Creating a Workflow for Integrating Live-action and CG in Low-cost Stereoscopic Film Production

Abhishek Kala

A thesis submitted to

Auckland University of Technology

in partial fulfillment of the requirements for the degree of

Master of Creative Technologies

November 2010

Table of Contents

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Candidate's signature

Abhishek Kala

Intellectual Property Rights

Abstract

This thesis proposes a workflow that can be effectively used to integrate CG elements into live-action stereoscopic footage, while ensuring a comfortable viewing experience for the viewer. In stereoscopic 3-D footage, each eye is presented with a slightly different image of the same scene in order to create the perception of depth. The key question explored in this research is: how do stereoscopic images affect the integration workflow of CG elements into live-action stereoscopic 3-D footage?

Stereoscopic 3-D has been around since the 1800's with the first public viewing of a stereoscopic film occurring in 1922. The recent surge in visual effects intensive stereoscopic 3-D films is attributed to advancements in digital acquisition, digital post-production and digital projection of stereoscopic films. Studios across the world are using and testing different techniques to solve problems in the visual effects film industry; currently it is a very active area of research and experimentation. The motivation for this project came from the fact that there is no documentation available for a visual effects workflow that can be effectively used by studios and visual effects artists when transitioning into stereoscopic film production. This project will provide them with a workflow that they can use to integrate CG elements into live stereoscopic footage.

Brief History

Stereoscopic cinema can be briefly divided into four general periods through which the grammar of stereoscopic cinema has evolved (Zone, 2007).

- The Novelty Period. From 1838 to 1952 the emphasis was on the technology of 3-D and the "gimmick" of the three dimensional shots rather than on the narrative. During this period there was a constant battle between the technical and narrative demands of the medium. Some of the movies released during this period were: The Power of Love in anaglyph (1922), Robinson Crusoe (1941), etc.
- An Era of Convergence. From 1952 to 1985 the motion cameras of the 1950's converged on the 3-D subject matter, which characterized this period. More than fifty stereoscopic films were released between 1952 and 1955, allowing cinematic storytellers to experiment with various techniques to explore the narrative palette that the stereographic motion picture presented. Friday the 13th Part III from Paramount Pictures was the first movie that was widely released to over 100 silver screens in North America.
- The Immersive Era. 1986 to present. The inauguration of the 15/70-mm IMAX 3-D large format (LF) led to the next era of the stereoscopic cinema. Transitions, a 1986 IMAX 3-D film produced by Colin Low for the Vancouver Expo marked the start of the "immersive" era. The large seven-storey high image projected on the giant screen eliminates the awareness of the edge of the frame, as the viewer's peripheral vision is completely immersed in the stereoscopic image.
- Digital 3-D Cinema. 2005 to present. On 4 November 2005, the release of Chicken Little 3-D marked the beginning of the fourth stage of stereoscopic cinema. It was

released in 84 digital 3-D cinemas along with the wider release of its “flat” mode, 2-D version in 2,500 theaters.

Background Research

Binocular Discovery

Binocular vision had been the subject of scientific speculation for centuries till 1838, when Charles Wheatstone described stereopsis, the perception of depth in human vision, in a historic paper entitled “Contributions to the Physiology of Vision, Part the First: On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision”. He included a number of line drawings in the form of stereoscopic pairs that he prepared as a proof to his theory. These drawings were meant to be stereoscopically viewed in his new invention: the Stereoscope. This device used two centered mirrors positioned at a 45-degree angle to each eye, which reflected right- and left-eye images. It was the first instrument ever designed to view stereoscopic images and produce a 3-D effect (Zone, 2007).

This image
has been removed by the author of this thesis for copyright reasons.

Charles Wheatstone’s Steresocope (1833) (Zone, 2007)

This image
has been removed by the author of this thesis for copyright reasons.

Charles Wheatstone's Stereo Drawings (Zone, 2007)

Human Depth Perception

Stereopsis is the ability to perceive the world in 3-D. It relies on a number of cues that are either physiological (binocular cues) or psychological (monocular cues) in origin. Psychological cues include: perspective, relative size, texture gradient, occlusion/interposition, lighting and shading, aerial perspective, and motion parallax. Physiological cues include: binocular disparity and the vergence position of the eyes (Ukai, 2006).

This image
has been removed by the author of this thesis for copyright reasons.

(Corporation, 1997; Stereographics Corporation-Background on Creating Images for CrystalEyes and SimulEyes," 1997)

Binocular Disparity

Binocular Disparity is the difference in the two projections onto the left and the right retina caused by the horizontal separation (interocular/interpupillary distance) of our eyes. The average adult interpupillary (IPD) distance is around 63mm with a range between 50 to 75mm, the wider range of 45 to 80mm includes almost all the adults, and the minimum IPD for children (down to 5 years old) is around 40mm (Dodgson, 2004). Retinal point-to-point disparity variation across the two projections provides the information about the relative distances of the objects to the observer and the depth structure of objects and environments. The line that can be drawn through all the points in space that stimulate corresponding retinal points for a given degree of convergence is called the horopter. Points that are not on the horopter have retinal disparity. Points in front of the horopter are said to be crossed and disparities behind the horopter uncrossed. As long as the disparities do not exceed a certain magnitude, the two separate viewpoints are merged into a single percept (fusion). The small region around the horopter within which disparities are fused is called Panum's fusional area.

This image
has been removed by the author of this thesis for copyright reasons.

Panum's fusional area (Palmer, 1999)

If disparities are large the images will not fuse and double images will be seen, a phenomenon that is known as diplopia. The largest disparity at which fusion can occur depends on a number of factors such as the subject tested (i.e., individual variations), the amount of training a subject received, the criterion used for diplopia (unequivocal ”singleness” of vision compared with unequivocal ”doubleness” of vision), and conspicuous disparity (Duwaer and van den Brink, 1981).

Vergence-accommodation

In the natural environment, the distance at which your eyes converge is the same as the distance at which your eyes should focus. This is not the case with stereoscopic 3-D, where the images for both eyes are projected as two separate images on a screen. Assume we project a 3-D object that is meant to appear to be in front of the screen. Your left eye turns to look at the left-eye image of the object, and your right eye turns to look at the right-eye image of the object. Put together, your eyes converge (vergence) as if the object exists in front of the screen. As your eyes converge, your brain sends instructions to the eyes to focus the way they normally would for a real object at that convergence distance (accommodation). But in 3-D displays the “object” is in focus on the screen, which is behind this convergence point. So your brain keeps working your eyes until the “object” is in focus. That inescapable difference between how we naturally see the real world and how we see 3-D movies is called the vergence-accommodation conflict and has been suggested as a potential cause of visual strain (Hiruma & Fukuda, 1993; Schreer, Kauff, & Sikora, 2005; Wann, Rushton, & Mon-Williams, 1995; Yano, Emoto, & Mitsuhashi, 2004)

This image
has been removed by the author of this thesis for copyright reasons.

Difference between accommodation distance and vergence distance when looking at a stereoscopic image (Schreer, et al., 2005)

Factors Affecting Viewing of Stereoscopic Images

Asymmetrical Binocular Combination

Binocular summation is the process by which vision with two eyes is enhanced over what would be expected with just one eye (Steinman, Steinman, & Garzia, 2000). However large differences in the stimuli (pattern direction, contour, contrast, illumination, etc.) presented to the two eyes can cause binocular mixture, binocular rivalry and suppression, and binocular lustre (Schreer, et al., 2005).

- Binocular mixture typically occurs when a uniform field in one eye is combined with a detailed stimulus in the corresponding part of the other eye.
- Binocular rivalry may occur when corresponding parts of the two retinas receive very different high-contrast images. In a way, this is opposite of fusion, as the two monocular images will alternate repetitively, in whole or in part, with unseen portion somehow suppressed.

