-	-	-	1,80	0,52
navigation	1x	5,00	11x	55,00	6x	30,00	1x	5,00	1x	5,00	-	-	2,50	0,89
selection of objects	-	-	5x	25,00	9x	45,00	2x	10,00	3x	15,00	1x	5,00	3,30	1,17
manipulation of objects	-	-	1x	5,00	8x	40,00	4x	20,00	6x	30,00	1x	5,00	3,90	1,07

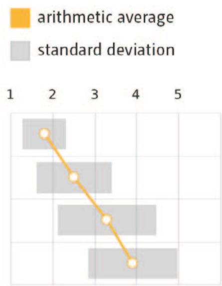


Fig.39: Results for interaction techniques

When furthermore evaluating the gesture control and the graphical user interface in terms of intuitivity, the grades differ considerably as well. While the overall principle of selecting and manipulating objects by the use of hands achieves a grade of 3.30, the menu is rated with a 1.85. In comparison to previous results, the GUI performs well, indicating that the partition into submenus and the arrangement of buttons appear somewhat comprehensible, even if the menu is difficult to handle. Generally speaking, the results of the questionnaire correspond to the outcomes of the personal review as they confirm the functionality of the interface prototype in principal but also reveal the shortcomings of certain components. The head-mounted display and the GUI seem viable, whereas the gesture control is still in sore need of improvement and might be replaced by an alternative input method.

	excellent (1)		good (2)		not so good (3)		not so bad (4)		bad (5)		awful (6)			
	Σ	%	Σ	%	Σ	%	Σ	%	Σ	%	Σ	%	Ø	±
gesture control	-	-	7x	35,00	9x	45,00	3x	15,00	1x	5,00	-	-	2,90	0,85
graphical user interface	6x	30,00	11x	55,00	3x	15,00	-	-	-	-	-	-	1,85	0,67

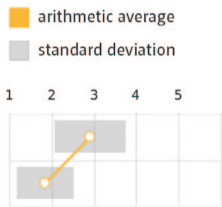


Fig.40: Results for intuitivity

In the last part of the survey the users are asked to contribute some thoughts about how the interface prototype could be improved. Some of the participants request a more reliable gesture recognition system or even call for a totally different device, like a hand-held controller, that may provide a more practical interaction for the fairly complex manipulation tasks. Apparently, the

Leap Motion sensor is not accurate enough and loses sight of the hands all too often. Related to that, some users complain about the menus being quit immediately once the left hand disappears from the volume by mistake and suggest an alternative approach in which both hands have to be taken out of the frame concurrently to leave the menu. In general, a broader variety of distinct gestures seems desirable in order to avoid accidental maloperations. One surveyor just pointed out that gesture control will never be as intuitive as an interaction with mouse and keyboard. Others recommend different mapping functions or an alternative selection method in case objects are far away. Several users also miss some kind of learning session in which one gets to know the single features of the interface before applying the entire range of functions in combination. Ideally, such a tutorial is not even performed with the Oculus Rift but with a standard desktop screen, at least for taking the first steps.

According to the surveyors, a market-ready solution should furthermore enable the user to record the changes and attach annotations. Besides that, it would be helpful to have an option for snapping the virtual objects to a ground plane. One participant proposed a feature that allows multiple operators to work simultaneously on a single scene, sharing ideas and suggestions in real-time. Additionally, a voice control or a multi-touch interface fixed to the HMD could introduce a much more intuitive human-computer interaction. Advanced techniques for light editing or on-set compositing would round off an ideal toolset for a modern virtual production environment. However, before dealing with nice-to-have gimmicks one should address the main problems by reducing the latency and increasing the resolution, finally accomplishing a live video stream which uses the full field of view of the head-mounted display.

Chapter 4

Conclusion

Initially, this paper was to examine the formation conditions, which led to the development of virtual production approaches. Due to the current VFX crisis, the industry is in sore need of advanced and more efficient procedures, while the pipelines coming along with virtual productions seem to cope perfectly with the challenges of modern digital filmmaking. By introducing common standards, all production steps from preproduction to postproduction can profit from a smooth data flow, resulting in a collaborative environment, to which all departments can contribute in a most interactive and intuitive manner. Game-like real-time technology furthermore allows the decision-makers to explore a virtually augmented reality live on set, enabling them to come up to informed decisions and artistically appealing results. However, despite the apparent qualities, one should not be mistaken about the actual stage of development, as most of the used technologies are either prototypical or lack essential features, while the appropriate workflows have still to be formed. To verify the theoretically explained advantages, challenges and limitations in practice, a fully functional prototype of a virtual production environment was developed and tested in an experimental shooting. The resulting setup did not only build on novel tracking technologies for previewing computer-generated characters in combination with live action footage but also introduced an innovative interface for editing the virtual set in real-time. The shooting provided revealing insights into the work on a virtual production set and helped to get a better understanding of the related workflows. In fact, the advantages of virtual filmmaking became apparent quickly. Despite the rudimentary state of the prototypical setup, the creative professionals really used the given techniques for virtual scouting, framing, actor guidance and decision-making. Once set up, the system worked reliably and was of great benefit for the realization of the planned shots. Furthermore the virtual production workflow proved to be scalable. Although the studio setup, including the Alexa camera, a motion capture system, innovative camera tracking techniques and novel input and output devices, can definitely not be described as a low-budget solution, the shooting was only little more expensive than a comparable traditional production, thus exposed that even student projects or independent films can make use of this novel way of filmmaking.

However, the use case also demonstrated that virtual production environments heavily depend on a carefully devised implementation and combination of different software and hardware components. If the on-set team did not succeed in providing a continuous and smooth flow of data, the setup would get widely unusable, at the same time undermining the general concept of real-time filmmaking.

