

Creation of a Complete Stereoscopic 3D Workflow for SoFA

Allison Hettinger and Ian Krassner

1. Introduction

1.1 3D Trend

Stereoscopic 3D motion pictures have recently risen to popularity once again following the success of films such as James Cameron's *Avatar*. More and more films are being converted to 3D but few films are being shot in 3D. Current available technology and knowledge of that technology (along with cost) is preventing most films from being shot in 3D. Shooting in 3D is an advantage because two slightly different images are produced that mimic the two images the eyes see in normal vision. Many take the cheaper route of shooting in 2D and converting to 3D. This results in a 3D image, but usually nowhere near the quality as if the film was originally shot in 3D. This is because a computer has to create the second image, which can result in errors. It is also important to note that a 3D image does not necessarily mean a stereo image. 3D can be used to describe images that have an appearance of depth, such as 3D animations. Stereo images refer to images that make use of retinal disparity to create the illusion of objects going out of and into the screen plane.

Stereo images are optical illusions that make use of several cues that the brain uses to perceive a scene. Examples of monocular cues are relative size and position, texture gradient, perspective and occlusion. These cues help us determine the relative depth positions of objects in an image. Binocular cues such as retinal disparity and convergence are what give the illusion of depth. When two slightly different images are presented to each eye, the brain combines these images using the cues above. When creating stereo images, it is important to make sure that there are no conflicting cues because this is where eyestrain is introduced. Long term exposure to problems like this will cause the headaches that many audiences complain of. One of the problems with 3D films today is that there are not enough filmmakers and technicians educated about these conflicting cues so the final film is headache inducing to the audiences.

On going research and new developments in 3D technologies are always trying to improve the quality of 3D films and part of our project was to look into what makes "good" and "bad" 3D films, in addition to creating a useable 3D workflow for RIT.

1.2 3D for SoFA

Currently, RIT's School of Film and Animation is not fully equipped to produce 3D films. The goal of this project was to create an entire workflow that can be easily utilized by SoFA students. A camera rig was built to be used with two Canon 5D Mark II cameras, which are readily available to SoFA students. A workflow then had to be created, from image capture to exhibition, that incorporated as many components as possible of the current workflow most SoFA students are accustomed to. Qualitative and quantitative test scenes were shot, which after analysis, provided helpful insight into what to avoid to produce quality stereo images. The next step was to educate students on how to use the rig along with how to properly shoot 3D films by using the results from the experiments. In the end, this project that a 3D workflow could be built for SoFA and be able to be utilized by future students who (with a little training) are interested in producing quality stereoscopic 3D films.

2. Background

2.1 History of 3D Technologies

Stereoscopic 3D images have been around since the mid 1800s. A paper was published by Charles Wheatstone in 1838 about how the brain can fuse two images together and shortly after, the first stereographs were made¹. Stereographs were two slightly different images placed side by side and were viewed through a stereoscope, invented in the 1850s, to produce a three dimensional effect. Stereographs were very popular, especially because they were cheap enough that most could afford them.



Figure 1. Example of a stereograph and a stereogram¹.

The first patent combining stereograms and motion was filed by Jules Duboscq in 1852². Duboscq put twelve image pairs onto a vertical rotating disk and he called it the Bioscope. Unfortunately, the Bioscope was not a success because there was not easy way to reproduce the disks needed. Throughout the late 1800s, many tried to create motion stereoscopic movies in a similar style to Duboscqu's disk, most involving animations instead of photographs. Even the Kinetoscope patent mentioned an attachment for stereograms².

The first known use of an anaglyph type system is attributed to Charles d'Almeida in 1858². Anaglyph takes advantage of complementary colors to send a different image to each eye. Charles d'Almeida used anaglyph to project two images using red and green filters while the audience wore red and green filtered glasses. Louis Ducos du Hauron furthered d'Almeida's work to fully understand how to use color filters in combination with colored images to send two different images to each eye.

The first feature length film to be commercially shown in stereoscopic 3D was *The Power of Love* in 1922. Two synchronized projectors were used and the two projections were superimposed. The audience wore red and green filtered glasses. To capture the images, one lens had a green filter over it and the other a red filter. When processed, the red film was tinted red and the green film was tinted green. This eliminated the need for filters over the projector lenses.

Anaglyph became the popular method to exhibit 3D films because of the ease of producing and projecting the films along with the ease of producing anaglyph glasses. MGM even released an anaglyph short, *Audioscopiks*, in 1936². This short was extremely successful and even won an Academy Award in 1936 for Best Short Subject, Novelty. MGM manufactured 3 million pairs of red/green anaglyph glasses for this film so that it was able to be widely released. MGM also worked with Technicolor to produce a two-color process to produce multiple anaglyph prints for distribution.

Also in 1922, Laruens Hammond introduced Teleview, an alternate way to view 3D films³. Two projectors alternately projected left and right eye images. The audience would look

through a viewer that had a rotating shutter that was in sync with the projectors so that the left eye saw and left image and the right eye saw the right image. The viewers were attached to the seats, which made this system not easy to move to other theaters. The Televue system was put to the test in New York City with a showing of *The Man from M.A.R.S* in 1922. Unfortunately, the showing did not draw a large audience.



Figure 2. Illustration of an audience using the Televue.

The idea to use polarized light as a way to exhibit stereoscopic films wasn't new in the 1900s but Edwin Land was the first to build synthetic polarizers, which would later be used to show 3D films. Land and his business partner, George Wheelwright III formed the Polaroid Corporation and experimented with using polarizers for 3D projection of films. By polarizing the light coming from the two projectors such that the two projectors are polarized 90 degrees apart, the audience can then wear glasses that are similarly polarized so that different images are seen by each eye. Land and Wheelwright held demonstrations in 1936 in New York City and opened an exhibit in the New York Museum of Science and Industry that showcased polarized stereoscopic films². The use of polarizers instead of colored filters meant that full color films could now be projected. Previously, anaglyph films were colored to work with the anaglyph glasses, but they were black and white films (also, color was only beginning to be introduced to the industry).

As 3D films became more and more popular, a way was needed to easily and reliably capture the images. Before the 1930s, two projectors were placed side by side which resulted in two separate film strips that needed to be in sync for projection. In 1935, Otto Vierling developed several methods for image capture. Using a single strip of 35mm film and a beam splitter camera, two images would be recorded on the film, slide by slide. There were different arrangements that included side-by-side vertical images and side by side horizontal images, with one flipped 90 degrees.



Figure 3. Example of Vierling's single strip format²

Stereoscopic 3D films experienced a surge in popularity in the 1950s sparked by the 1952 release of *Bwana Devil*, one of the top grossing films of the time⁴. Studios rushed to create 3D films and titles such as *House of Wax* and *Dial M for Murder* had 3D releases. All these films were projected using polarized filters, not anaglyph filters. By the mid 1950s the 3D trend died down, due to issues such as cost of equipment and difficulty of perfectly syncing the images.

3D films did not enter the mainstream film industry again until the 1980s. This time, a spaghetti Western called *Comin' at Ya!* reignited the 3D trend in 1981⁴. The years between the 1950s and 80s allowed time for stereoscopic capture and projection technologies to become more reliable and produce higher quality films. StereoVision, developed by Future Dimension, was a single strip over-under film format that became widely used to shoot 3D movies. Films such as *Friday the 13th Part III* and *Jaws 3D* brought in a lot of money and sold out numerous showings. An interesting trend from the first boom in the 1950s that continued into the 1980s was that most films made for 3D were horror films, taking the suspense from characters jumping out at the audience to a new level. Most of these films were also projected using polarizers, only a few were in anaglyph. Similar to the first surge of popularity in the 1950s, popularity of 3D films in the 1980s had died out by the end of the decade.

But once again, 3D made a comeback in the mid 2000's, with films such as *The Polar Express*, released in 2004. This time around, most films were either digitally animated to be 3D or digitally converted from 2D to 3D. The ability to digitize films is a major cause of the recent resurgence of 3D films. It's costly to shoot films in 3D but it's easy to convert them digitally to 3D. The films are shot in 2D and the second "eye" is created using a computer algorithm. This process requires the computer to recreate parts of the scene to get a slightly different perspective from the original scene captured by the camera. Because the scene has to be recreated, there is error introduced that a human has to manually fix. James Cameron's *Avatar* also helped to jump-start the 3D industry. Unlike most other 3D films in the 2000s, *Avatar* was shot in 3D, which led to a higher quality film than what most audiences were used to seeing.

In addition to viewing stereoscopic 3D films in theaters, 3D has begun to spread to consumer's homes as well. Companies such as Samsung and Sony have developed consumer level 3D televisions. The challenge to taking 3D outside of the movie theater is that the image is no longer projected but emitted from a display. Bringing 3D into the home means that new standards have to be created, similar to when HD programming was introduced. Aside from delivering stereo content through television programming, 3D Blu Ray movies have been released which allows for viewing of stereo content, even if you do not have access to 3D television channels. Currently, both active and passive systems exist on 3D TVs. Active shutter glasses are very expensive so passive systems are gaining popularity. There is not much stereo content available for 3D TVs, only a few channels broadcast in stereo and not many movies are released in 3D Blu Ray. This technology is still relatively new so more and more content is becoming available and this technology is still being improved upon to make it more accessible to the public.