- Binocular lustre occurs in areas of uniform illumination in which the luminance, or colour, of the reflected light is different for the two eyes. In this instance, the images are stable and fused, yet appear to be shimmering or lustrous and cannot be properly localized in depth. It results from reflections being perceived differently by the two eyes due to the different position of the eyes, the lustrous appearance of surfaces like a car body or a waxed tabletop is largely attributed to binocular lustre (Schreer, et al., 2005).

Individual Differences

Some of the individual differences that affect stereoscopic viewing are IPD (interpupillary distance), vision disorders, and ageing.

- IPD (interpupillary distance) is the distance between the centers of the pupils of the two eyes. The mean IPD for the majority of the population is somewhere near 63mm, with extremes being 40 to 80mm, which takes into account children down to age of 5 (Dodgson, 2004). Stereoscopic images designed with an IPD of 60mm might have too large of a depth for children, possibly causing divergence leading to visual fatigue.
- Vision disorders that affect stereoscopic viewing are estimated to affect 5-10% of the population (M. T. M. Lambooij, Ijsselsteijn, & Heynderickx, 2009). According to psychophysical tests done by Richards (1970), four percent of his subjects (Members of MIT community) were unable to use the information provided by disparity and another 10% had great difficulty in seeing depth in Julesz stereograms (Schreer, et al., 2005). Two of the most common vision disorders that cause stereo blindness are:

- Strabismus
    - Strabismus is a disorder in which the eyes do not line up in the same direction when focussing. Commonly known as “crossed eyes”, it prevents proper binocular vision and depth perception. Prevalence of Strabismus is about 3-5 % (Graham, 1974; Noorden, 1990) and it is usually present at birth or develops early in life or may also occur secondary to diseases such as cardiovascular accidents, tumours, or trauma.

- Amblyopia
    - Amblyopia is a disorder of the visual system in which the visual acuity in one eye is reduced compared to the other eye without any apparent eye disease. Prevalence of amblyopia is about 1-2% (Kleinstein, 1984) and it usually develops during the critical period in development of the visual system (which occurs before the age of 4).

- Age is also a determining factor in stereoscopic viewing because visual abilities vary with age as a result of changes in the structure of the eye. Accommodative ability decreases with age, with a decline starting at 40 years continuing to about 55 years of age, after which point little or no accommodation remains (Ostrin & Glasser, 2004). Conversely, the visual system of children is not fully developed till the age of seven (Rushton & Riddell, 1999) and some ophthalmologists remain concerned that viewing stereoscopic images may cause strabismus in young children. There is no evidence for or against the hypothesis that viewing the stereoscopic images causes strabismus, except for a report by Tsukuda and Murai

(Tsukuda & Murai, 1988). They reported a case of a 4 year, 11 month old boy who manifested esotropia after viewing stereoscopic animation at a cinema using an anaglyph. Children should be cautioned about stereoscopic images because they may not subjectively perceive a problem even if an eye is deviated (Noorden, 1990). Due to the same concern, lots of companies manufacturing new hardware are printing warnings indicating that their products should not be viewed by children (e.g. Fujifilm cameras).

Keystone Distortion

Keystone Distortions are introduced in a stereoscopic image when it is captured with a converging camera setup where the left and the right cameras are toed in. This introduces trapezoidal distortion because the two images in effect have placed one portion of the object closer to the camera than the other resulting in differential magnification across the image field. Keystone distortion is most noticeable in the image corners and increases with increasing camera base distance, decreasing convergence distance, and decreasing lens focal length (Schreer, et al., 2005).

The vertical parallax created due to the converged camera setup can induce eye strain and visual discomfort (A. Woods, Docherty, & Koch, 1993).

This image
has been removed by the author of this thesis for copyright reasons.

Unwanted horizontal and Vertical parallax due to converging camera setup (exaggerated for illustration purposes) (Schreer, et al., 2005).

Shear Distortion

While viewing stereoscopic images if the observer changes position or moves to the side, the stereoscopic image appears to follow the observer causing the perspective to appear distorted. This distortion is called shear distortion. Objects with positive parallax move in the opposite direction to the observer's movement and objects in the negative parallax move in the same direction (A. Woods, et al., 1993).

Cross Talk

Cross talk results from imperfect separation of the left- and the right-eye views (i.e. when the right-eye view leaks through to the left-eye view and vice versa). It is perceived as ghosting, blurring or double contours and is one of the main factors responsible for visual discomfort while viewing stereoscopic images. The perceptibility of cross talk increases with contrast and disparity values, even small amount of cross talk can lead to problems(Pastoor, 1995). Cross talk is present in various levels with most stereoscopic displays but it is mostly evident with anaglyphic 3-D glasses (A. j. Woods & Rourke, 2004), it is also quite evident with linear polarization techniques if the head position of the observer is tilted (e.g. IMAX screen on Queen Street in Auckland, New Zealand) (Schreer, et al., 2005).

Picket Fence Effect and Image Flipping

Picket fence and image flipping are artifacts found in multi-view auto-stereoscopic displays, both the artifacts are perceived when observers move laterally in front of the display. Picket fence is the appearance of vertical banding due to the black mask between the columns of pixels in the LCD. Image flipping is the noticeable transition between the two viewing zones (Sexton & Surman, 1999).

2-D to 3-D Conversion

To meet the increasing demand for the 3-D content, real-time 2-D to 3-D conversion seems to be a promising method. Research has demonstrated that with 3-D television, content depth has to only approach reality to create an acceptable 3-D percept (Meesters, Ijsselsteijn, & Seuntiens, 2004). The development of algorithms for generating depth maps is based on an assumption that an accurate depth map is not required to create an acceptable depth impression which is based on assumptions, estimations, and heuristic cues (Battiato, Curti, LaCascia, Tortora, & Scordato, 2004; Redert, et al., 2007; TARN, et al., 2007). The inaccuracy of depth maps can lead to artifacts like spatial and temporal inconsistencies, e.g. parts of objects get assigned incorrect depth values leading to incorrect depth layering. This can lead to unnatural visualizations; e.g. flickering of the image and turbulence around edges.

Unnatural visualizations may also result from dis-occlusion. This happens when the occluded area in the original 2-D image becomes visible in one of the virtual views. Since no information is available of the occluded objects in the original image, the missing areas (often referred to as holes) must be replaced with useful colour information (Fehn, 2004). There are different algorithms proposed for this hole-filling process (Mark, McMillan, & Bishop, 1997; Shade, Gortler, He, & Szeliski, 1998), but they all have the same shortcoming: the occluded area is never fully correct, but is always interpolated from existing information. Hence 2-D to 3-D conversion cannot be fully accurate and the artifacts introduced due to this process may induce visual fatigue (M. Lambooij, Ijsselsteijn, Fortuin, & Heyderickx, 2009).

Definitions

Vocabulary and concepts used frequently during this project.

Parallax:

While viewing stereoscopic footage two different slightly offset images are presented to the eyes to produce disparity so that our brain can perceive depth in the stereoscopic image. Parallax and disparity are similar entities, the only difference is that parallax is measured at the display screen and disparity is measured at the retina. While watching stereoscopic footage parallax becomes retinal disparity, thus producing stereopsis. There are four kinds of parallax that one may encounter while viewing stereoscopic images.

Zero parallax is when the homologous image points of the two images lie exactly on top of each other, part of the image with zero parallax is perceived on the screen plane. This is the only setting when the accommodation of the eyes is at the same point as convergence and the image is comfortable to view. When the image points have zero parallax, they are said to have zero parallax setting (ZPS).

Positive parallax is when the eye axes are between zero and parallel (eye axes are parallel when viewing distant objects). In this setting the part of the image is perceived behind the screen plane.

Divergent parallax occurs when the homologous image points of the two images are further apart than the interpupillary distance (i.e. distance between the two eyes) it is unnatural and is a major cause of eye strain while viewing stereoscopic footage.