Regarding the interface prototype, the general idea of an interactive environment in which the user is empowered to perform changes that are propagated through the entire on-set pipeline immediately, turned out to be practicable. The survey pointed out that a head-mounted display for navigation and a 2D menu for system control are both widely suitable, while the gesture control still suffered from inaccuracy when applied for manipulating virtual objects. By further refining the pipeline and integrating novel technology, the Dreamspace project will continue to develop techniques and workflows which might one day lead to an overall integrated setup for collaborative real-time editing in virtual productions.

To this day, the VFX industry is in an early stage of adoption of virtual production methodologies. Although it appears most evident that all departments on set and in postproduction profit enormously from the collaborative and interactive environment introduced by the novel workflow, there are still way too many ingrained and somewhat outdated practices, which have not yet been overcome, both mentally and technically. Having said that, one must not disregard the fact that the majority of virtual production techniques are far from being standard or routine, as customized technologies are often developed and applied only for particular projects instead of advancing the workflow in favour of the entire industry. The existing hardware is expensive and requires specially educated personnel, thus appears unsuitable for small or independent productions or projects with tight schedules and low budgets in general (Knop 2014, p.74). Considering the big studios, it is a widespread myth that novel filmmaking technologies can help to reduce costs. By applying virtual production methods, films with large-scale visual effects will definitely not become less expensive, whereas the quality as well as the amount of VFX shots will continue to rise considerably (Legato 2012). Furthermore, the existing prototypical solutions always carry the risk of failure and postponement – a fact that might alienate major studios, which have to stick to fixed deadlines at all costs.

Despite all current pitfalls and deficiencies, a virtual production environment constitutes a powerful solution for coping with the requirements of a virtual filmmaking process in the digital age. The novel approach does not only lead to a much higher visual quality but helps to meet the demands of VFX departments as well. Indeed many facilities are quite unhappy with the contemporary working conditions and ask for more reliable task descriptions so that the artists know what is actually expected of them before they set out to do the work (Patel 2009, p.18). By applying virtual production techniques, each creative professional involved in the project is able to contribute to the vision, helping the decision-makers to come up to distinct and consistent conclusions, which again allow the contributors to work on a common knowledge-base. Hence, one does not simply put the blame on the missing communication between the departments but works towards a higher level of collaboration (Sylwan 2012). Technology-wise virtual production environments will have to hide their complexity, getting small, mobile and easy to use (Trumbull 2012). In the near future, as techniques progress, the qualities are certainly going to outweigh the disadvantages and will make the novel workflows an irreplaceable part of production. Thus, the prototypical interface developed in the context of this thesis made a small but notable contribution to the continuing evolution of virtual filmmaking.

Outlook

In the future, the idea of a virtual production will have gained acceptance among the filmmakers as the related workflows will deliver highest quality at reasonable cost. In addition to that, the required technology will also be widely accessible to productions outside the U.S., including independent projects, which usually focus more on the actors and less on breath-taking visual effects (Kilkenny 2012). The great success of films like 'Avatar' or 'Tintin' will further bring studios to make use of the novel procedures, whereas James Cameron stated that "creating the virtual production pipeline on 'Avatar' was a groundbreaking process that only enabled us to scratch the surface of what is possible" (quoted by Giardina 2012).

It is an exciting time to be a filmmaker. Augmented reality techniques find their way into production pipelines, while light stages, LIDAR scanners, time of flight and lightfield cameras change the way visual effects used to be made, enabling the creative professionals "to devote more of […] [their] energy to the creative side of the moviemaking process, and dig deeper into all that is possible with virtual production" (Cameron quoted by Giardina 2012). However, innovative technology will not only affect the methods of film creation but also introduce new channels of consumption. Mobile devices come up with an ever growing performance and evoke the formation of new interactive narratives, which can be accessed on demand all over the world at any time. Carl Rosendahl thus ventures a rather fatalistic prognosis of the future by saying that "entertainment will become integrated even more tightly into our lives. But not as movies" (Rosendahl 2012).

Chapter 5

Appendix

5.1. Schedules and Tables

Equipment List

Kamera	
Arri Alexa Plus/ Arri Amira	1
Alura (18-80)	1
Prime (24, 50, 85)	3
Astro Kontrollmonitor	1
Kompendium	1
Preston Funkschärfe	1
ND-Filter	3
SDI Kabel (0.5m)	3
SDI Kabel (10 m)	2
Netzkabel Arri	1
V-Mount Akkus	3
SxS Karten (64GB)	3

Kamerabühne / Kamera-Grip	
Kamerastativ (150) + Kopf	1
Bungeerig	1
Handgriffe	1
Bühnenkiste	2
Panther Dolly + Studioräder	1
Easyrig 4 (max. 40 kg)	1

Video Village	
Referenzmonitor (17")	2
Workstation	1

Virtual Production Setup	
OptiTrack Motion Capture System	1
Mocap Workstation	1
VCS Rig	1
Oculus Rift (+Webcam)	1
Leap Motion	1
Leap Motion Halterung	1
nCam Bar (+Zubehör)	1
nCam Server	1

Licht-Grip	
Stahlstativ	4
C-Stand groß	6
C-Stand Boden	2
Alustativ groß	3
Alustativ heavyduty	3
Kurbelstativ	2
Goboarm	2
Gobokopf	6
Magicarm	4
Uniklemme	30
Cardellini	4
Styrogabel oder -halter	2
Butterflyhalterung	2
Traverse (hängend)	1