There are many other ways stereoscopic images are being adapted to today's technology, such as 3D gaming systems and 3D screens that don't require glasses. Using either a parallax barrier or a lenticular lens, each eye sees a different image. For most autostereoscopic displays at the moment, the viewer must be in a specific position relative to the screen to see the stereo image. This technology has already entered the consumer market with the Nintendo 3DS.

There are also improvements that can still be made to 3D films shown in theaters. Research is always on going to try to find the next best technology and to lower expenses associated with 3D exhibition.

2.2 Human Vision and 3D

2.2.1 Perceiving Depth

The human visual system relies on two eyes to collect visual information. Stereo vision refers to the way a binocular visual system captures two slightly offset images to then better determine relationships between objects and create a sense of space for the viewer. The employment of two eyes in addition to depth cues such as perspective play a major role in our vision and is the first step in our development of depth perception. Of course simply capturing two images does not automatically yield perception of space, much processing of the different depth cues in the scene must be preformed further back in the visual system. With so many varying components in each distinct visual system it should be clear that no two people see the world exactly the same as one another. However, these slight differences among the population are expected and do not cause a great deal of problems when the visual system is fully developed and working correctly.

Single lensed, or one eye, vision systems are complicated enough on their own, needing to determine changing points of focus, varying illuminations, and all the other factors present in an environment as mercurial as the natural world, adding a second lens only complicates things further. After all a single image such as one achieved with monocular vision does offer clues as to the depth of a scene. Occlusion or the ordering of objects and how they block one another is perhaps the strongest visual cue when an individual is trying to determine the relative depth of an object (Figure 4). Using other parts of our brain and tapping into knowledge we have accumulated, characteristics such as the size of an object relative to other objects can dictate where it is in space. Knowing the general size of a car, our visual system can understand that a car in the distance which appears smaller than one in front of us is farther away, not just smaller and equally close to us. The texture gradient and relative height of an object can also help to properly establish depth as patterns that continue away from the viewer become compressed and eventually unable to be clearly distinguished. Familiarity of the size of an object also helps to determine placement in space as does linear perspective and the knowledge that the geometry of the world is Euclidean in form. Lines in our field will converge in the distance at a single vanishing point. Haze in a scene or aerial perspective is another monocular cue that our visual system recognizes as proof that an object is recessed far enough for the atmospheric haze to be noticeable. Motion parallax also denotes depth as objects closer to the viewer move quicker then objects in the distance when the observer is moving. This is clear in the way that trees in the distance barely seem to move when looking out the window of a car yet the lines on the road fly past the viewer. These are all cues we get from simply having only one angle of an image such as in a photograph. The addition of a second lens seems almost unnecessary since we can very accurately determine depth from monocular cues alone.

When two eyes are available both eyes must work together as the brain must now fuse these two images into one. This can be a difficult process. However, while this second eye may be problematic, it also greatly enhances the capabilities of the visual system allowing for the two

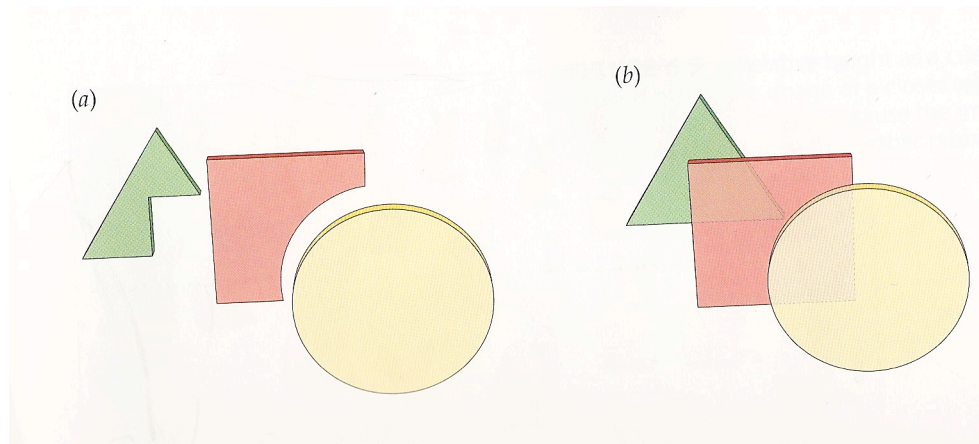


Figure 4. Occlusion and its potential failure⁵

images captured to be compared and differences to be used to better and more correctly determine depth. It also allows the field of view of an individual to widen to encompass the entire hemisphere in front of the two eyes, assuming the base of the hemisphere runs through the center of both eyes. The combined two eyes create a field of view actually slightly greater than a hemisphere, about 190° in front and to the side of the viewer. Both eyes together cover 110° of that field. For example, using the monocular cue occlusion, our eyes may be tricked to believe an object is in front of another when in reality the object that appears farther is at the same distance from the viewer (Figure 1). Relative size, familiar size and relative height can all also be misinterpreted if an object is simply made different from its counterparts or is different from what we are familiar with. This is where a binocular system has the advantage and cues such as motion parallax begin to better define our surrounding environment. By moving around an object we can better see the scene and more accurately determine the depth of the scene. Even without motion, the slight disparities between the left and right eye images help to remove any confusion and see another angle just offset enough to understand relationships of the objects in space. This allows us to make quick decisions to avoid bumping into objects and plan our paths of movement instantly without even realizing we are consciously making judgments about our environment. Our visual system also takes advantage of oculomotor information. As our eyes move in our head, they converge onto the point of interest that we are looking at. This angling of the eyes toward one point is done by the ocular muscles of the eye. Information about the depth of an object is also obtained from the information about the degree of the angle of convergence. An object closer to the viewer will cause a greater angling of the eyes.

Referring to eyes as lenses leaves quite a bit out of the picture. The eye consists of the cornea, which first acts to focus the light which then enters the pupil, which can control how much light enters the eye. The light then meets the lens, where muscles determine the focusing power to focus the light onto the retina, a process known as accommodation. The retina is composed of a mosaic pattern of photoreceptors known as rods and cones. These photoreceptors determine the wavelength of the light by their own spectral sensitivity and thus the color of the incoming light. The rods and cones then send messages through the optic nerve to the brain for further processing.

In order to properly capture the environment, both eyes must work together. As the retinal images are compared with one another to determine depth, both eyes must be focused at the same point in the scene. This means that the eye lens must be adjusted for focus and the eyes must be angled towards the point of interest. The extraocular muscles work to properly angle the

eyes so that they converge upon the point of focus. The focus can then be adjusted by changing the thickness of the eye lens using the ciliary muscles. This of course is done automatically in a properly working visual system and the viewer does not even know the change is occurring. These muscles are crucial since without first aiming the eye at the correct place, how can the brain hope to properly process the images to allow for stereo vision.

Once properly focused on the retina, the image is then sent to the brain via the optic nerve. Its first stop is at the lateral geniculate nucleus (LGN). The way that the information is transmitted is crucial in understanding how it is processed and where along the pathway problems can occur. It is in the LGN that the visual field is split and the left half of the visual field is sent to the right LGN and the right half of the visual field is sent to the left LGN⁵ (Fig. 5). The two sides are first processed separately and then recombined to the same side of the LGN and visual cortex to allow binocular columns and comparator cells to work together to extract even greater information about the scene. There exist many cells that are responsive only to stimulation in both eyes at the same time. In the visual cortex at more complex processing levels, binocular cells work only when they are stimulated by imagery in designated areas of the right and left eye retinas. Since an object only exists in the center of both retinas when the main object is focused on, disparities between other parts of the scene and their placement on the different retinas is useful information. Specific cells and pathways are developed to only be sensitive to these specific binocular pairs as they appear on the retina. This was determined when testing done with cats found cells previously unstimulated that responded only to binocular retinal stimulation⁶. The optic chiasm is where stereo vision is really begun as it is the point where the two signals are crossed and binocular fibers and their information begins to be determined.