Negative parallax is when the eyes' axes are crossed. Parts of the image with negative parallax appear to be between the observer and the screen, giving the appearance of objects coming out of the screen plane into the viewer's space.

While viewing stereoscopic 3-D footage, the screen is more like a window into another world and it can be used for creative advantage in various ways. For instance, the objects can be behind the stereoscopic window, at the window (screen plane), or in front of the window (also referred to as "off screen" effects).

Floating Window:

Watching stereoscopic footage on a 3-D monitor is like looking through a window, where position of the objects in 3-D space is defined by the relative position of the edges of the screen. Objects can be behind the screen and can also come out of the screen into the viewer's space.   If an object is in the viewer's space, for example, if the object is in front of the screen and is occluded/cut or intersected by one of the edges, it creates an effect called stereoscopic window violation (SWV), and it is also one of the main causes of viewers discomfort while viewing stereoscopic footage.

This image
has been removed by the author of this thesis for copyright reasons.

If an object hits the side frame when in front of the screen, it generates a stereoscopic window violation (Mendiburu, 2009).

Stereoscopic window violation can be fixed with the help of floating windows. A floating window is created by adding a mask on either side of the stereoscopic pairs (left/right images), which creates a virtual screen that can be pulled into the viewer's space to bring the action closer to the audience. This floating window can be static or animated, or asymmetric. The virtual screen can be shaped and manipulated in many different and creative ways with the help of floating windows.

Masking the right side of the right image brings the right edge of the screen towards the audience, if the left side of the left image is masked at the same time, the whole screen is moved towards the audience (Mendiburu, 2009).

Screen Size and Viewer Distance:

Target screen size and viewer's distance are two very important considerations to consider before starting any stereoscopic production. Native pixel parallax (NPP) is the term widely used to determine the maximum positive pixel parallax in the background on a given screen between the two images (left/right) that is equal to the interocular separation (63 mm), it ensures that viewing the stereoscopic footage will not cause your eyes to diverge, which is one of the main factors that cause visual discomfort while viewing stereoscopic footage. NPP can be easily calculated by dividing the resolution of your screen by its width and then multiplying it by 2.5. So for an LG screen that is 20 inches wide and has full HD resolution (1920 multiplied by 1080 pixels) NPP will be $(1920 \div 20) \times 2.5 = 240$ pixels. The same footage projected on a 30-foot cinema screen with a NPP of (1920 ÷

360) × 2.5 = 13.5. 13.5 pixels will cause instant eye strain and headache, and the footage will be un-fusible because the distance between the two images will be almost 457 mm!

Viewer distance is another important aspect that has to be considered before starting a stereoscopic production. It has been noted by Lambooij that horizontal disparity limits are closely related to depth of field, and beyond the one degree limit there is a zone of increasing visual discomfort up to a value where diplopia appears (M. Lambooij, et al., 2009). Currently this limit is regarded as a standard reference value. Others have recommended lower limits of 0.82 degrees as well (Yano, et al., 2004).

If stereoscopic footage is produced for an HD monitor (20 inches wide), and has a positive parallax of 2.5 inches in the background, then the safe parallax angle of 1 degree will be achieved at a distance of 12 feet, but usually the distance of the viewer on a computer HD monitor is around 2 feet! Around 0.35 inches will be the safe maximum positive pixel parallax, if the viewer's distance is 2 feet.

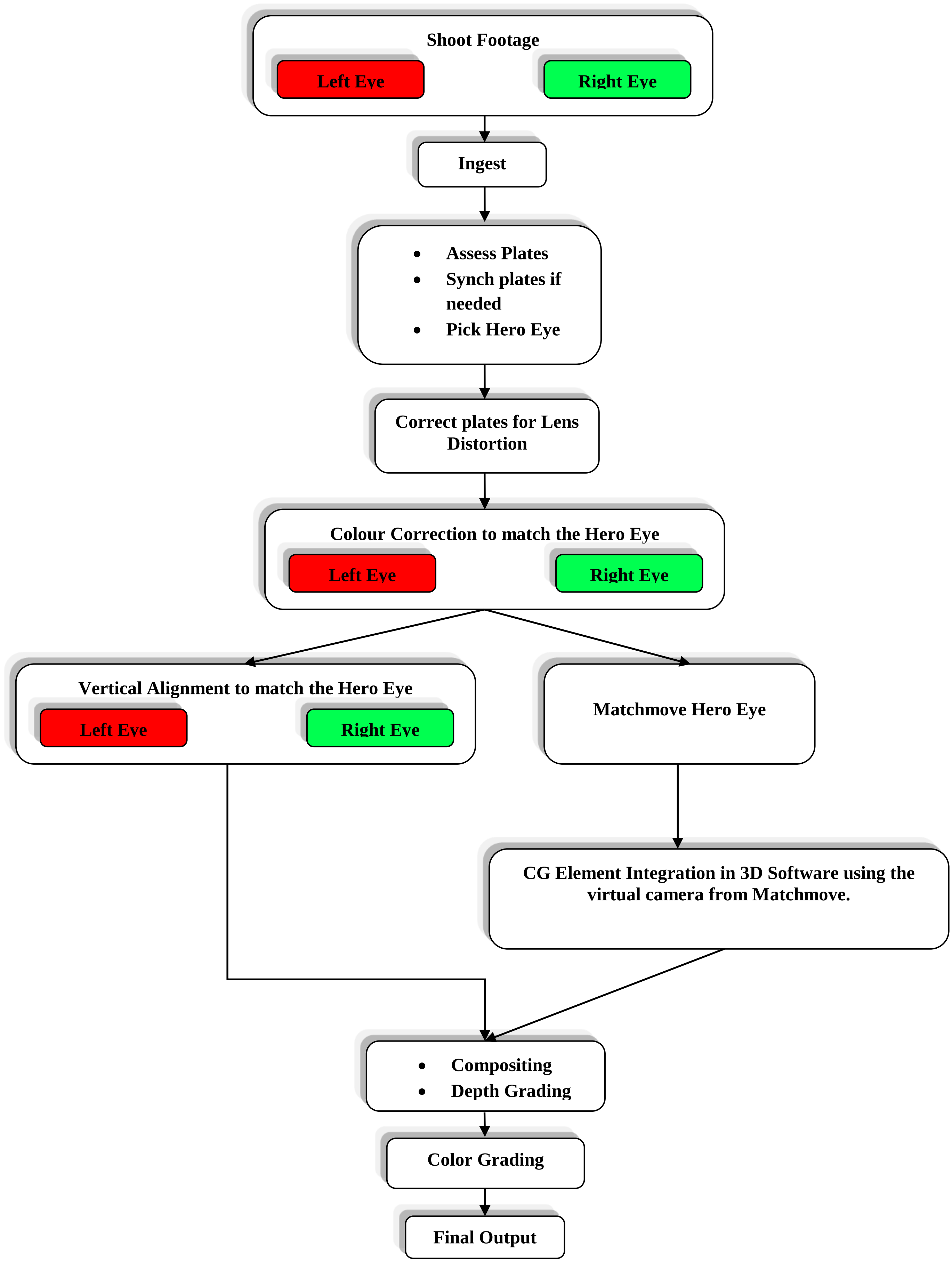
$$\beta = 2 \arctan \frac{P}{2d}$$

β = Parallax angle

P = Pixel Parallax

D = Viewer distance from the screen

Recommended Workflow for Integration of CG Elements into Stereoscopic Footage



Shooting

Process

> "Like many others who came before me, I had no idea how terribly difficult it is to make such [stereoscopic] equipment perform properly. The naive position is that since one is simply using two of this and two of that, stereoscopic film-making ought to be fairly simple. I had no notion that stereoscopic photography was in many ways substantially different from conventional, or planar, photography. The danger with stereoscopic film-making is that if it is improperly done, the result can be discomfort. Yet, when properly executed, stereoscopic films are beautiful and easy on the eyes (Lipton, 1982)".