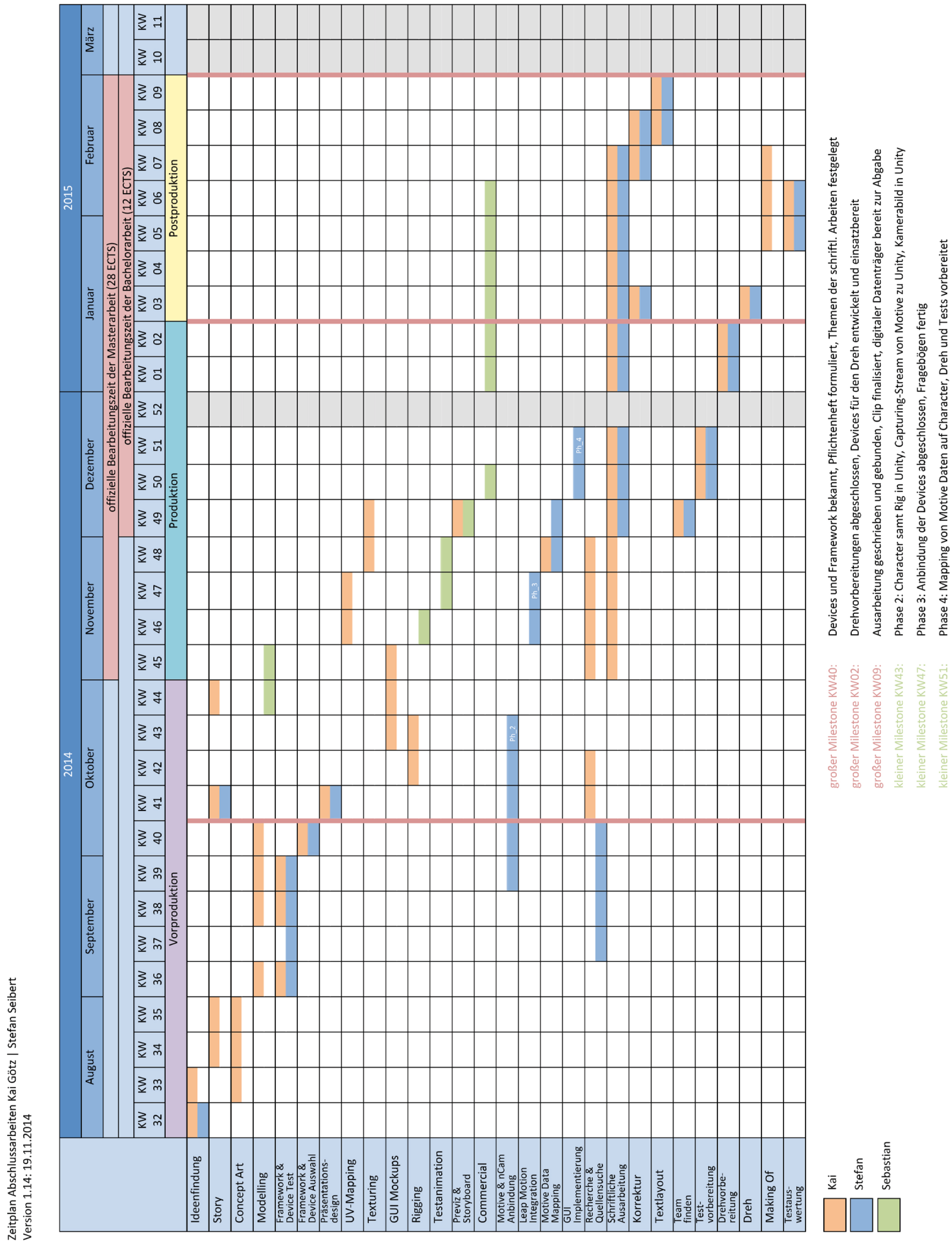
Making Of	
Kamera	1
Aufsatzmikro	1
Laptop (+ext. Festplatte)	1

Sonstiges	
Laptops für Fragebögen	2
Tonangel	1
Mikro	1
Recorder	1
XLR Kabel (2m)	1

Strom	
Schuko Verlängerung (10m)	6
Schuko Verlängerung (20m)	4
Mehrfachstecker	6
Kabeltrommel	2
Verteilerbox (32A)	1
Drehstromkabel	1

Licht	
Lichtpaket 2	1
Arrilight 800	1
Compact 575	2
Arrilight 300	2
Spacelights	4
Vistabeam	2
Kinoflo Fourbank	2
Chimera M	2
Chimera L	2
Styro	2
Pebble	1
6x6 Butterfly	1
Folien (WD, ND, CTB, CTO)	8
Frostrahmen	2