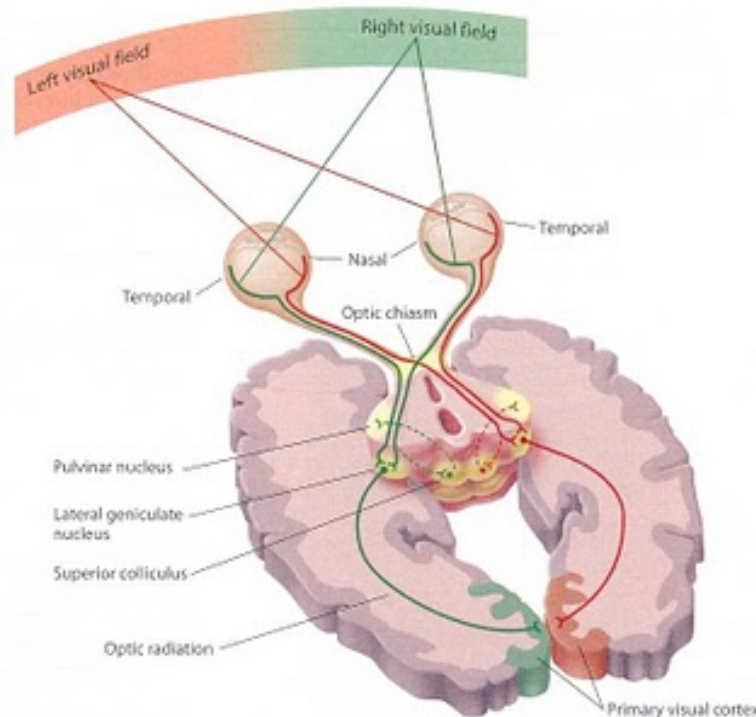


Figure 5. LGN Processing Pathway⁵

The information of the scene is also contained in terms of how the field was imaged onto the retina. This information passes to the primary visual cortex where it will be further processed in terms of spatial frequency, orientation motion, direction, speed or temporal frequency and other elements of vision. The amount of area designated for processing is directly related to the importance of the information within the visual field, a process known as cortical magnification. The center of the visual field is where the highest focus is placed and the most detail captured. This acuity fades off as we get further off angle towards the periphery of the field. For this reason, the visual cortex takes the information from the LGN, which retains location information, and gives the center of the field of view the greatest area and thus most processing power and gives less area to the periphery. This is the path that images follow when light reflects off of object allowing us to see what is there. Each side of the visual field is processed on the opposite side of the brain meaning the left part of the visual field is processed by the right side of the brain. However, this does not mean the entire field captured by each eye is “reflected” to the opposite side for processing, instead the left half of both the left and right eyes are sent together and the right half of both eyes are sent together. At any point in this chain issues can arise but since the images from the retina onward are mixed between the two eyes and are designated by sides of the field of view, it seems unlikely that damage to one part of the brain would affect only one eye. In this case at least, it seems damage would cause trouble for one half the visual field, but not one eye in specific. The images are compared and processed together through the use of the corpus callosum, which connects the two hemispheres of the brain and allows for disparities between the two now processed images to be analyzed and depth perception created⁵.

2.2.2 Human Vision and 3D Films

The concept of shooting a movie for 3D projection in theaters or at home is modeled after the human visual system. The theory behind 3D movies is that if one can discretely capture the two offset images in the same manner as the eyes would, then by showing the left eye one image and the right eye the other, the brain will process the two and create the perception of depth. As long as the images presented to both eyes are correctly captured, the brain should be able to process the aforementioned cues in the scene and the brain is tricked into perceiving depth from a scene that is projected flat. While this may seem simple enough, there are many factors that must be considered when using cameras to capture stereo images as opposed to our eyes.

The brain is much more intelligent than a camera and will consistently adapt to maintain correct stereo vision. As we look around a scene, our focus changes to different objects and people. We choose to ignore certain changes and make note of others, all in the attempt to maintain correct vision. A camera, on the other hand, cannot adapt like the brain can and this is what can cause 3D movies to fail and become painful to watch. A large problem present in 3D movies is that different depth cues may conflict and force the eyes and brain to work unnaturally. Invariably, a camera cannot capture as much information as our eyes can, nor are we given a full world to look around in.

The directors tells the audience where to look by setting the focus and convergence onto a specific area or object. While looking at this specific object, the audience should experience a situation similar to viewing that object in the real world. However, if the viewer decides to look away from the object, problems can arise. In a normal situation, if the viewer looks to the background, the eyes will adjust and now the background will be in focus and the main object

may go out of focus. In a 3D film, the background remains out of focus, even if a viewer looks directly at it (unless the director decided to use a large depth of field, where everything is in focus). Also, since the convergence was set at the object, not the background, the viewer is forced to try to fuse two disparate images. In a normal real world situation, the eyes constantly converge on the object they are looking at.

An example of this situation would be to look focus on your finger. The background behind your finger is out of focus and is actually a double image. As soon as you bring your focus to the background, the background becomes a focused, single image while your finger becomes a blurry, double image. In 3D films, the focus and convergence of the scene remain constant, regardless of where the audience looks.

When shooting and projecting in 3D there are three different areas of depth that exist: the space in front of the screen, which is the negative depth, the space behind the screen, which is the positive depth, and the space at the screen plane, which is the zero point. The offset between specific objects at specific depths in the two images is what establishes the depth perceived for those objects in the scene. This is similar to how our eyes work, as objects that are farther away from our point of convergence and focus have a greater distance between them on our two retina. Our visual system is set to force these objects out of focus, which helps us to disregard any double vision and pay attention only to what is in focus.

2.2.3 Stereo-Blindness and 3D Imaging

Having noted the key elements and structures used in creating a sense of depth through perception, it becomes clear how common ailments can inhibit the creation of stereopsis and result in an individual being considered stereo blind. Of course the most obvious cause of stereo blindness is complete blindness in one eye.

When vision is present in both eyes, stereo blindness may be caused by other defects. Amblyopia is a condition where the individual cannot fuse the two images from the separate eyes due to an overbearing disparity between the two images. Common causes of amblyopia can result from physical defects such as a weakness in the extraocular muscles in which the eyes are not aligned correctly. Therefore one eye seems to drift, forcing it to capture an image too dissimilar to the other eye. This ailment known as strabismus can cause stereo blindness, as the brain in these scenarios makes the decision to only process and proceed with information from the stronger eye. This is most commonly onset in childhood because from birth to about two years old, the extreme neuro-plasticity of the brain suggests it adapts to suppress the “incorrect” image. As the child grows, studies suggest that after age seven the individual will have a nearly impossible chance of regaining stereo vision. The suppression of the image causes one part of the brain’s visual processing to be underdeveloped, which is why even though the fields of view from both eyes are mixed, selected areas of the cortex pertaining to only one eye can be underdeveloped.

The eye does not have to be “lazy” for stereo blindness to occur, such as with anisometropic and occlusion amblyopia. In these cases the refractive indexes of the lenses do not match causing a blurring in one of the eyes or the substance filling the eye is too opaque and light has difficulty passing through to the retina. In either scenario, the stronger clearer image is preferentially chosen and the other image suppressed. If untreated, the disparity causes one eye’s processing area within the LGN and cortex to be underdeveloped. This can even occur before birth, at the prenatal stage. Monocular pathways begin to develop within the womb as spontaneous retinal activity, which is present in a growing fetus, causes pathways to be fortified

and developed. When this process is uneven in the developing fetus, certain pathways and ocular dominance columns from one eye will become stronger and more dominant. This creation of an ocular dominance column can lead to one eye being weaker, not in terms of capture as with the previous amblyopia, but at the processing level. Even perfectly captured stereo images cannot be fused.

To correct for disparities between the two eyes in terms of focus, corrective lenses can be worn. In the case of strabismus, the extra-ocular muscles can be surgically adjusted to align both eyes. For all of these cases except for occlusion amblyopia, especially in children, the dominant eye can be covered to force development of the weaker area of the brain. Adults have less success with this, and when amblyopia sets in at a later stage in life, instead of suppressing the image the individual may experience double vision since the brain is already developed and the transduction pathways are already fully developed. Here the only fix is to wear glasses.

In terms of the film industry, all leading 3D technologies rely on stereopsis to create illusions of depth. Using polarizers or alternating shutters such as in active glasses, different images are displayed to the two eyes just as they would be in a normal visual situation. Technologies accommodating stereo blindness today attempt to either remove the stereopsis component from the process or attempt to create it using only one eye for stereo-blind clients. Methods using only one eye for stereopsis attempt to shutter the images showing the different images at an offset to each eye. In this method the dominant eye can be shown the image designated for it and then the image for the opposite eye can be shown to that same dominant eye. With this method, the glasses must be calibrated to react differently than normal to show the dominant eye the two different images at an offset long enough not to cause blurring by flashing different images too quickly.

While these are two options to help bring 3D to stereo blind clients, the majority if not all manufacturers today seem less than concerned with the loss of this section of the population. Their response to questions is that 3D TVs work just as well in 2D with the 3D turned off and stereo blind clients can watch that way. The only true solution comes from attempts to make 3D televisions without glasses. Using as many as 20 or 30 cameras to capture a scene at all different angles and then projecting the corresponding camera's capture to a viewer at that given angle can create a 3D view into a space. This method captures all conceivable angles so that a viewer standing at offset angles from the center can be shown the image the camera saw of the scene at that angle. This method allows viewers to move around and perceive depth on a flat screen. However, the extreme control of the light can only be achieved using a highly diffuse screen to allow blending of the angles for the viewer. This is a highly experimental method and nowhere near commercial quality, but it is one of the methods available that allow stereo blind clients to experience 3D from a flat screen.

2.2.4 Visual Discomfort

Visual discomfort in stereo images arises when the image was not properly created. Conflicting depth cues along with problems such as excessive vergence forces the eyes to perform unnaturally and can lead to headaches and eye strain. Basic depth cues such as the visual cues previously mentioned help the viewer determine the depth of the scene. When these cues do not work together, the resulting image can be confusing to look at. Cues such as perspective and occlusion all work together to help the brain determine the depth objects are at. When viewing a stereoscopic 3D image, all the monocular cues that are used in 2D films are still present. Retinal disparity is introduced in 3D films to create the perception of depth. Sometimes, the depth

determined from the retinal disparity does not agree with the depth determined by monocular cues.