Shooting of stereoscopic footage requires two cameras configured in such a way that both the cameras are aligned and calibrated in every respect, such as the white balance, shutter speed, shutter angle, exposure, aperture, clock (genlock), frame-rate, gain, etc. (Mendiburu, 2009). The only difference between the two cameras should be the horizontal offset, which is also referred to as interaxial separation or stereo base, changing the interaxial separation affects the depth of the scene, it also needs to be adjusted according to the screen size that the shot is intended for (the shoot and tests for this project were intended to be displayed on a computer monitor or an HDTV, therefore the 1/30th rule was applied for the interaxial distance). Perceived depth in a shot is directly proportional to the interaxial separation, for example, when the interaxial separation (separation between the two cameras) is larger than the interocular separation (separation between the two eyes) the resulting effect is called hyperstereoscopy, which results in depth exaggeration and miniaturization in the scene. Hypostereoscopy, or cardboarding is when the interaxial separation is smaller than the interocular distance.

Challenges

Shooting in 3-D is a very complex task, at least in the current scenario where not many ready 3-D camera rigs are available. The first, and most important decision that needs to be made is for what screen size the output is intended? To illustrate this better, consider an example: footage is shot for a 30-foot screen and the positive parallax of the object placed at infinity is set to 2.5 inches apart. When the same footage is now projected on a 60-foot screen the positive parallax of the object at infinity will be 5 inches! This will cause the viewer's eyes to diverge, causing visual fatigue and possibly a headache to develop within minutes.

Decision also needs to be made in regard to the type of rig configuration to use: either a side-by-side rig or a mirror rig (beam-splitter rig)? For this project, the decision was taken in favour of a mirror rig because it allows for an interaxial separation of 65mm and less for close up shots, whereas it was not possible to achieve the same interaxial separation with the side-by-side rig due to the size of the cameras. Although mirror rigs allow for a closer interaxial separation, they have disadvantages. For instance, the size and weight is the most problematic disadvantage, as it becomes really difficult to maneuver the rig. Mirror rigs are also susceptible to lens flares and reflections because of the half-silvered mirror. Additionally, if one camera shoots off the reflection, image quality suffers, as it reduces the exposure level in the image that is being shot off the reflection. Also, these rigs are prone to dust interference.

Common challenges pertaining to both the rigs are:

Zoom: In prosumer cameras one has to rely on the LCD screen to read the zoom values and even though the values in the two cameras are the same there might be slight inconsistencies.

Focus and Iris: In prosumer cameras, it is really difficult to match the focus and iris of the two cameras manually, in spite of the LCD screen displaying identical values consistencies are inevitable.

Polarization, Reflections and Flares: Polarization, reflections and lens flares are a 3-D hazard because they are perceived differently by the two cameras due to the different position of the two cameras.

Synchronization Issues: In prosumer cameras, it is really hard to synch the start of the cameras. Therefore, clapper board is a must so that the shots can be synched in post-production.

On-set Live Monitoring: One of the most crucial necessities of a stereoscopic shoot is to have an on-set live monitoring system. Lack of on-set live monitoring affected this project as well, as lots of problems surfaced in the shot footage during post-production that could have been avoided while shooting with the aid of on-set live monitoring.

<u>Solution/Conclusion:</u>

Shooting in 3-D is a very arduous process, especially with a basic rig. For this project various steps were taken during the prep work that made it possible to shoot the stereoscopic 3-D shot.

The footage for this project was shot using the 1/30th rule, which states that your interaxial separation should be 1/30th of the distance from your camera to the closest subject. If your cameras are 2.5 inches apart then your closest object should not be closer than 75 inches, or 2 meters. This rule works well for small size screens such as computer monitors and HDTV's, for big screens it's typically 1/60th or 1/100th.

Both the cameras used for this project were identical and gen lockable with identical lenses. The cameras were clearly marked as left and right using different coloured tape and all the cables for each camera were colour coded as well. All the settings on both the cameras were set to manual mode, and steadishot was deactivated on both the cameras. Once the cameras were mounted on the mirror rig, interaxial distance was set to 6.5cm using a ruler.

When mounting the mirror it is really important to check the side that will be reflecting the top camera. This can be done by touching a fingernail to the mirror. If the fingernail joins the reflection of itself then it is the right side, if there is a gap between the fingernail and its reflection then the mirror needs to be flipped. It is important to set the mirror so that it is set at a 45 degree angle and to achieve that a set square was used.

Light proofing the rig was one of the most difficult challenges. Lots of black tape was used to cover the gaps and a black cloth was used to cover the top camera to avoid light leaks. A white board was used to white balance both the cameras.

Even though there were measures taken to achieve a good stereoscopic shot, because of the bad mechanics of the rig the acquired footage needed lots of post-processing to make the shot comfortable to view. Professional rigs can help in acquiring better quality footage, thereby minimizing the post-production required to fix the problems that lead to visual fatigue. Rigs such as the one developed by Vince Pace, President and CEO of Pace (http://www.pacehd.com), or that of P+S Technik (http://www.pstechnik.de/en/3d-rig.php) can help in minimizing the problems as they are motorized and have options to synch the two cameras. One factor that still remains is that these rigs are really bulky; the solution lays in development of cameras like the Panasonic AG-3DA1 (http://pro-av.panasonic.net/en/3d/ag-3da1/index.html) an integrated twin-lens full HD 3-D camera recorder. The only issue with the Panasonic camera is that one cannot change the interaxial separation of the lenses because they are fixed, therefore limiting its use, as it cannot be used for extreme close-ups and while shooting landscapes because the interaxial separation cannot be increased to increase the roundness factor of the objects in the scene.

The on-set live monitoring tool is a definite must for stereoscopic shooting as it allows a live monitoring of two gen locked cameras on a single 2-D monitor. Viewing an anaglyph output of the two cameras live on a 2-D monitor can greatly help in aligning and monitoring the two cameras while shooting. The quality of the footage from the shoot for this project could have been much better even with the basic rig had such a monitoring tool been used. Stereobrain, a 3-D processor from Inition (http://www.inition.co.uk/inition/dispatcher.php?URL_=product_stereovis_inition_stereobrain_p&model=products&action=get&tab=summary) and Transvideo 3-D view monitor (http://transvideointl.com/pages/english/products/index.htm) are currently two of the major players in on-set live monitoring of stereoscopic 3-D.

Ingesting and Assessing Stereo Plates

Process:

The shot for this project was ingested using Adobe Premiere CS5. At this stage the shot has to be carefully assessed for synch, flares, polarization, exposure differences, colour differences, etc. The decision has to be made to choose the hero eye, for this project the right eye (right camera) was chosen as the hero eye because the footage was shot using a mirror rig, and the right camera was shooting straight through the mirror, whereas the left camera was mounted on top and was shooting the reflection. The image shot of the reflection is usually soft, and has exposure difference as well due to the mirror.

Challenges:

The biggest challenge at this stage was synch! The two cameras had to be started manually, one after the other, because the controller that fires both the cameras in synch was not available. There was considerable colour and exposure difference between the two eyes.

Solution/Conclusion:

A clapperboard was the saviour when it came to synching the shot. It is highly recommended that no stereoscopic footage should be shot without a clapperboard in the beginning, as the clap can be used to synch the shot by moving one eye in the timeline in Adobe Premiere or any other editing program to match the clap in the other eye. Ingesting and assessment of the stereo shots could also be done using a few advanced tools available like Cineform's Neo3D, which offers some really creative tools to help in synching the shots (clapperboard still needed) and allows a real-time workflow in Adobe Premiere for stereoscopic editing. There are other high end tools like Quantel Pablo (http://www.quantel.com/list.php?a=Products&as=Stereo3D), Mistika (http://www.sgo.es/products/sgo-mistika-4k-2k-hd-sd/), and Assimilate Scratch (http://www.assimilateinc.com/scratch.html).