Thesis and Production Schedule



Evaluation Plan

#	experiment	question	evaluation
01	parallelized mocap: concurrent capturing of performer and camera in separated spaces mocap performer is captured inside the volume virtual character is positioned outside the volume camera is tracked outside the volume, looking at the virtual character	Which advantages and disadvantages come along with the concurrent capturing of camera and performer in separate spaces?	advantages and disadvantages set up workflow performer's creative freedom DOP's creative freedom
		Is this workflow suitable at all?	qualitative evaluation quality of final outcome
	set editing: manual transformation of virtual character virtual character has to be located at the correct position using both traditional and novel interface	Can a creative professional who is not well educated in DCC tools transform an object in 3D space to the correct position with mouse and keyboard?	qualitative evaluation quality of navigation quality of selection quality of manipulation
		Can a creative professional who is not well educated in DCC tools transform an object in 3D space to the correct position with our novel interface?	qualitative evaluation intuitivity of GUI quality of gesture control quality of navigation quality of selection quality of manipulation
02	serial mocap: capturing of performer and camera spatially and temporally separated mocap performer is captured inside the volume virtual character is positioned somewhere in the studio later the camera is tracked outside the volume, looking at the virtual character	Which advantages and disadvantages come along with the spatially and temporally separated capturing of camera and performer?	advantages and disadvantages set up workflow performer's creative freedom DOP's creative freedom
		How reliable is the serial mocap setup?	qualitative evaluation quality of final outcome
	virtual scouting DOP and director walk around and explore the set with a HMD DOP and director explore the set with the nCam-tracked camera and preview a combination of real and virtual footage	Does it make sense that the director and the DOP can explore the virtually augmented studio with a head-mounted display?	qualitative evaluation quality of navigation efficiency of decision making
		Do the real and virtual worlds match?	qualitative evaluation quality of visuals
		Does creativity benefit from this approach? Which benefits?	advantages and disadvantages brainstorming framing blocking staging
		Does the director access the real-time preview?	qualitative evaluation frequency of usage
03	capturing of rigid body a rigid body is captured together with the performer the measured data is transferred to the virtual counterparts	Does the simultaneous capturing of rigid bodies and actors work?	qualitative evaluation quality of combination
		How convenient is the assignment of captured transformation data to virtual objects?	qualitative evaluation complexity of workflow
04	simultaneous mocap: concurrent capturing of performer and camera in same space mocap performer is captured inside the volume camera is captured with the nCam system and looks towards the real actor the virtual character is rendered on top of the actor and occludes the actor	Which advantages and disadvantages come along with the simultaneous capturing of performer and camera?	advantages and disadvantages set up workflow performer's creative freedom DOP's creative freedom
		How reliable is the simultaneous mocap setup?	qualitative evaluation complexity of setup quality of final outcome
	interaction with real person the actor collides with another person the virtual character is rendered on top of the actor and seems to bump into the second person	Is this preview believable, thus usable at all?	qualitative evaluation quality of interaction quality of combination
		Does the captured motion data look the same when assigned to the virtual character?	qualitative evaluation quality of transferred motion
05	set editing: manual transformation of virtual objects virtual object has to be located at the correct position using both traditional and novel interface	Can a creative professional who is not well educated in DCC tools transform an object in 3D space to the correct position with mouse and keyboard?	qualitative evaluation quality of navigation quality of selection quality of manipulation
		Can a creative professional who is not well educated in DCC tools transform an object in 3D space to the correct position with our novel interface?	qualitative evaluation intuitivity of GUI quality of gesture control quality of navigation quality of selection quality of manipulation
		Is it possible to position two virtual objects in a way that they meet the requirements of the framing?	qualitative evaluation quality of matching
all	virtual production 2.0 innovative interfaces on set applied by traditional educated personnel	Can creative professionals with a traditional knowledge background perform tasks in 3D space?	qualitative evaluation effort to fulfil the objective quality of final outcome
		Do the theoretically derived advantages of virtual production prove to be true in practice?	qualitative evaluation more efficient workflow collaborative work creative environment shared expertise less expensive productions higher overall quality reliable decisions integrative environment balanced workload faster results

5.2. Glossary

3D ARTIST. Digital artist engaged in work related to 3D computer graphics. Working on models, rigs or animations in DCC tools.[1]

ACCELEROMETER. Device for measuring the proper acceleration. Can be used together with gyrometers and magnetometers to determine the orientation of a device.[1]

ANIMATIC. Filmed storyboard. Storyboard frames combined to a video, sometimes including rudimentary dialogues and music, offering valuable clues about camera movements, timing and pace.[1]

ANIMATION. Moving Imagery that is created on a frame-by-frame basis. Accomplished via the use of computers or more traditional cel animation techniques.[2]

ANIMATRONIC. Robotic device to emulate mimics, gestures or locomotion for an otherwise inanimate object.[1]

ASSET. Any 3D object created with 3D modelling and animation software.[5] Term usually used for digital props.

BLOCKING. Animation technique in which key poses are created to establish timing and placement of characters and props. Blocking is most commonly used in 3D computer animation.[1]

CHROMA-KEYING. A keying technique to separate an object from its background based on colours that are unique to either the foreground or the background.[2]

CINEMATOGRAPHY. Science or art of motion picture photography.[1]

COMPOSITING. Process of combining visual elements from separate sources into single images, often to create the illusion that all those elements are parts of the same scene. Today achieved through digital image manipulation.[1]

CENTRAL PROCESSING UNIT. CPU. Electronic circuitry within a computer that carries out the instructions of a computer program by performing the basic arithmetic, logical, control and input/output operations.[1]

CUE. Trigger for an action to be carried out at a specific time.[1]

DIGITAL CONTENT CREATION TOOL. DCC tool. Software used for creation of electronic media.[1]

DEGREES OF FREEDOM. Number of independent motions that are allowed to a body.[1]

FRAMING. Presentation of visual elements in an image, especially the placement of the subject in relation to other objects. Framing can make an image more aesthetically pleasing and keep the viewer's focus on the framed objects.[1]

GRAPHICS PROCESSING UNIT. GPU. Specialized processor for rendering 3D graphics.[5]

GRAPHICAL USER INTERFACE. GUI. Type of interface that allows users to interact with electronic devices through graphical icons and visual indicators.[1]

GREEN SCREEN. Green monochrome backdrop. Used for replacing the backdrop with material from a different image applying chroma-keying techniques. Green is the ideal colour, as image sensors in digital video cameras are most sensitive to green, due to the Bayer pattern allocating more pixels to this channel.[1]

GYROSCOPE. Device for measuring orientation.[1] Can be used together with accelerometers and magnetometers to determine the orientation of a device.