A major depth cue is the screen window. Our brain recognizes that the screen is at a physical distance and the stereo images can be in front, behind or at the screen plane. Knowledge of where the screen plane is affects how a stereo image should be converged. If an object with a perceived depth is in front of the screen, it should not intersect the edge of the screen. If part of an object bisects the edge of the screen, then monocular visual cues tell us the object's depth should be behind the screen. However, the convergence of our eyes along with accommodation tells us the object should be in front of the screen. This is called breaking the window and causes confusion, which leads to visual fatigue. Most educated stereographers know to look for and avoid these problems but they still do pop up in hastily created stereo images.

One of the biggest problems when viewing 3D films is that accommodation and convergence of the eye are forced to act independently. Accommodation is when the lens in the eye adjusts to focus on an object and vergence is when the eyes converge on the object in focus. They usually act together when viewing a real scene but 3D films can force these two actions to be separate, which causes eye strain. When a viewer is watching a 3D film, the eye lens focuses on the screen plane, which is constant. However, the eyes converge to the action in the scene, which can may or may not be at the screen plane as its depth can vary. If the perceived depth of the action is not at the screen plane, then accommodation and convergence are forced apart. This is why setting the convergence of the main action of the scene at the screen plane will make viewing 3D films much more comfortable.

Panum's fusion area relates to the mismatch between vergence and accommodation⁷. If vergence and accommodation are separated too much, the two retinal images won't fuse together. This results in a double image, or diplopia. Instead of diplopia, there may be binocular rivalry, where instead of double vision, the viewer sees the image from one eye for a moment and then the image from the other eye is seen for a moment. What the viewer sees will continue to alternate. Both of these results will cause the viewer some eye strain and confusion when looking at a stereoscopic 3D image. Figure 6 shows Panum's fusion zone which is defined by the inner and outer limits of single binocular vision.

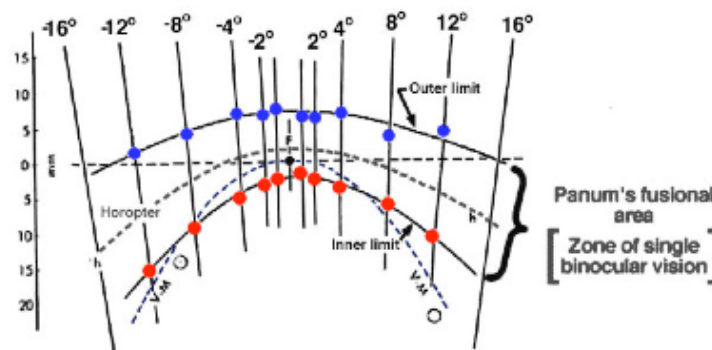


Figure 6. Panum's Fusional Area

Excessive divergence and convergence is one problem that limits the fusion of two images. A scene that contains too much divergence is usually a scene that has a large depth of field with a convergence point that is near the front of the scene. Objects in the background begin to diverge so much that if the viewer were to look at them, the viewer's eyes would have to physically diverge to fuse the image. Similarly, if a scene has too much convergence, the viewer

may be forced to cross their eyes to fuse the image to a point that it becomes painful. This may occur in a scene where the convergence point is set far back so that objects closest to the camera have a large convergence and the viewer wouldn't be able to fuse the images unless they crossed their eyes an excessive amount.

When looking at a scene with excessive vergence, some viewers will try to unnaturally diverge or converge their eyes to fuse the image, but doing so is painful. So instead they don't look directly at the object, which results in the retinal images being outside of Panum's fusion area. This produces double vision or binocular rivalry, both of which can be very distracting. Also, excessive vergence leads to accommodation-vergence conflicts. If the perceived depth of an image is too far in front or behind the screen, eye strain is introduced for reasons already mentioned.

Percival's zone of comfort suggests vergence and accommodation limits for comfortable viewing of a stereo image⁷. Figure 4 shows Percival's zone of comfort along with an example screen plane, which is the horizontal line running through the graph. The focal distance is constant because the viewer is focused on the screen. Accommodation is the variable that needs to be limited as shown below. The accommodation range that is within Percival's range is about +/- 0.5 to 1D from the normal convergence/accommodation pairing at screen distance.

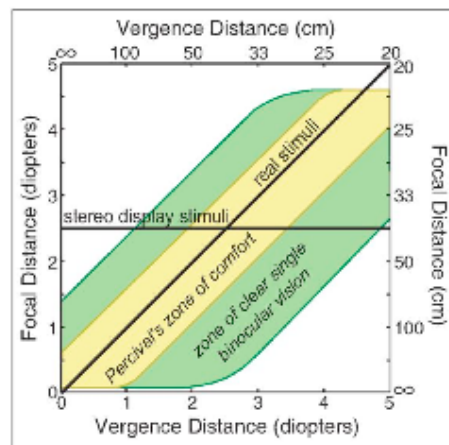


Figure 7. Percival's Comfort Zone.

In a recent series of experiments, Hoffman et al.⁸ looked at the affects of images with a vergence-focal conflict. They presented their viewers images that forced vergence and accommodation to act separately and only 6 out of 11 viewers could fuse the images without significant discomfort. The viewers who couldn't fuse the images had proper eyesight and could perceive depth in 3D stereo images, just not when the image was outside of Percival's comfort zone. The fact that nearly half of the viewers had a problem with images with mismatched vergence and accommodation shows that 3D filmmakers need to be careful about how much depth they put into a scene so as not to cause eye strain to half or more of their audience.

In an additional study⁸, Hoffman and others presented their viewers with images that had vergence-focal conflicts and asked questions such as "how tired are your eyes?" and "how tired and sore are your neck and back?" It was found that after 45 minutes of exposure to stereoscopic images that did and didn't have vergence-focal conflict, the observers reported mild to severe symptoms to the question "how tired are your eyes?" and mild symptoms for "how tired and sore are your neck and back?" and "how do your eyes feel?" In this experiment, the vergence had a

range of +/- 0.65 diopters, which is near the threshold of Percival's comfort zone. From both experiments, it can be concluded that many viewers will experience eye strain if an image has excessive convergence or divergence, which leads to a vergence-focal conflict.

Another problem viewers have when fusing images is that it takes time to fuse an image. An example of this is holding your finger in front of your eyes and quickly bringing it closer while trying to focus on your finger. Once you stop moving your finger, you'll see a double image of your finger for a second before the brain can fuse the image. The more excessive the disparity between the left and right eye images, the harder the images are to fuse and the longer it will take your eyes to fuse them. In 3D films, shots of an object quickly flying toward the audience are hard to pull off because the audience cannot fuse the image quick enough. Instead, they see a double image of the object and the stereo effect is lost.

Another problem that can lead to visual discomfort is ghosting. Ghosting occurs when one eye sees information meant for the other eye. In anaglyph and polarized systems, this is typically due to the filters in the glasses being imperfect. The result is that some objects appear to have a ghost, similar to the effect produced by double vision. To counteract ghosting, a technique known as ghost busting is used, which mathematically subtracts part of the left image from the right image (or vice versa). This helps to solve the problem of ghosting but not completely and it sometimes leaves behind artifacts. The best way to currently prevent ghosting is to be careful about having highlights in your scene. Areas with a high intensity are more likely to leak through a filter made to stop it.

2.3 Methods for Viewing Stereoscopic Images

There are many ways to exhibit stereo content so that a different image is delivered to each eye. This paper only looks at the use of polarizers and color anaglyph to deliver stereo content.

2.3.1 Polarized Glasses

Once the two images have been properly captured, the next step is to project them and ensure that right eye sees one image and the left sees the other. There are many different ways to achieve this but the method that maintains the best image quality and has become the industry standard is through the use of polarized glasses.

The light from the projector is polarized either linearly or circularly with the images for the two eyes set at opposite polarizations. The use of circular polarization has become the more prevalent method over linear as it is more resilient than linear polarization. With a linear polarization scheme, if a viewer tilts their head while wear polarized glasses, the glasses' polarization no longer matches the polarization of the light. This results in severe ghosting because a portion of both left and right eye images are able to make it through the tilted, polarized filter. With circular polarization, even if the viewer tilts his head, the illusion is better maintained, even as the orientation of the glasses changes. This is because circular polarizers add a wave plate to a linear polarizer, which shifts the phase of the linear polarized light and it becomes circular polarized light. To use polarizers to show stereo content, the projector will polarize the light from the left image one way and polarize light from the right image 90 degrees from the left image.

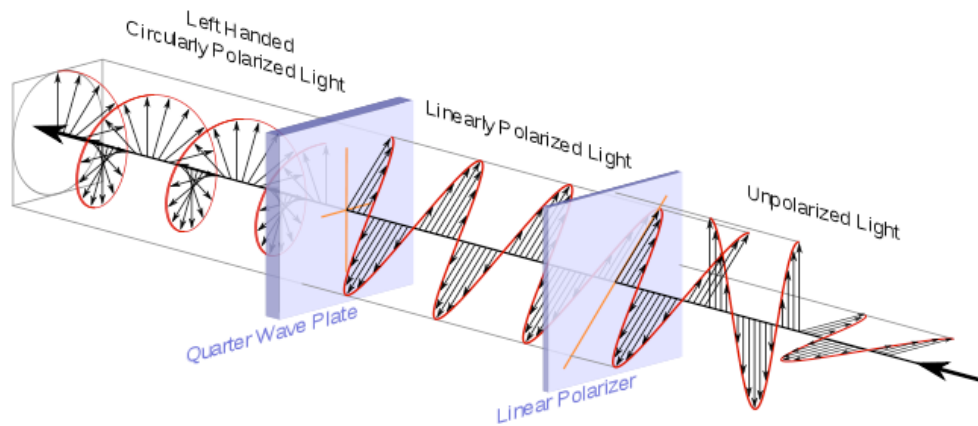


Figure 8. Example of how a circular polarizer affects unpolarized light.