Correcting Lens Distortion

Process:

Lens distortion is a type of aberration that causes images to become stretched or compressed around the edges of the frame (Dobbert, 2005). It is more prominent with wide angle lens and around the edges of the frame.

This image
has been removed by the author of this thesis for copyright reasons.

Lens Distortion Pipeline (Dobbert, 2005)

It is more visible in shots with straight lines, such as shots that include objects like buildings. The shot used for this project was shot in a forest where the effect of lens distortion is not noticeable because of the absence of straight lines in the shot. Correction for lens distortion was not used throughout this shot completion; this workflow needs to be further tested with architectural stereoscopic 3-D shots where correction for lens distortion will be definitely needed to integrate the CG element seamlessly with the background plate.

Colour Matching

Process:

Colour correction in a traditional 2-D film is used to correct exposure, evoke mood or enhance the look of the film but in stereoscopic footage there is one more added complex challenge. How to match one image to the other (left to right, or vice versa)?

Sample image showing the colour and exposure difference in the footage shot for this project.

Challenges:

A mirror rig was used to shoot the footage for this project and the right eye was treated as the hero eye. This decision was made based on past research that shows that 2/3rd of the population is right-eye dominant (Chaurasia & Mathur, 1976; Ehrenstein, Arnold-Schulz-Gahmen, & Jaschinski, 2005). Picking the right eye as the hero eye means that all the corrections in post-production will be made to the left eye to match the right eye. If a side-by-side rig is used it wouldn’t matter which eye is chosen as the hero eye, but with a mirror rig it is important as the camera shooting off the reflection of the mirror loses sharpness and suffers from exposure and colour problems. Therefore, all the shots were shot with the right-eye camera looking straight through the mirror and the left camera mounted on top shooting the mirror’s reflection.

Standard colour correction and grading techniques could be used to match one eye to the other; masks can be created by using Bézier shapes to separate different sections of the image so that all the elements in the image could be colour corrected locally. One of the main drawbacks of this technique is that the Bézier shape needs to be animated throughout the shot so that it follows the contours of the element in the moving image sequence, which would be very time consuming, as it would be a similar task to rotoscoping.

Solution/Conclusion:

Foundry's Ocula suite of plug-ins (http://www.thefoundry.co.uk/products/ocula/) were developed while being used in many large stereoscopic productions, for example, Avatar was used for this project to aid in colour correcting. A plug-in that comes with Ocula is O_ColourMatcher, which provides various options to speed up the colour matching process. It is capable of doing a global colour match by matching the histogram of one image to another, or block-based matching—where it divides the image into number of user-specified block sizes, and histogram of each block is matched to the corresponding block in the other image by utilizing advanced image processing algorithms.

Basic global matching

Block-based colour matching

Although block-based matching gives better results it does not give a perfect solution, the results of the block-based matching method rely on Ocula being able to accurately pick the two corresponding points in the two images (left and right) in every frame of the moving sequence. It is able to produce a reasonably good result in an image with clear identifiable features. In an environment such as a forest, where all the leaves and trees having the same colour information, it is quite easy for Ocula to get confused while trying to pick two corresponding points in the two images in all the frames. This can lead to temporal inconsistencies, which introduces flicker in the image.

Block-based matching was tried initially for this project but because of the busy environment it suffered from temporal inconsistencies, therefore, O_ColourMatcher's basic global matching was used to bring the colour and exposure of the left image closer to the right image and then standard 2-D colour correction techniques were used to match it better.

An ideal solution will be a camera with perfectly calibrated exposure, gain, and colour. New cameras like Panasonic AG-3DA1 (http://pro-av.panasonic.net/en/3d/ag-3da1/index.html) will definitely reduce if not eliminate the colour and exposure inconsistencies.

Matchmove

Process:

Matchmoving is one of the most crucial processes in visual effects for integration of CG elements into live action footage. Matchmoving enables insertion of CG elements into live action footage so that the position, scale, orientation and motion of the CG elements are relative to the shot footage. The process of matchmoving involves generating a virtual camera by tracking the movements of a camera through a live-shot scene, so that the tracked virtual camera can be used inside 3-D programs to match CG elements to the shot footage.

This image
has been removed by the author of this thesis for copyright reasons.

Basic matchmoving workflow (Dobbert, 2005)

The above flowchart shows a typical pipeline for matchmoving that starts with the evaluation of the footage. There are several ways to solve a matchmove, therefore scrutinizing the shot is very important so that complications are identified that might arise during the process of matchmoving. After analyzing the plate any additional data (camera information, set measurements, and survey data) if available from the shoot is fed into the matchmoving software to aid it in calculating an accurate camera track. Once the camera is tracked and solved it is very useful to construct a spatial layout of the environment on the live-action plate. This helps to see if the proxy geometry fits in with the geometry in the live-action plate as well as helping to check if the geometry locks on to the geometry in the live-action plate. If it doesn't then a slide in the geometry will be visible when the camera moves across the scene. Once a good camera solve is achieved it is exported into a format that can be read by other artists on their respective software.

This image
has been removed by the author of this thesis for copyright reasons.
(Dobbert, 2005).

Challenges:

Matchmoving a stereoscopic shot should be easier and more accurate because the matchmoving software has more information available to solve a track (for example, two camera viewpoints of the same scene). But the challenge is that all matchmoving software requires very accurate information about the cameras interaxial separation, which is fine if a motorized rig is used, which will provide accurate measurements, but when a basic rig is used it is really difficult to get accurate measurements as a measuring tape is relied on for most of the measurements. Another problem is the vibration between the two cameras on moving shots due to the mechanics of the rig.

Solution/Conclusion:

The solution implemented for tracking shots for this project was to track only the hero eye (right eye) and create the other camera in the 3-D software (Maya) based on the measurements taken on-set while shooting. The ideal solution would be a very sturdy professional camera rig with the information of the interaxial difference encoded in the metadata of the shot footage.

CGI Integration

Process:

Another virtual camera is created to simulate the left eye which is set horizontally apart from the main tracked camera (right eye) based on the measurements taken at the shoot location. Based on the camera solve information, a CG element is placed at the right position. Scale and perspective of the CG element is also matched to the live-action plate. The CG element is modeled, textured, and lit to match the scene lighting using HDRI images and additional lights in the 3-D program. Once the CG element is complete with all the details to make it look seamless in the live-action plate, it is then rendered as an image sequence with various passes. These image sequences are then taken into a compositing application (Nuke). Compositing is the process of taking images from variety of sources and combining them in such a way that they appear to be shot at the same time under the same lighting conditions with the same camera.

Challenges:

In Maya, creating the other camera based on the measurements did not give the correct result, as the CG element appeared to be at a different depth than the object that it was supposed to be sitting on in the live-plate. In addition to camera, scale, perspective, and lighting, stereoscopic footage introduces depth as another element that has to be matched between the live-action plate and the CG element so that the CG elements appear spatially correct.

Solution/Conclusion:

To correct the depth of the CG element the left camera was shifted horizontally, until the pixel separation of the two corresponding pixels of the CG element matched the pixel separation of the two corresponding pixels of the live-action plate (located at the depth where the CG element is supposed to be integrated).

Vertical Disparity Correction

Process:

Vertical parallax, as mentioned in the stereoscopic background section, can induce eye strain and visual discomfort (A. Woods, et al., 1993). Vertical disparity in stereoscopic footage may come from misalignment of the two cameras, difference in focal length, key stone distortion, or from non-linear distortions (Ronfard & Taubin, 2010). Vertical disparity can be fixed by scaling, rotating or shifting the left or the right eye image so that the only difference between the two images is horizontal mis-alignment.

Challenges:

As mentioned earlier, due to the bad mechanics of the rig and absence of an on-set live monitoring tool, it was very difficult to analyze and avoid camera alignment problems. There were huge vertical parallax problems in the footage shot. Consequently, the shots were not fusible and would cause almost instant eye strain and headache if viewed for long periods of time. Therefore, vertical parallax had to be fixed in post-production to make the shot comfortable to view.