HEAD-MOUNTED DISPLAY. HMD. display device, worn on the head or as part of a helmet. Has a small display optic in front of one (monocular HMD) or each eye (binocular HMD).[1] Modern HMDs provide build-in position and orientation sensors.

JAMES CAMERON. Canadian film director, film producer, screenwriter, editor, inventor, and engineer who has directed the two biggest box office films of all time, 'Titanic' and 'Avatar'.[1]

JITTER. Erroneous tracking deviation. Inaccurate measurement caused by reflections, oversized volumes or wrong calibration.

KABUKI. Video of a human face projected as basic texture onto lowpoly head geometry. Used as previsualization of facial captures.

KERNEL. Piece of code that will be run at each position in the iteration space.[3]

KEYING. Process of algorithmically extracting an object from its background.[2]

LEAP MOTION CONTROLLER. Gesture Recognition Sensor developed and distributed by the American company Leap Motion. Small USB peripheral device designed to track the human hand.[1]

LIVE-ACTION. Refers to cinematography or videography that is not animated.[1]

LOWPOLY. Adjective describing models with a small polygon count.

MACHINIMA. Animation technique that uses interactive real-time computer-generated imagery and games as underlying render engines or interactive virtual production studios.

MAPPING. Process of transferring values from one system to another. Often used for describing the assignment of motion capture data to the virtual character.

MATCH MOVE. Techniques for extracting camera motion information from a motion picture. Needed for a correct insertion of computer graphics into live action footage with correct position, scale, orientation, and motion relative to the photographed objects in the shot.[4]

MATTE. An image used to define or control the transparency of another image.[2]

MOTION CAPTURE SYSTEM. Hardware setup for capturing an actor's performance. May apply different techniques such as optical, inertial, electromagnetic, acoustic or mechanic tracking.

MOTION CAPTURE. Technique whereby an actor's performance is captured and translated for driving a CG character's performance.[2]

MOTION CONTROL CAMERA. Method of using computer-controlled mechanism to control the position, orientation, and lens settings of a camera so that its movement is continuously repeatable.[2]

MOTION GRAPHICS. Animated graphic imagery that is done primarily to achieve a specific visual design rather than to produce realistic images.[2]

NCAM. Markerless real-time camera tracking system providing complete position and rotation information, focal length and focus.

OCULUS RIFT. Virtual reality head-mounted display. Available as development kits DK1 and DK2.[1]

PERFORMANCE CAPTURING. technique whereby an actor's performance is captured and translated for driving a CG character's performance.[2] Often used for full body captures including facial capturing.

PETER JACKSON. New Zealand film director, producer and screenwriter. Best known as the director and producer of 'The Lord of the Rings' trilogy and 'The Hobbit' trilogy.[1]

PIPELINE. Well designed set of processes for achieving a certain result.[2] Sequence of tasks throughout the film production.

PLATE. A piece of original photography that is intended to be used as an element in compositing.[2]

POST-PRODUCTION. Work done once principal photography has been completed.[1] General process of working on the on-set recorded material to achieve the final deliverable output. Also commonly referred to as the offline process.[4]

POSTVIS. Rough combination of digital elements and production photography to validate footage selection and refine effects design. Provides placeholder shots for editiorial.[2]

PREVIS. Collaborative process that generates preliminary versions of shots or sequences, predominantly using 3D animation tools and a virtual environment. Enables filmmakers to visually explore creative ideas, plan technical solutions, and communicate a shared vision for efficient production.[2]

PROXY. Scaled-down image used as a stand-in for higher resolution original.[2]

RAYTRACING. Technique for generating an image by tracing the path of light through pixels in an image plane and simulating the effects of its encounters with virtual objects.[1]

REAL-TIME. Computational processing that appears to be nearly instantaneous.[2]

RENDERING. Process of creating synthetic images from 3D data set.[2]

ROTOSCOPING. Process of creating imagery or mattes around objects on a frame-by-frame basis.[2]

SHOT. Unbroken continuous image sequence.[2] Smallest part of a filmic scene.

SPECIAL EFFECTS. SFX. Practical optical effects created in front of the camera.[2]

TECHVIS. Technical diagrams or breakdowns for analysis and communication of technical specifications. Done prior to the shooting.

TENT-POLE FILM. Film that supports the financial performance of a studio or television network. Accompanied by large budgets and heavy promotion. Expected to turn a profit in a short period of time.[1]

TRACKING. Process of determining the movement of objects or cameras in a scene by analysing the captured footage.[2]

VIRTUAL CAMERA SYSTEM. Setup for handling a virtual camera based on familiar camera form factors including controls for focal length, zoom, and depth of field as well as common film-specific setups such as virtual dolly and crane rigs.[5]

VIRTUAL FILMMAKING. Virtual Production. Visually dynamic, non-linear workflow, blending virtual camera systems, advanced motion and performance capture, 3D software and practical 3D assets with real-time render display technology. Enables filmmakers to interactively visualize and explore digital scenes for the production of feature films and game cinematics.[5]

VIRTUAL PRODUCTION COMMITTEE. Joint initiative assembled by the American Society of Cinematographers (ASC), the Art Directors Guild (ADG), the Visual Effects Society (VES), the Previsualization Society, and the Producers Guild of America.[2]

VISUAL EFFECTS. VFX. Broad term referring to anything that cannot be captured using standard photography techniques. Accomplished via a number of different digital postproduction processes.[2]

VOLUME. Motion capture stage.[2] volume of space defined by a series of remote sensing devices capable of detecting the exact orientation and movement of the tracking controller.[5]