The main problem that the use of polarization faces is the light loss that is innate to the system. All single projector stereo exhibition suffers from a 50% loss of all light instantly as part of the duty cycle which exists as at all times only one eye is receiving luminance information. The projector is forced to switch between the left and right eye images, which integrates to a loss of 50% of all luminance. The polarizing filter that is placed in front of the projection then reduces the light further so that only about 20% of the original light makes it to the screen. The light is bumped up via the gain of the screen but only by a factor of about 2 or a little more. Because a silver screen must be used for projection, a gain is introduced due to the metallic properties of the screen. The glasses the viewer must wear also have polarizers for lenses and reduce the light again, meaning the entire system is only about 36% efficient. The efficiency can be raised in a two projector system because the duty cycle no longer exists and you start with twice the luminance since you are using two projectors instead of one. But as both projectors still need their own polarizers, the final outcome means the system only ends up being about 43% efficient.

In these scenarios, the light's polarization must be maintained at all times which is why a white matte screen can not be used, as it is made to diffuse the light and remove polarization. Polarized stereo projection requires the use of a silver lenticular screen, which as its name suggests, often contains silver and works to maintain polarization of light. However, with this type of screen the inherent qualities of the silver and the maintained polarization do have some flaws. The metallic properties of the screen creates a gain which, means that as viewers move off angle from the center, the luminance drops off dramatically. It also means that the center of the screen where the projector is pointed will have a hit spot that is brighter than the periphery of the screen even when seated in the center of the theatre. The screen does offer an increased brightness with a gain factor of around 2.0 or above which helps to compensate for some of the light lost due to the polarization of the glasses and projectors.

The other option for those not interested in silver screens is the use of active glasses. These glasses are made up of LCDs that are activated by a sensor on the screen to turn on or off. The on state aligns the LCDs so as to allow light through and the off state reverses the LCDs so that no light is let through. The image on screen will alter between left and right eye images and the glasses will shutter accordingly. Through the use of an infrared sensor, the glasses coordinate their shuttering with the projector ensuring that as the projector switches between left and right

eye images, so do the glasses. Even without polarizers in front of the projector, the glasses' material have a reduced transmission so the shuttering combined with the light loss of the glasses leads to a system that is only 17% efficient. This reduced efficiency is also due in part to the use of a matte screen during projection.

2.3.2 Anaglyph

Creating stereo images using anaglyph is one of the simplest and cheapest ways to produce 3D images. The left and right images are differentiated by color; the left and right image colors are complementary colors. Filtered glasses are used to separate the two images so that the left eye sees the left image and the right eye sees the right image.

There are many ways to produce anaglyph images. McAlliser, Zhou and Sullivan outline a method of creating anaglyph images by converting RGB values into CIE L*a*b* color space⁹. In this space, they can then find pixel values for the two images so that the colors in the left image have the smallest possible Euclidean distance from the colors in the right image, when viewed through the red/cyan filters. This means that the colors the right and left eyes see when wearing anaglyph glasses are as similar as possible. Retinal rivalry happens when the two images (or colors) in each eye are so different that the brain cannot fuse them. This method tries to lessen the retinal rivalry by delivering more similar colors to each eye when looking through the filtered glasses. This method is computationally intensive and assumes knowledge of the display device and filters used.

A simpler approach to generating an anaglyph image is to stay in RGB color space and use simple matrix math. There are pros and cons to every method; converting to CIE color spaces typically give better results with better color reproduction, but involve complicated computations and knowledge that may not always be accessible to anyone wanting to view an anaglyph image.

Several methods of converting images to anaglyph using matrices have been outlined by Peter Wimmer¹⁰. The most basic way to create red/cyan color anaglyphs is to use the equation below, where the subscript 1 refers to the left image and subscript 2 refers to the right image.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} R1 \\ G1 \\ B1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R2 \\ G2 \\ B2 \end{pmatrix} = \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

This method simply takes the red channel from the original image to create the left image and the green and blue channels from the original image to create the right image. This method reproduces color relatively well but it leads to retinal rivalry. Retinal rivalry occurs when the two eyes see very different images. For example, in a red/cyan image, a cyan object will cause retinal rivalry. The left eye will see black because the red filter should theoretically block all light from a cyan object. To the right eye, the object will appear white, since all the light from the object passes through the cyan filter. Figure 9 illustrates this.

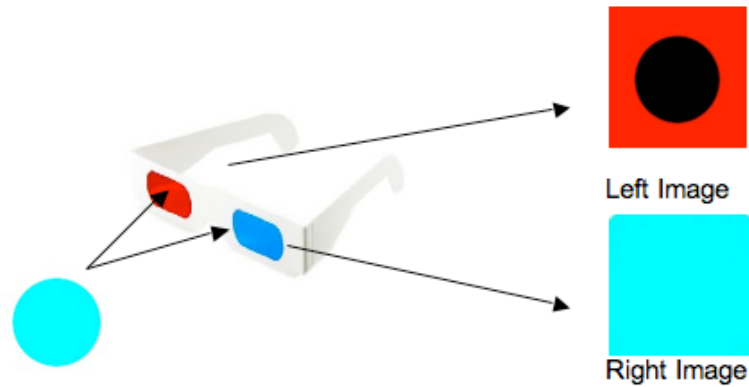


Figure 9. Example of retinal rivalry in anaglyph systems

The black and white circles have very different luminances, which is very difficult for a human observer to fuse into a single image. The equation below is one solution Wimmer proposes to reduce retinal rivalry.

$$\begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} R1 \\ G1 \\ B1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R2 \\ G2 \\ B2 \end{pmatrix} = \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

The change to the matrix forming the left image takes the luminance of the left eye image and turns that into the red channel. This reduces retinal rivalry. For example, a cyan object still appears white to the right eye but doesn't appear completely black in the left eye, because some of the information from the original green and blue channels are now in the red channel of the left image. This method makes it easier to fuse cyan and red objects but at the expense of color reproduction.



Figure 10. Half color anaglyph on the left and full color anaglyph on the right.

Many dislike anaglyph for viewing stereoscopic images for several reasons. First, color reproduction is very poor in anaglyph images. To form colored images for the left and right eyes, color information is lost in the left and right eye images. Also, viewing the images through colored filters changes the colors from the original scene. The combination of complementary

colors for the filter helps the viewer see as wide a spectrum as possible but color reproduction is still poor. Also, there is a lot of ghosting in anaglyph images and it is very hard to control the amount of cross talk. To have a perfect image with no ghosting, the primaries of the display must match the transmission of the filters on the glasses. If the system used red/cyan filters, then the red filter must block all cyan light and vice versa. This is almost never the case because every display will have slightly different primaries. Also, glasses from different manufactures will have different transmissions.

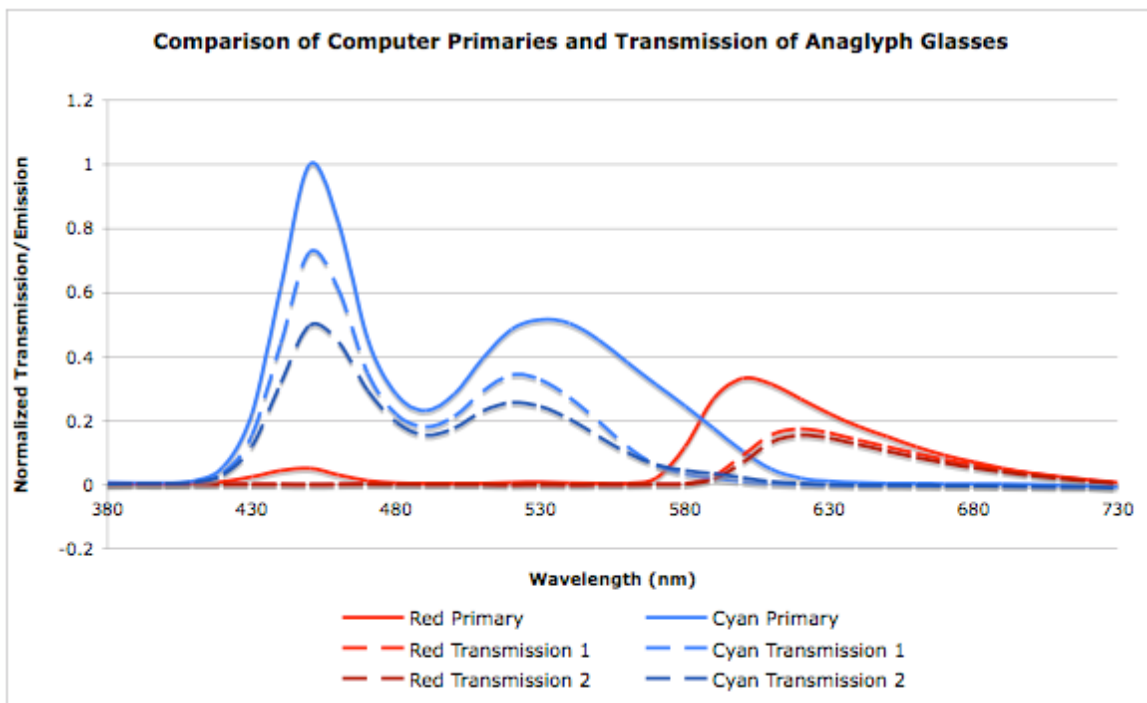


Figure 11. Comparing transmission of two different anaglyph glasses and a computer's primaries.