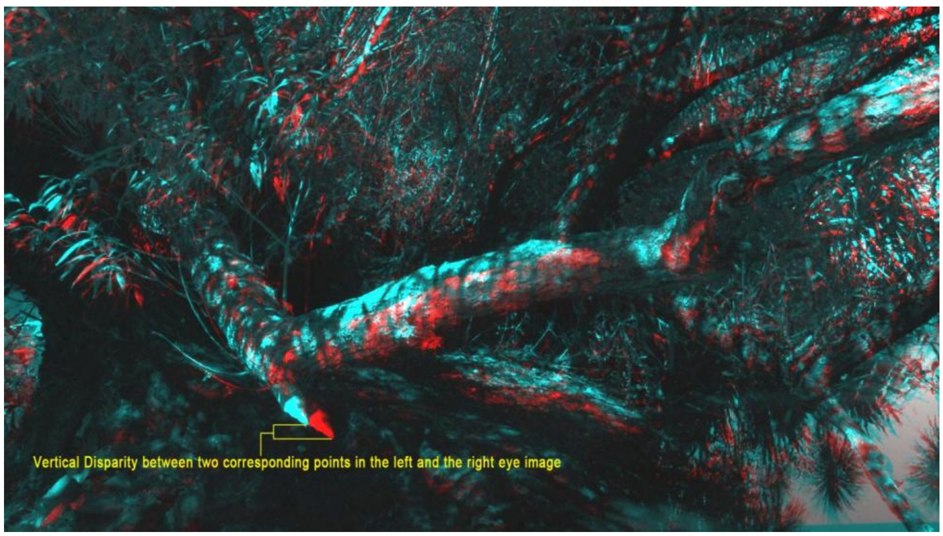


Image displaying vertical disparity between the two corresponding points in the left and the right eye image (anaglyph mode).

Solution/Conclusion:

Vertical disparity can be fixed in post-production by rotating, shifting and scaling one image to match the other image till the only difference between the two images is horizontal offset. Camera rotations and slightly different lens distortion from the two lenses contributed to the fact that vertical disparity was more prominent around the edges than in the centre of the image. Therefore the image had to be warped in order to fix the vertical disparity. Warping the image manually was not a very viable solution, because the warp would need to be animated, which is a very tedious job and might also result in temporal inconsistencies. The decision was made to try the Ocula suite of plug-ins from The Foundry to fix the vertical disparity in the footage. The suite includes a set of plug-ins (such as O_Solver and O_VerticalAligner) to correct the vertical mismatch between two images (the left and right eye images).

O_Solver defines the geometric relationship between the two images used as input (left and right eye) by detecting features in one view and then locating the corresponding features in the other view.

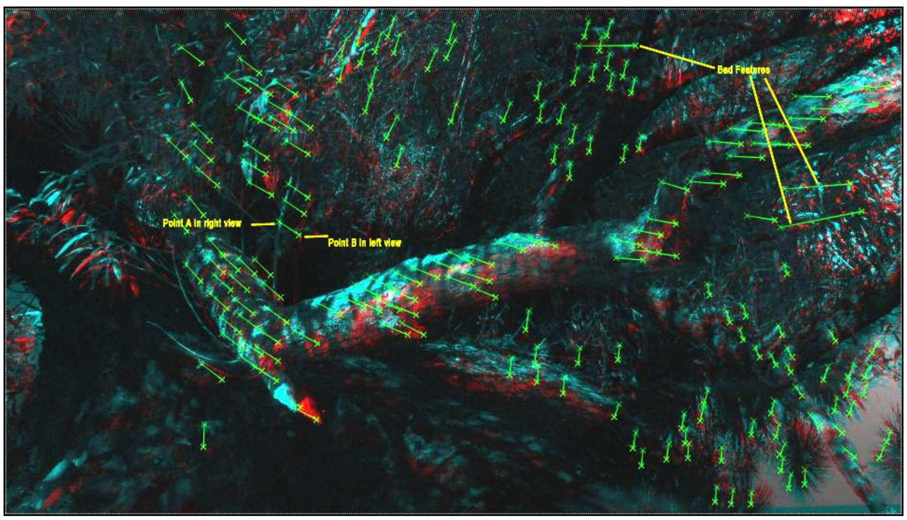
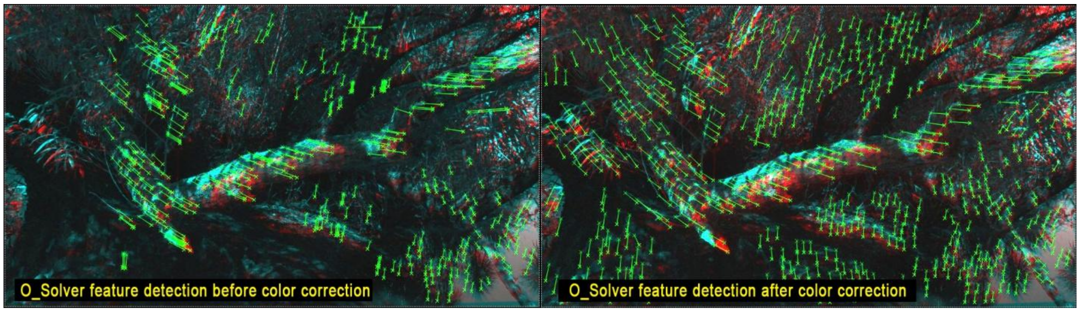
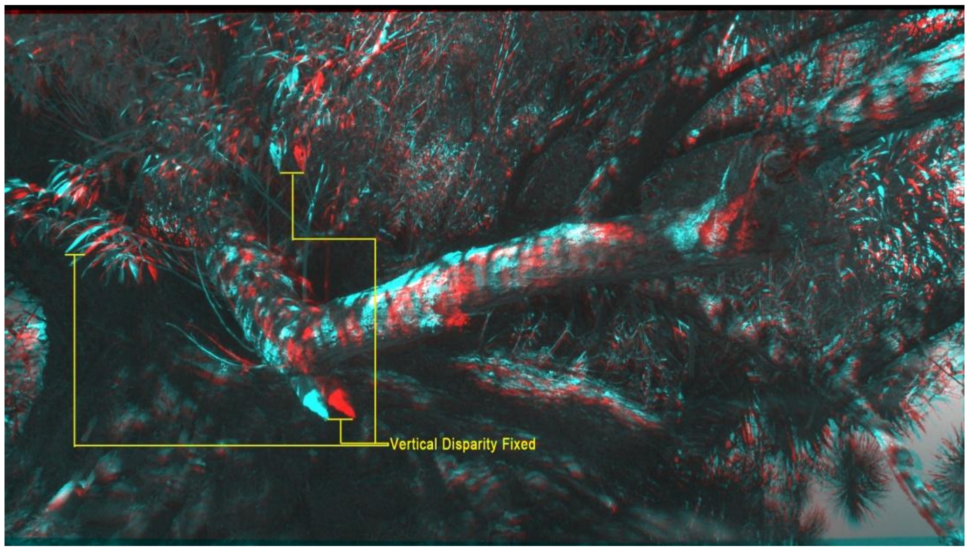


Image displaying features detected by O_Solver in the two images.

The accuracy of the solver is negatively affected by colour and exposure differences in the two images. The accuracy is additionally compromised by the fact that one eye (the left image) is softer because of the mirror being used in the rig. This was the main reason for doing colour correction first and then moving to vertical disparity correction, because colour correction increases the accuracy of the O_Solver plug-in, thus improving the quality of the vertical correction.



Improvement in the O_Solver feature detection due to colour correction.