WETA DIGITAL. Digital visual effects company based in Wellington, New Zealand. Founded by Peter Jackson, Richard Taylor, and Jamie Selkirk in 1993. Known for 'The Lord of the Rings', 'Avatar', 'The Hobbit', 'Planet of the Apes' and many more.[1]

WIDGET. 2D or 3D graphical control element. Element of interaction in a graphical user interface (GUI).[1]

WITNESS CAMERA. Camera used to record the action from a viewpoint differing from the viewpoint of the primary production camera.[2]

[1] Wikipedia. (2015). http://www. wikipedia.org.
[2] A. Okun, S. Zwerman (Eds.). (2014). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures. Focal Press.
[3] Introduction to Blink Kernels. (n.d.). http://www.thefoundry.co.uk.
[4] Knop, T. (2014). Virtual Production - Methods Guidelines Scenarios. http://www.dreamspaceproject.eu/documents.
[5] Patel, M. (2009). Autodesk Whitepaper - The New Art of Virtual Moviemaking. Autodesk. Retrieved from: http://www.autodesk.com.

5.3. Bibliography

Adib, F., Kabelac, Z., Katabi, D., Miller, R. (2014). 3D Tracking via Body Radio Reflections. Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation, 317-329. Retrieved from: http://dl.acm.org.

Altman, R. (n.D.). Zoic's Mike Romey discusses Previs App ZEUS:Scout. Postperspective. Retrieved from: http://postperspective.com.

Bachmann, D., Weichert, F., Rinkenauer, G. (2014). Evaluation of the Leap Motion Controller as a New Contact-Free Pointing Device. Sensors 2015, 214 - 233.
Retrieved from: http://www.mdpi.com.

Balakrishnan, G., Diefenbach, P. (2013). Virtual Cinematography: Beyond Big Studio Production. Proceeding ACM SIGGRAPH 2013 Studio Talks. Retrieved from: http://dl.acm.org.

Beck, M. (2014 a). Previs: Advanced Techniques. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 68 - 74. Focal Press.

Beck, M. (2014 b). What is Previs. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 45 - 48. Focal Press.

Bell, K. (2014). Leap Motion Can Track Your Fingers, Even When It Can't See Them. Mashable. Retrieved from: http://mashable.com.

Bennett, J., Carter, C. (2014). Adopting Virtual Production for Animated Filmmaking. Paper presented at the 7th Annual International Conference on Computer Games, Multimedia and Allied Technology CGAT 2014. Retrieved from: http://eprints.qut.edu.au.

Billington, A. (2008). A Brief Look at the Technological Advancements in James Cameron's Avatar. Firstshowing. Retrieved from: http://www.firstshowing.net.

Bowman, D., Hodges, L. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. Proceedings of the 1997 symposium on Interactive 3D graphics, 35 - 38. Retrieved from: http://dl.acm.org.

Bowman, D. et al. (2005). 3D User Interfaces – Theory and Practice. Addison-Wesley.

Clavadetscher, C. (2014). Real-time Visual Effects and Camera Tracking. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 193 - 200. Focal Press.

Conner, B. et al. (1992). Three-Dimensional Widgets. Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics, 183 - 188. Retrieved from: http://dl.acm.org.

Cox, M. (2014). Source Filmmaker + Razer Hydra [video file].
Retrieved from: https://www.youtube.com.

Dartfish - Discovering DartCaster, SimulCam & StroMotion in Action. (n.D.).
Retrieved from: http://www1.dartfish.com.

Derry, G. (2011). Interview with Glenn Derry, Virtual Production Supervisor, Videohawks & Technoprops (Part 1 of 3) [video file]. Retrieved from: https://www.youtube.com.

Derry, G. (2012). FMX 2012: System Engineering for Virtual Production [video file].
Retrieved from: http://area.autodesk.com/fmx2012.

Desai, P. R. et al. (2014). A Review Paper on Oculus Rift – A Virtual Reality Headset.
International Journal of Engineering Trends and Technology (IJETT),
Volume 13 Number 4, 175 - 179. Retrieved from: http://arxiv.org.

Dreamspace. (2015). Retrieved from: http://www.dreamspaceproject.eu.

Dreamspace DOW - Description of Work. (2014). [confidential file].

Duncan, J., Fitzpatrick, L. (2010). Avatar: Der Film - Das Making Of. Knesebeck.

Dunlop, R. (2014). Production Pipeline Fundamentals for Film and Games. Focal Press.

ElKoura, G., Singh, K. (2003). Handrix: Animating the Human Hand. Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation, 110 - 119.
Retrieved from: http://dl.acm.org.

Faber, R. (2009). Script as only starting point in the narrative media is outdated. Faberface.
Retrieved from: http://www.faberface.com.

Fink, M. (2014). Introduction. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 1 - 4. Focal Press.

Giardina, C. (2011). The Adventures of Tintin: What James Cameron Showed Spielberg and Jackson. The Hollywood Reporter. Retrieved from: http://www.hollywoodreporter.com.

Giardina, C. (2012). Avatar Sequels to Incorporate New Virtual Production Techniques.
The Hollywood Reporter. Retrieved from: http://www.hollywoodreporter.com.

Helzle, V., Spielmann, S. (2014). Draft Specification and Interface description of Virtual Production Editing Tools. Retrieved from: http://www.dreamspaceproject.eu/documents.

Introduction to Blink Kernels. (n.D.). Retrieved from: http://www.thefoundry.co.uk.

Johnson, P., Pettit, D. (2012). Machinima: The Art and Practice of Virtual Filmmaking. McFarland and Company.