Figure 11 shows how different the transmission of a filter from anaglyph glasses can be from the primaries of the display the images are viewed on. In this case, the display was a MacBook Pro. When red (255,0,0) is sent to the display, there is some blue light that is being emitted, due to the specific calibration profile of the computer. The red filter cuts out this information and is transmitted through the cyan lens. This can cause ghosting because part of an image, only meant to be seen by the left eye is instead seen by the right eye. This is not the only case where information meant for one eye is seen by the other eye in this example. The cyan and blue primaries overlap, so the left eye sees some cyan information and the right eye sees some red information. Also, the peaks of the transmission of the filters do not align with the peaks of the computer's primaries. This can cause problems because the filters are cutting out important color information.

This is just one example of problems with anaglyph systems. As stated earlier, every display can have slightly different primaries. The graph above only refers to a specific calibration setting on the MacBook, each color profile will have different primaries. Some profiles, like the one used above, will not have completely clean red, blue and green primaries, which will cause ghosting.

In addition to the widely used red/cyan anaglyph method, some have experimented with green/magenta and blue/yellow anaglyph images. Woods and Harris¹¹ conducted an experiment looking at cross talk in red/cyan anaglyph images as well as red/magenta and blue/yellow anaglyph images. They created an algorithm to predict the amount of cross talk and then compared their results to results found from human observers. For LCD and Plasma displays, they calculated red/cyan filters to generally have the least amount of cross talk and the blue/yellow filters to have the most cross talk. When human observers were asked to rank the different anaglyph glasses, red/cyan followed by green/magenta glasses were ranked to have the least amount of cross talk on a CRT display while on a plasma display, red/cyan performed the best followed by blue/yellow. Their results also seemed to suggest that red/cyan and green/magenta glasses are best for LCD and plasma displays and blue/yellow glasses perform best on plasma displays. These results may explain why red/cyan anaglyph images seem to be the most popular method, followed by green/magenta. Very rarely are blue/yellow anaglyph images seen.

It is very difficult to try to correct for any ghosting in anaglyph images unless the viewer will always be viewing the same display using the same glasses. This is very often not the case so it is impractical to implement anti-ghosting methods. A solution for one environment may not work for the next environment. The glasses themselves are also cheaply made and it would not be surprising if the transmission of the filters drifts during manufacture.

2.4 3D Camera Rigs

In order to capture the two offset images needed to create the stereo illusion in 3D, two slightly different images must be captured. The image offset is determined by the human visual system's designated distance between the two eyes on the human head. The perception of the image is based on this offset and misplacement of the cameras in the form of an over exaggerated or under exaggerated distance will effect how the scene is perceived. Just as in the real world, perception of size works in ratios. If the cameras are placed too close together, it is as if the viewer is small and looking at a gigantic scene, just as your eyes would be closer together if you were shrunk down to a miniature size or are a child. If you were a giant whose eyes were farther apart than the average human, or the cameras were placed too far apart, the scene would appear miniature in comparison.

This spacing, along with our eye movements and abilities, must be maintained in the fashioning of a rig. There are many different methods used to capture the two images needed for stereo cinema but the most prominent rigs used today are either rigs that use two cameras and a beam splitter to achieve the offset, or physical placement of the cameras next to one another in a side by side rig that mirrors our eyes. Beam splitter rigs split the light from the two offset areas of the scene, sending it to two cameras or a partitioned sensor. The benefits of this type of rig are its ability to work with larger, higher quality cameras which otherwise would not be able to achieve the correct offset distance for the two images if placed side by side. However, the beam-splitter rig risks malfunction if the prism is not held properly and firmly in place, which is why amateur rig builders can easily run into problems during production.

The other type of rig, a side by side rig, places two cameras side by side with the sensors of the cameras no more or less than the distance between the average human eyes for a realistic representation of the scene. At times the director may move the cameras closer or farther apart for artistic reasons, but to simply attempt to capture a real world scene in 3D, the interaxial

camera distance should be the same as the distance between the two human eyes. In these rigs the manufacturer has a choice with how he plans on positioning the cameras. Our eyes are set to both focus and converge on objects in a scene, meaning they physically angle in to get a clear view of an object. In a rig, this is not a necessity and the cameras can either toe in or remain parallel. When they are not angled in, the convergence of the scene is set at infinity and the images must be adjusted in post production.

There is an industry wide debate over which method is best, parallel or converged, as both had advantages and disadvantages. When shooting in parallel, the point of convergence must be set in post-production by shifting the two images until the focus point is perfectly overlaid on itself for both images. This allows the director greater creative control of the scene after it is shot but of course means longer post time in editing. Shooting converged best mirrors the actions of our eyes, but unlike our eyes, once convergence is set it cannot be so easily changed because of the introduction of exaggerated keystoneing. Perhaps the greater issue with shooting converged is that the two sensors are flat squares facing at two opposite angles to one another. Therefore when the two images are combined, they create a trapezoidal shape since the images were captured in a sort of “X” fashion and then rotated to a flat, parallel space. Parallel seems to reduce this chance of keystoneing problems in capture.

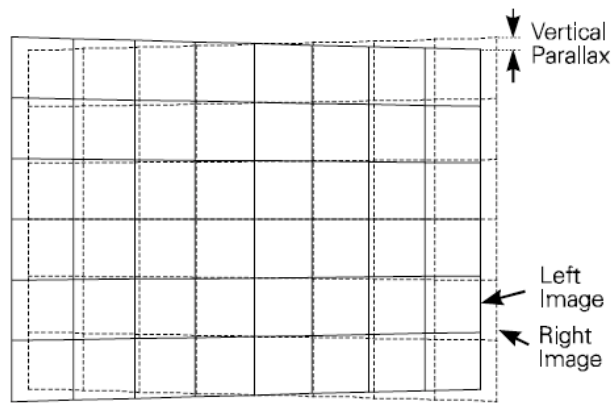


Figure 12. Example of keystoneing

Keystoneing occurs when the image is projected or captured at an angle. The image starts to distort either by moving back in space like a trapezoid, or flaring out to the sides. While shooting in parallel may yield greater post time, shooting converged means more time on set since for each shot convergence must be changed as the subject moves around in the space. The debate still remains present in the industry with strong supporters of both technologies, as both have been shown to work.

3. Workflow

3.1 Pre-Production and Production

3.1.1 Camera Rig

There are many different paths available when designing a rig to shoot in 3D, each with their own benefits and drawbacks. The goal for this workflow was to design a rig that would fit into a 3D workflow specifically for Rochester Institute of Technology students in the School of Film and Animation. To achieve this goal, a rig was designed to fit two Canon 5D Mark II's.

One camera is rotated by 180 degrees as shown in Figure 13 to achieve the smallest interaxial distance possible with these cameras (5 inches). Since one camera is flipped upside down, platforms are used to re-center the lenses vertically. The two back plates on the rig help to keep the cameras parallel and avoid any unwanted rotation, which would cause problems in post production.

This rig is simple and streamlined so that even the most novice cinematographer can use it. The ability to set convergence in post production gives the director and editor greater control over a scene and offers another learning experience for students interested in working specifically on 3D footage by providing them with a crucial job in post production. The final design of the rig with measurements is shown in Figures 13 and 14.



Figure 13. View of the camera rig with cameras.

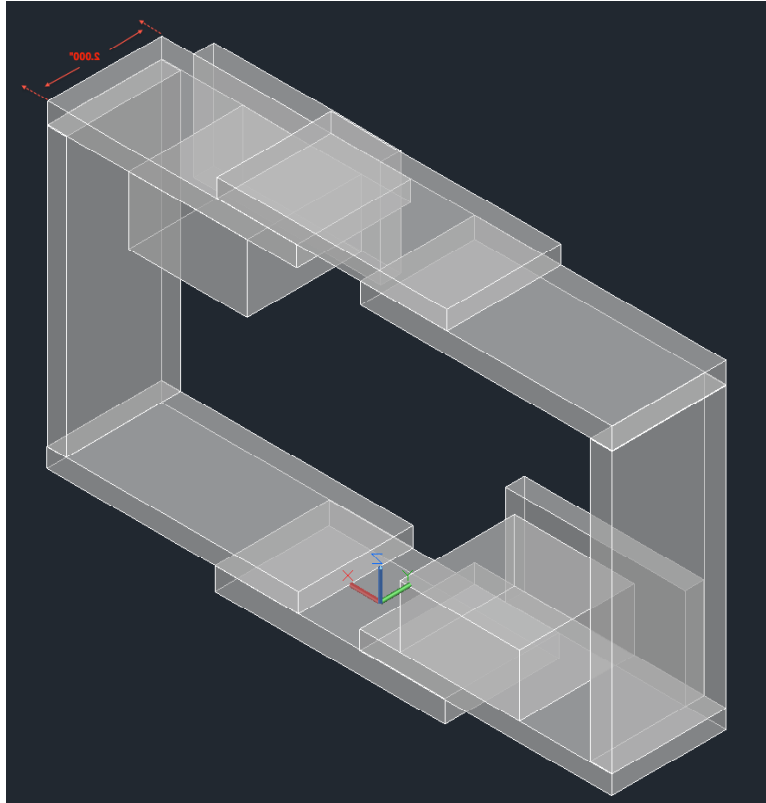


Figure 14. Another view of the camera rig

The need for a live onset preview when shooting 3D movies is clear given how many factors influence the development of a proper illusion of depth in a stereo scene. It is a challenge to maintain a good stereo scene while also focusing on blocking and art direction. In order to accomplish this, the space between the cameras allows for the attachment of two USB cords.