Vertical Disparity Fixed

Compositing/Depth Grading

Process:

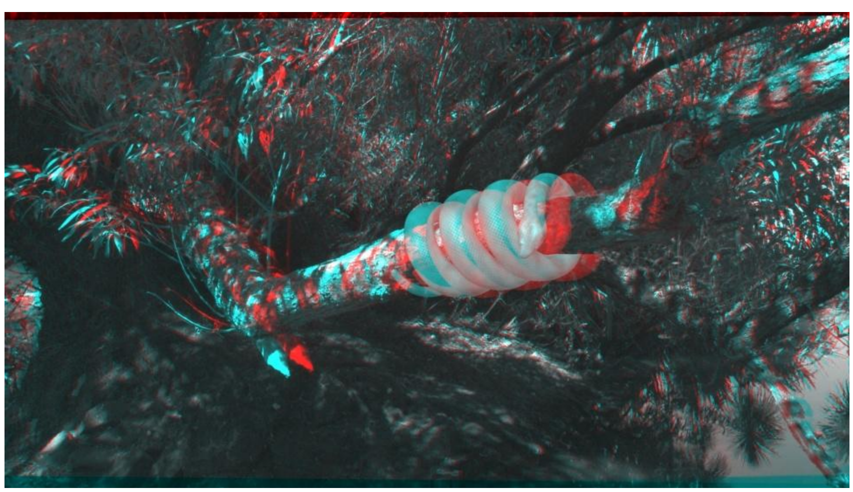
Compositing is the process of integrating and combining elements from various sources and manipulating them in such a way that they appear to have been shot at the same time under the same lighting conditions with the same camera.

Compositing of stereoscopic images introduces one more element,"depth", which has to be carefully dealt with for every element so that its 3-D shape is consistent with the original stereoscopic scene. The process of adjusting depth in a stereoscopic scene is called "depth grading". Adjustments could be for one element (or number of elements) in the scene or the whole scene. Adjustments are also made for different screen sizes in order to avoid divergence.

Another challenge is that instead of one image sequence there are two image sequences (left and right eye) for any given shot, which means that any process done for one eye has to be replicated to the other eye as well. Take rotoscoping as an example, which is a very tedious task, with stereoscopic footage post-production techniques have to be applied twice, once for each eye for each shot, which increases the shot completion time considerably.

Challenges:

The O_VerticalAligner plug-in warps the left image to fix the vertical disparity, which introduces artifacts at the top and the bottom of the left image. The footage for this project was shot using a parallel rig, causing there to be no ZPS (zero parallax setting), meaning the whole scene is in the viewer's space. Therefore, depth grading had to be done in order to set a convergence point in the scene. One side effect of setting a convergence point on the python (the CG element in the scene) is that all objects in front of the python are pushed into the negative space, that is, the viewer's space/in front of the screen, which creates a window violation that has to be treated using floating windows.



Huge horizontal disparity and artifacts caused by vertical aligner (top/bottom edge) are visible in the above image.

Solution:

The two images are scaled and cropped to remove the artifacts at the top and the bottom.

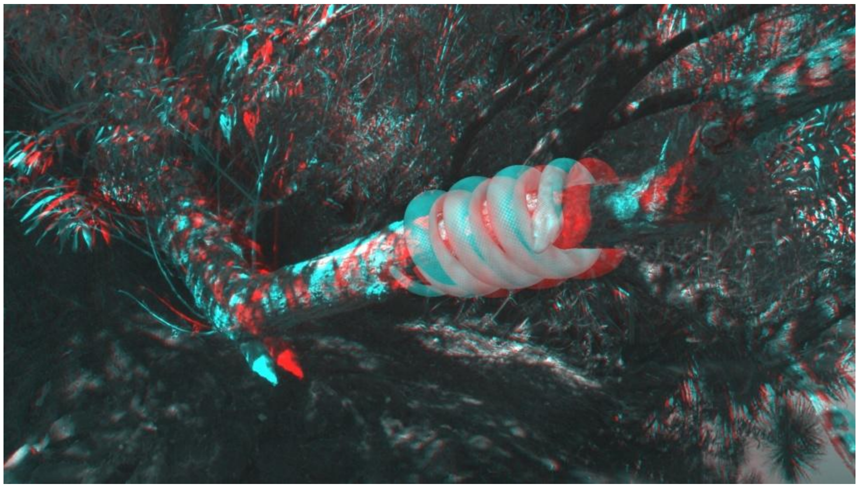


Image with the cropped edges in order to remove the artifacts caused by vertical alignment.

HIT (horizontal image translation) is applied to the images to set the plane of convergence in the scene.

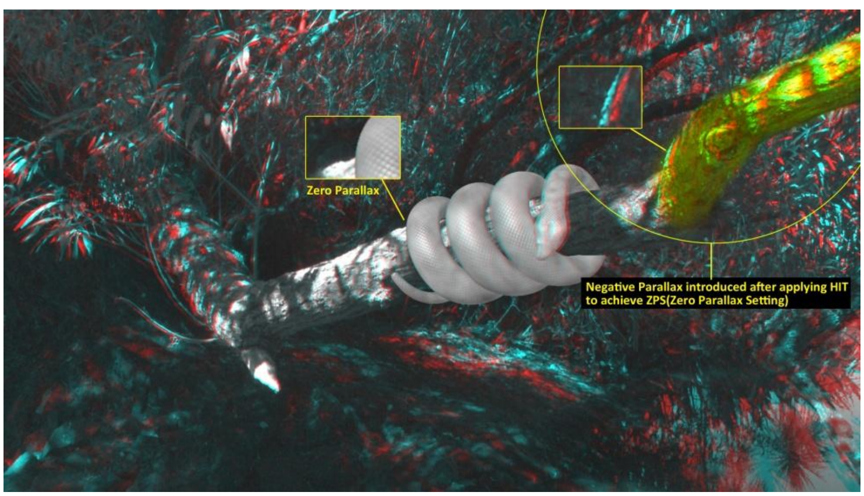


Image showing the marked area of the branch that is pushed into negative space once convergence point is set on the python.

The right image is masked on the right side to pull the screen window into the viewer's space in order to avoid the screen window violation. The mask of the colour depends on the physical border of the display, for a feature film the masks are black but to make a floating window for this document that has a white background the masked area is filled with white.

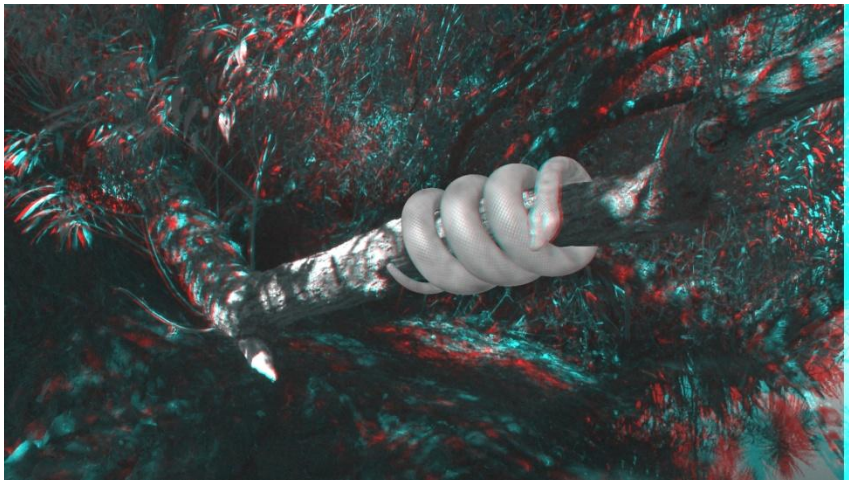


Image with floating window to pull the virtual window to avoid window violation.

Colour Grading

Colour grading for 3-D is a very complex task if it is going to be projected in a cinema due to different projection and display systems. Some recent stereoscopic film releases have included up to 14 different digital packages for single title (Mendiburu, 2009). The colour grading for this project was done for an HD TV using techniques that are similar to colour grading of standard 2-D footage.