Jones, A. (2012). FMX 2012: Worldbuilding for Independent Movies Q&A [video file]. Retrieved from: http://area.autodesk.com/fmx2012.

Jud, D., Michel, A. (2011). Motion Tracking Systems - An overview of motion tracking methods. Retrieved from: http://students.asl.ethz.ch.

Kaufman, D. (2013). VFX Crossroads: Causes & Effects Of An Industry Crisis. Creative Cow. Retrieved from: https://library.creativecow.net.

Kilkenny, J. (2012). FMX 2012: Virtual Production Opening Keynote [video file]. Retrieved from: http://area.autodesk.com/fmx2012.

Knop, T. (2014). Virtual Production - Methods Guidelines Scenarios. Retrieved from: http://www.dreamspaceproject.eu/documents.

Krause, J. (2014). STEM System: Demos zeigen Motion Tracker für Oculus Rift in Aktion. Next Gamer. Retrieved from: http://www.next-gamer.de.

Leberecht, S. (Director), Storm, C. (Producer). (2014). Life after Pi [video file]. Retrieved from: https://www.youtube.com.

Legato, R. (2012). FMX 2012: Closing Panel [video file]. Retrieved from: http://area.autodesk.com/fmx2012.

MacKenzie, S. (1995). Input Devices and Interaction Techniques for Advanced Computing. Virtual Environments and Advanced Interface Design, 437 - 470. Retrieved from: http://www.yorku.ca.

Mazalek, A., Nitsche, M. (2007). Tangible Interfaces for Real-Time 3D Virtual Environments. Proceedings of the international conference on Advances in computer entertainment technology, 155 - 162. Retrieved from: http://dl.acm.org.

McDowell, A. (2012 a). FMX 2012: Virtual Production Opening Panel [video file]. Retrieved from: http://area.autodesk.com/fmx2012.

McDowell, A. (2012 b). FMX 2012: Upside Down Worldbuilding for Independent Movies [video file].Retrieved from: http://area.autodesk.com/fmx2012.

McLuhan, M. (1964). Understanding Media: The Extensions of Man. Signet.

McSherry, T., Root, J., Fischer, R. (2014). Which Technology is right for a Project. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 388 - 397. Focal Press.

Montgomery, J. (2014). FMX 2014: day three/four report. FX Guide.
Retrieved from: http://www.fxguide.com.

Morin, D. (2012). FMX 2012: Virtual Production Opening Keynote [video file].
Retrieved from: http://area.autodesk.com/fmx2012.

Neufeldt, D., Baneham, R. (2011). The 3D Tech Behind Virtual Production [video file].
Retrieved from: https://www.youtube.com.

Nitsche, M., Kirschner, F. (2013). Hybrid Interface Design for Distinct Creative Practices in Real-Time 3D Filmmaking. Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, 303 - 306. Retrieved from: http://dl.acm.org.

Northam, L., Istead, J., Kaplan, C. (2012). A collaborative real time previsualization tool for video games and film. Proceeding ACM SIGGRAPH 2012 Posters.
Retrieved from: http://dl.acm.org.

Novy, D. (2014). Virtual Studio Technology. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 933 - 934. Focal Press.

Patel, M. (2009). Autodesk Whitepaper - The New Art of Virtual Moviemaking. Autodesk.
Retrieved from: http://www.autodesk.com.

Poupyrev, I. et al. (1996). The Go-Go Interaction Technique: Non-linear Mapping for Direct Manipulation in VR. Proceedings of the 9th annual ACM symposium on User Interface Software and Technology, 79 - 80. Retrieved from: http://dl.acm.org.

Root, J. (2014). What is Motion Capture. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 385 - 386. Focal Press.

Root, J., Edwards, G., Alexander, O. (2014). Facial Motion Capture. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 423 - 438. Focal Press.

Rosendahl, C. (2012). Carl Rosendahl, from PDI to the future of VFX. FX Guide. Retrieved from: http://www.fxguide.com.

Sargeant, B., Morin, D., Scheele, J. (2014). Virtual Production. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 443 - 445. Focal Press.

Sato, T., Nakamura, S., Terabayashi, K., Sugimoto, M., Hashizume, H. (2010). Design of Accurate Motion Capture System using Ultrasonic Communications. Proceedings of Symposium on Ultrasonic Electronics, 445-446. Retrieved from: http://www.use-jp.org.

Squires, S. (2014). Monster Sticks. In A. Okun, S. Zwerman (Eds.). The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures, 183 - 189. Focal Press.

Steinlechner, P. (2011). Razer Hydra ausprobiert: Bewegungssteuerung für PC-Spieler. Golem. Retrieved from: http://www.golem.de.

Sylwan, S. (2012). FMX 2012: Closing Panel [video file]. Retrieved from: http://area.autodesk.com/fmx2012.

Thacker, J. (2012). FMX 2012: In search of virtual production. CG Channel. Retrieved from: www.cgchannel.com.

Thomas, G. (2007). Mixed Reality Techniques for TV and their Application for On-Set and Pre-Visualization in Film Production. BBC Research White Paper. Retrieved from: http://www.bbc.co.uk/rd/publications.

Trumbull, D. (2012). FMX 2012: Closing Panel [video file]. Retrieved from: http://area.autodesk.com/fmx2012.

Williams, K., Rosendahl, C. et al. (2013). VES Whitepaper: The State of the Global Visual Effects Industry 2013 - An Analysis of Current Business Models and Better Business Practices. Visual Effects Society. Retrieved from: http://www.visualeffectssociety.com.