The rig, while designed as a parallel side by side rig for the Canon 5D MKII's, was made to be adjustable for different cameras and possible testing opportunities. As nothing is welded, the back plates can be removed to shoot converged footage, however the rig was not made to note the convergence exactly, which is a crucial part of converged capture. Also, the top and bottom plates holding the left and right halves of the rig together can be detached and different sized plates can be reattached to achieve different interaxial distances. The platforms can also be removed at any time and the popular Flip video cameras can be used either in parallel or converged for shooting.

3.1.2 On Set Preview

An on set live preview is crucial if quality stereo footage is to be captured. Not only is it necessary to be able to see what the cameras are capturing, but it is important to see the convergence of the shot and make any adjustments if necessary. Problems such as excessive divergence and convergence may occur that can easily be corrected on set but not as easily corrected in post production.

To monitor two camera feeds simultaneously (in addition to image processing), the cameras are connected to a Mac computer via mini B USB connection. To bring up a live feed on the computer, a software program called DSLR Remote Pro is used. This program allows Canon DSLR users to directly control the camera from a computer, in addition to displaying live

images on the computer. Two instances of DSLR Remote Pro can be open at the same time, so two windows with the two feeds from the cameras can be viewed simultaneously.



Figure 15. Example of live feed window using DSLR Remote Pro¹²

To take the live feeds and turn them into anaglyph images, Quartz Composer is used. A patch in Quartz Composer takes a continuous screen capture of a section of the computer's monitor. This patch is used to take capture the area of the desktop running the two live feeds.

The program written in Quartz composer takes the two images, flips the feed from the left camera (since the camera is upside down) and applies an anaglyph filter. Two matrices are used to create the anaglyph left and right images. The Wimmer matrices described before are used so that there is minimal ghosting in the images. The user can then use the arrow keys to fix any vertical misalignment and to set convergence. The user is also given live feedback of how many pixels the image has been translated, which will be used later on to set convergence in post production. Allowing the user to adjust convergence on set is extremely helpful because the parallel camera rig has no set convergence (convergence is at infinity). The live feed also allows the user to preview every shot, which reduces the number of mistakes that may or may not be able to be fixed in post. This includes all the things directors monitor for while 2D as well as any mistakes such as excessive depth or conflicting cues, which are problems when shooting 3D.



Figure 16. Screen capture of live preview using Quartz Composer

Even though the live preview is viewed on a computer, the footage is still recorded to the camera and will need to be transferred to a computer. Also, before every shot, a clapperboard is needed to make a synchronization point so that the left and right images can easily be synched in post. If the images are not properly synched, the left and right eye images may be a second or two off from each other.

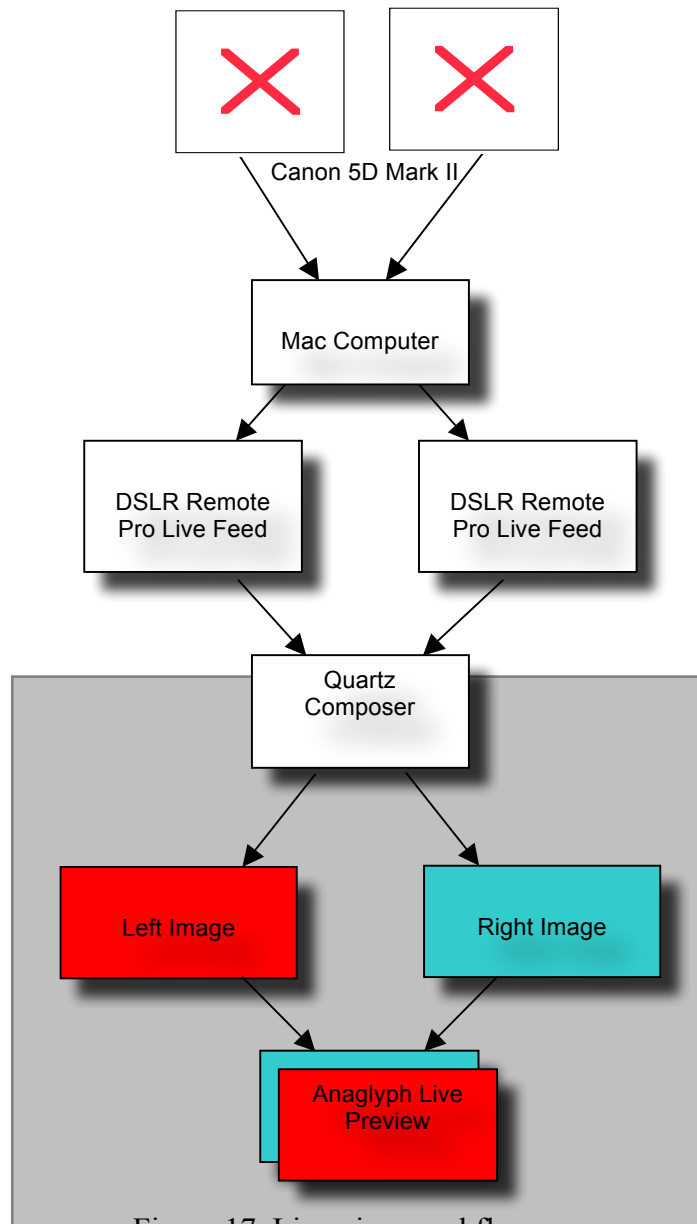


Figure 17. Live view workflow.

3.1.3 Test Chart

An on set live preview also allows the user to check for camera alignment using a test chart. We created a test chart that would check for zoom alignment, vertical alignment, rotational alignment, color mismatch, focus mismatch and any keystoneing. To use the chart correctly, the center cross must line up with the center of the camera rig, so that the center of each lens lines up with the crosses in the yellow and green squares. Also, the chart needs to be horizontally level and parallel to the camera rig. The rig itself also has to be level on the tripod, or else a vertical disparity will be introduced in the images. Every time the camera is moved, regardless of whether the test chart is being shot or not, the rig needs to be leveled.

The yellow and green squares are used to check zoom alignment. Since they are spaced 5.5" apart, the same distance between the center of the lenses on the two cameras, the yellow square is centered in the left camera's frame and the green square is centered in the right camera's frame. When the two frames are overlaid with no convergence adjustment, the two squares will lie on top of each other. If the zoom of the two cameras does not match, the two squares will not lie perfectly on top of each other and one square will be larger than the other. The squares may also not align perfectly if the cameras are not vertically and rotationally aligned.

To check for vertical alignment, the rulers on either side of the chart are used. The two images can be shifted in the live preview so that the left side ruler from the left frame overlaps the right side ruler from the right frame. If the rulers do not match up perfectly, there is a vertical misalignment. If the lines of the ruler all have the exact same offset, then only a vertical pixel shift is needed in post to correct this. If the rulers are skewed at a certain angle, then the cameras are not rotationally aligned. The cameras may not be properly secured in the rig, which can lead to rotational misalignment. This mismatch can also be fixed in post by counter rotating the image from the rotated camera. However, this should not be necessary because when the cameras are positioned properly in the rig, there is no noticeable rotational misalignment in the images.

The color bars and grayscale can be used to find any color differences between the two images. Even when the cameras have the exact same settings, there may be a slight color difference between the two images, which is due to the sensors not having the exact same responsivities. If this problem arises, it will have to be corrected in post. In some cases, the cameras may not have the same color settings so this chart will alert the user to check the camera settings if the colors appear very different in the two cameras.

It is also important that the two cameras are focused at the same distance. The focus fans on the chart help the user to set the focus on both cameras to the same distance.

Finally, the large dashed square helps the user detect any keystoneing issues. Currently the rig only allows for parallel convergence, so there should be no keystoneing issues. However, the rig can easily be adjusted to fit two converged cameras, which is when keystoneing problems are introduced. The two images can be shifted in the live preview so that the two test charts lay over each other with the center crosses on top of each other. If the dashed squares do not exactly line up, (and there is not rotational or vertical misalignment), then there may be some keystoneing problems.