Conclusion

The outcome of this research is a suggested workflow that can be effectively used by studios and individual professionals to integrate CG elements into live stereoscopic footage. The main finding of this research is that in a stereoscopic visual effects shot, the correct depth placement of the CG element is the most salient aspect in regard to producing a shot in which CG integration is believable.

During this project it was also found that the acquisition of good quality live-action stereoscopic footage is a crucial and difficult aspect in producing a stereoscopic visual effects piece. There are many factors that can make this difficult, such as the unavailability of commercial camera rigs tailored for stereoscopic production.

The viewer's distance to the screen and final viewing screen size introduce constraints that necessitate the careful planning of every shot. The depth of the CG element has to be carefully matched in relation to the live footage, in order to avoid visual and perceptual discomfort. These constraints enforce that to be able to produce a good stereoscopic visual effects shot one has to understand the basic physiology of binocular vision.

Future Work

This workflow needs to be further tested with different kinds of shots and it also needs to be tested in production studios. This research has answered quite a few questions relating to the workflow required to integrate CG elements into live-action stereoscopic shots, but at the same time it has raised many more questions that need to be addressed in future research.

For instance, all the existing composition rules like golden ratio, rule of thirds, etc., are made for 2-D composition, how can these rules be extended into the third dimension? Stereoscopic 3-D is currently being used predominantly in movies, but how will it transfer across to different mediums, such as music videos? Fast cuts that are synchronized to the beat of the music is common in music videos, but fast cuts in stereoscopic 3-D are avoided because it can cause visual discomfort due to quick and frequent changes in vergence accommodation relationship. The element of depth that is introduced by stereoscopic 3-D needs to be applied to all the different visual mediums and tested in order to create a new visual grammar for stereoscopic 3-D.

References

Duwaer A and van den Brink G. (1981).What is diplopia threshold? Perception and Psychophysics 29, 295-309.

Battiato, S., Curti, S., LaCascia, M., Tortora, M., & Scordato, E. (2004). Depth-map generation by image classification. Spie 5302, 95-104.

Chaurasia, B. D., & Mathur, B. B. L. (1976). Eyedness. Cells Tissues Organs, 96(2), 301-305.

Corporation, S. (1997). Background on Creating Images for CrystalEyes and SimulEyes. The Stereographics Developer's Handbook.

Dobbert, T. (2005). Matchmoving: The Invisible Art of Camera Tracking: Sybex.

Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. In A. J. Woods, J. O. Merritt, S. A. Benton & M. T. Bolas (Eds.), Stereoscopic Displays and Virtual Reality Systems Xi (Vol. 5291, pp. 36-46). Bellingham: Spie-Int Soc Optical Engineering.

Ehrenstein, W. H., Arnold-Schulz-Gahmen, B. E., & Jaschinski, W. (2005). Eye preference within the context of binocular functions. Graefe's Archive for Clinical and Experimental Ophthalmology, 243(9), 926-932.

Fehn, C. (2004). Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV. Paper presented at the Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI.

Graham, P. A. (1974). Epidemiology of Strabismus. British Journal of Opthalmology, 58(3), 224-231.

Hiruma, N., & Fukuda, T. (1993). Accommodation response to binocular stereoscopic TV images and their viewing conditions. [Journal]. SMTPE, 102, 1137-1144.

Kleinstein, R. N. (1984). Vision Disorders in Public Health. 5, 369-384.

Lambooij, M., Ijsselsteijn, W., Fortuin, M., & Heyderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. Journal of Imaging science and Technology, 53(3).

Lambooij, M. T. M., Ijsselsteijn, W. A., & Heynderickx, I. (2009). Visual Discomfort and Visual fatigue of Stereoscopic Displays: A Review. [Journal]. 53(3).

Lipton, L. (1982). Foundation of the Stereoscopic Cinema: Van Nostrand Reinhold Company Inc.

Mark, W. R., McMillan, L., & Bishop, G. (1997). Post-rendering 3D warping. Paper presented at the Proceedings of the 1997 symposium on Interactive 3D graphics.

Meesters, L. M. J., Ijsselsteijn, W. A., & Seuntiens, P. J. H. (2004). A survey of perceptual evaluations and requirements of three-dimensional TV. Circuits and Systems for Video Technology, IEEE Transactions on, 14(3), 381-391.

Mendiburu, B. (2009). 3D Movie Making.

Noorden, G. K. V. (1990). Binocular Vision and Ocular Motility: Theory and Management of Strabismus (4 ed.): Mosby-Year Book.

Ostrin, L. A., & Glasser, A. (2004). Accommodation measurements in a prepresbyopic and presbyopic population. [doi: DOI: 10.1016/j.jcrs.2003.12.045]. Journal of Cataract & Refractive Surgery, 30(7), 1435-1444.

Palmer, S. (1999). Vision Science: Photons to Phenomenology. Cambridge: MIT Press.

Pastoor, S. (1995). Human factors of 3D imaging; Results of recent research at Heinrich-Hertz-Institut Berlin. ASIA Display'95 Conference, 14(3), 66-72.

Redert, A., Berretty, R.-P., Varekamp, C., Geest, B. v., Bruijns, J., Braspenninq, R., et al. (2007). challenges in 3DTV image processing. SPIE6508.

Ronfard, R., & Taubin, G. (2010). Image and Geometry processing for 3-D Cinematography(Geometry and Computing) (1 ed.): Springer.

Rushton, S. K., & Riddell, P. M. (1999). Developing visual systems and exposure to virtual reality and stereo displays: Some concerns and speculations about the demands on accommodation and vergence. Applied Ergonomics, 30(1), 69-78.

Schreer, O., Kauff, P., & Sikora, T. (2005). 3D Videocommunication: Algorithms, concepts and real-time systems in human centred communication: Wiley.

Sexton, I., & Surman, P. (1999). Stereoscopic and autostereoscopic display systems. Signal Processing Magazine, IEEE, 16(3), 85-99.

Shade, J., Gortler, S., He, L.-w., & Szeliski, R. (1998). Layered depth images. Paper presented at the Proceedings of the 25th annual conference on Computer graphics and interactive techniques.

Steinman, S., Steinman, B., & Garzia, R. (2000). Foundations of Binocular vision: A Clinical Perspective (1 ed., pp. 153-170): McGraw-Hill Medical.

Stereographics Corporation-Background on Creating Images for CrystalEyes and SimulEyes. (1997). The Stereographics Developer's Handbook.

TARN, #160, James, W., VAZQUEZ, Carlos, SPERANZA, et al. (2007). Surrogate depth maps for stereoscopic imaging : Different edge types (Vol. 6490). Bellingham, WA, ETATS-UNIS: Society of Photo-Optical Instrumentation Engineers.

Tsukuda, S., & Murai, Y. (1988). A case report of Manifest Esotropia After Viewing Anaglyph Stereoscopic Movie. Japanese Orthoptic Journal, 18, 69-72.

Ukai, K. (2006). Human factors for stereoscopic images, Toronto, ON.

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. [doi: DOI: 10.1016/0042-6989(95)00018-U]. Vision Research, 35(19), 2731-2736.

Woods, A., Docherty, T., & Koch, R. (1993). Image distortions in stereoscopic video systems. Paper presented at the Proceedings of SPIE: Stereoscopic Dispalys and Applications IV.

Woods, A. j., & Rourke, T. (2004). Ghosting in Anaglyphic Stereoscopic Images. Spie 5291, 354-365.

Yano, S., Emoto, M., & Mitsuhashi, T. (2004). Two factors in visual fatigue caused by stereoscopic HDTV images. [doi: DOI: 10.1016/j.displa.2004.09.002]. Displays, 25(4), 141-150.

Zone, R. (2007). Steereoscopic Cinema and the Origins of the 3-D Film, 1838-1952: The Unoiversity Press of Kentucky.