Wolfe, J. (2014). ZEUS:Scout Brings Virtual Production to the iPad. Animation World Network. Retrieved from: http://www.awn.com/news.

Workman, M. (2014 a). Writing the Future of Virtual Production - Q/A with Girish Balakrishnan. Cinematography Database. Retrieved from: http://www.cinematographydb.com.

Workman, M. (2014 b). The Cinematographer's Role in Virtual Production. Cinematography Database. Retrieved from: http://www.cinematographydb.com.

5.4. Image References

Fig.1: Virtual production keywords
Götz, K. (2015).

Fig.2: Virtualization of filmmaking
Götz, K. (2015). Based on Morin, D. (2012). FMX 2012: Virtual Production Opening Keynote [video file]. Retrieved from: http://area.autodesk.com/fmx2012.

Fig.3: Development of film budgets (left), increasing number of VFX shots (right)
Götz, K. (2015). Numbers derived from http://www.upcomingvfxmovies.com/svfx-shots-race.

Fig.4: The vacuous green void
http://www.photosquare.club/start-creating-your-own-green-screen-studio.

Fig.5: Traditional pipeline
Götz, K. (2015).

Fig.6: Scenes from ‘Avatar’
left: http://james-camerons-avatar.wikia.com/wiki/File:Jake_Sully_18_HD.png.
right: http://www.hfrmovies.com/tag/jon-landau.

Fig.7: Mocap approaches
Götz, K. (2015).

Fig.8: Inertial mocap suit from XSens
http://animationdigitaldigest.com/index.php/xsens-releases-redesigned-motion-capture-product-line-for-professional-animators.

Fig.9: Facial capturing in ‘Planet of the Apes Revolution’
2 x left: http://marketsaw.blogspot.de/2014/05/must-watch-new-performance-capture-bts.html.
2 x right: http://www.itsartmag.com/features/dawn-of-the-planet-of-the-apes-previs-by-mpc.

Fig.10: Jackson and Spielberg working with a virtual camera system (left), Cameron (right)
left: http://www.a113animation.com/2012/09/how-spielberg-is-saving-tintin.html.
right: http://german.fansshare.com/gallery/photos/376145/Lr-James-Cameron/?displaying.

Fig.11: NCam system
Götz, K. (2015).

Fig.12: Simplified virtual production pipeline
Götz, K. (2015).

Fig.13: Scenes from ‘The Adventures of Tintin’
left: http://www.blu-raydefinition.com/reviews/the-adventures-of-tintin-blu-ray-review.html.
right: http://collider.com/the-adventures-of-tintin-review.

Fig.14: Smart VCS system
left: http://replay.drexel.edu/research.html.
right: http://girishbalakrishnan.com/?attachment_id=7225.

Fig.15: Zeus Scout on tablet
http://presskitchen.com/2014/07/29/filmmaking-pre-visualization-reaches-peak-levels-zeusscout.

Fig.16: Generic editing workflow
Götz, K. (2015). Based on Helzle, V., Spielmann, S. (2014). Draft Specification and Interface description of Virtual Production Editing Tools. Retrieved from: http://www.dreamspaceproject.eu/documents.

Fig.17: Razer Hydra (right), Sixense STEM system (left)
left: http://nl.hardware.info/reviews/2455/razer-hydra-bewegingscontroller-review.
right: http://www.dailytech.com/Sixense+Aims+to+Bring+Immersive+VR+to+the+Masses+with+Its+STEM+System/article34124.htm.

Fig.18: Leap Motion controller
http://www.mjdinteractive.com/mjd-technology-spotlight-leap-motion.

Fig.19: Oculus Rift DK2
http://www.playm.de/2014/09/oculus-rift-spezifikationen-stein-gemeiselt-und-preis-zwischen-200-und-400-dollar-177966.

Fig.20: Icon design
Götz, K. (2015).

Fig.21: Object menu (left) and light menu (right) for mockup
Götz, K. (2015).

Fig.22: Mockup
Götz, K. (2015). background: http://www.photosquare.club/start-creating-your-own-green-screen-studio. tablet: https://www.sonynordicimages.net/vaio-duo11.

Fig.23: RET-VP prototype
Spielmann, S. (2015).

Fig.24: RET-VP interaction techniques
Götz, K., Seibert, S. (2015).

Fig.25: Interaction and menu scheme for set editing
Götz, K. (2015).

Fig.26: Arri Alexa with NCam system
left: Bolbeth, M. (2015).
right: Trottnow, J. (2015).

5.5. Acknowledgements

Let me take this opportunity to gratefully acknowledge the assistance and contribution of those who supported me throughout this project.

First and foremost I offer my sincerest gratitude to my supervisors, Prof. Katja Schmid and Simon Spielmann. Without their assistance and dedicated involvement in every step throughout the process, this paper would have never been accomplished. I would also like to thank Volker Helzle for his guidance and the opportunity to write this thesis, while Diana Arellano deserves great thanks for her overall support and the proofreading.

Moreover I would like to express my gratitude to my fellow Stefan Seibert. It was a pleasure to work with you in such a professional and ambitious way.

Special thanks go to our DOP Conrad Lobst and the entire on-set team, including Tobias Müller, Jonas Trottnow, Sebastian Moreno, Adrian Meyer, Patricia Griotto, Louis von Klipstein and Maik Bolbeth.

Furthermore I would like to acknowledge the crucial role of the staff of Filmakademie Baden-Württemberg, who gave the permission to use all required equipment and the necessary material to do the shooting.

I also have to thank my housemates for cheering me up after a long day of work. You are great!

Last but not least, and most importantly, none of this could have happened without my family. Especially to my parents, my uncle and my sister – Danke!

Kai