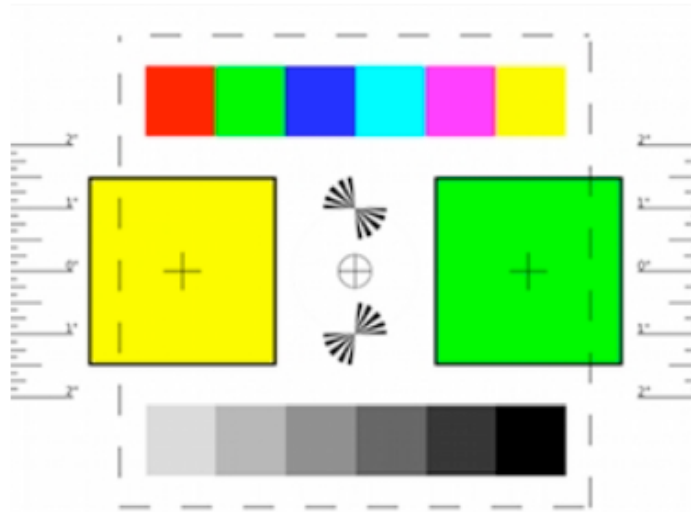


Figure 18. Test chart

3.2 Post-Production

3.2.1. Working in Final Cut Pro

After material has been shot, it needs to be edited in Final Cut Pro. We choose this software over other editing programs because it is an established part of the current SOFA workflow. In Final Cut, the user can edit their footage together conventionally in full color by viewing either the left eye or right eye or they can edit while viewing the image in anaglyph 3D. Figure 19 shows the workflow for post production, with numbers for each step. The User Manual describes this process in more detail.

Before bringing the footage into final cut, each shot needs to be edited to so they all begin at the same mark, using the clapper in the scene for reference. The two video files from each scene must also have identical file names. This syncs the footage and allows for easy use of an EDL later in the process.

Once the footage is brought into FCP, a 2D edit can be made using the right eye footage (1). The user can also preview shots in 3D by using the 3D anaglyph plugin. Once the final 2D edit is made, an EDL is exported (2) and brought back into FCP. In the new sequence created by the EDL, the media needs to be reconnected to the left eye footage, not the right eye footage (3). The new left eye sequence is then brought into the right eye sequence so that there are two video tracks, one for each eye.

To set convergence for each scene, apply the anaglyph filters to each video track and using the wireframe, move the left eye image until you have the desired convergence for each shot (4). If the live preview was used when shooting, the convergence data from Quartz Composer can also be applied to set convergence. To crop out the floating windows left by setting convergence, a calculator in excel was created that centers both the right and left eye images and scales them so the floating windows are outside the final image area. The calculator gives the vertical and horizontal shifts along with the scale but the user has to input this data into FCP for each shot (5, 6).

We created a workflow that allows for the user to primarily edit in 2D, while previewing select shots in 3D. However, the user can decide if they want to use this workflow or create their

own and edit in 3D. Editing in 2D and creating an EDL seems to be the easier, more organized way to move the footage through post production.

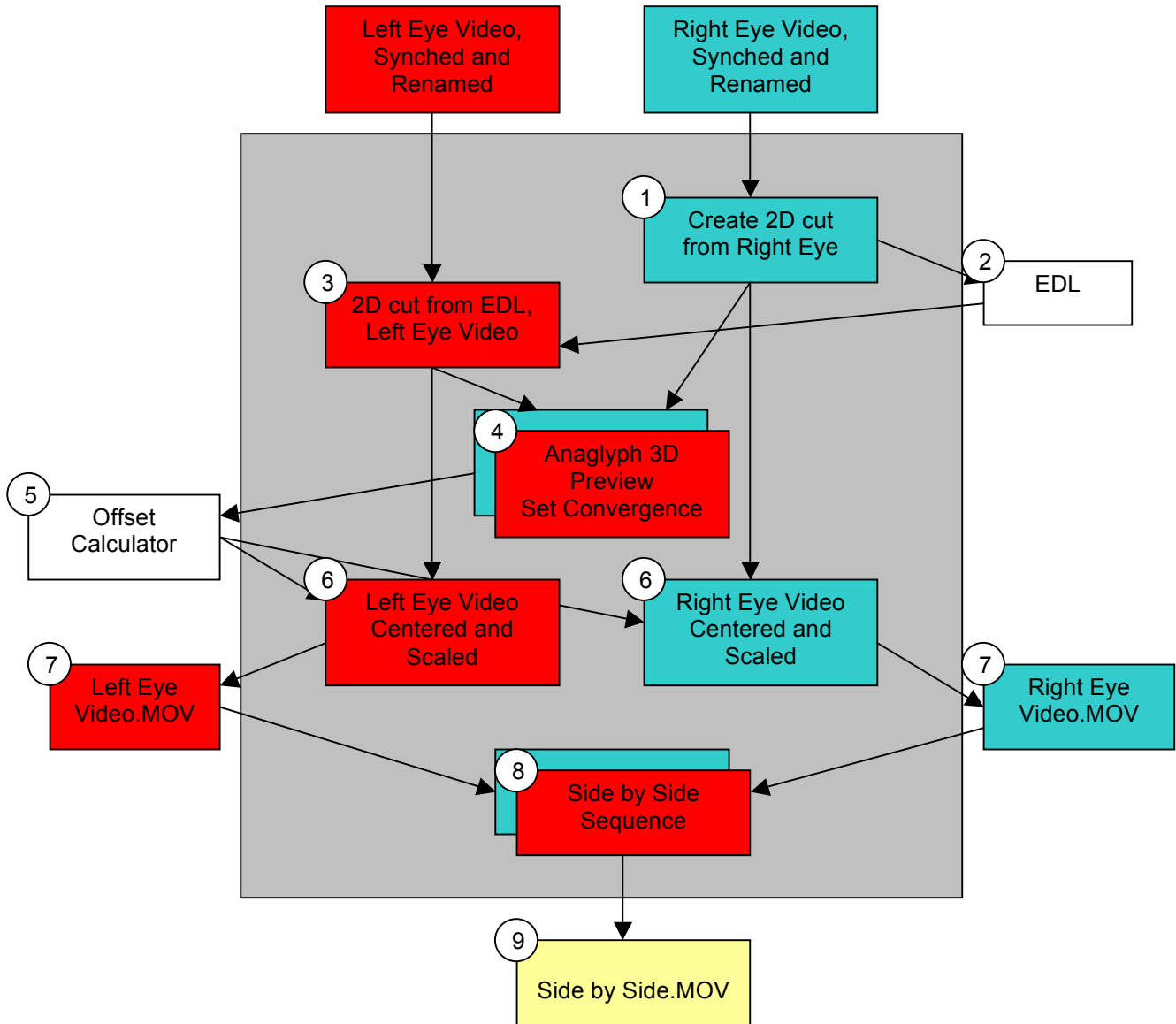


Figure 19. Diagram of post production workflow in Final Cut Pro.

3.2.2. Exporting

Exporting a stereo video from Final Cut Pro happens in two steps. First, two video files need to be exported, one for the left eye video and one for the right eye video (7). Because of the scaling/cropping done in the previous step, the video files need to be exported and then brought back into Final Cut. One both files are brought back in, the side by side final video needs to be

made. A new sequence with dimensions 3840x1080 needs to be created and the left eye image is shifted 960 pixels to the left and the right eye image is shifted 960 pixels to the right (8). This creates a video file twice the size of the original and contains both the right eye and left eye video files. Finally, the sequence needs to be exported (9).

3.3 Exhibition

3.3.1 Driving Two Projectors

There are three main aspects to focus on when projecting 3D footage in a two projector system. Aligning the two projectors, correctly polarizing the light, and properly syncing the video feeds must all be perfected for the viewer to experience the stereo illusion. In order to project stereo footage, two Panasonic AU200 projectors are used, with one placed on top of the other. Each projector has a linear polarized filter placed in front of the lens. The polarizers are set to 45 degrees and 135 degrees from normal to match the linear polarized glasses each viewer must wear. It is important to keep track of which filter is placed in front of which projector. The filter corresponding to the left eye filter in the glasses must go in front of the lens of the projector projecting the left eye image. If the filters are switched, the stereo illusion will be lost and eye strain will be introduced.

The projectors must be carefully aligned, as any misalignment in the projectors will exaggerate the horizontal disparity between the right and left eye images and introduce unwanted vertical disparity. The projectors can be aligned by projecting an image of a grid from both projectors and adjusting the vertical and horizontal positions of the projectors until the two grids are aligned. The zoom and the focus of the projectors must also match.

To drive the projectors, a computer with two DVI outputs is connected to an HDMI converter, which is connected to the projectors. The computer is set up so that the two projectors act as dual monitors, with one projector showing the main desktop and the other projector showing the extended desktop. To play a movie, the side by side Quicktime file exported is stretched across both screens so that half of the file (the 1920x1080 left image) fills up the image area of the left eye projector and the other half of the file fills up the image area of the right eye projector. The side by side file is 3840x1080, which matches the resolution of the combined projectors. This allows for easy playback since the user simply has to play the Quicktime file back at full resolution.

3.3.2 Silver Screen

Silver screens contain silver content that is embedded in the screen so that the screen maintains polarization of the light bouncing off the screen. Silver screens are necessary for stereo projection so that the polarized light from the projectors is still polarized when it hits the glasses the viewer is wearing. Because of the metallic properties of silver screens, these screens tend to have a hot spot that normal white screens do not. This hot spot is described as the screen's gain. The specific screens at SoFA have an on axis gain of 2.8. As soon as the viewer goes off axis, the gain decreases so that the image becomes more and more dark. Figure 20 illustrates this point. These properties of the silver screen therefore limit the viewing environment the screen will be in.

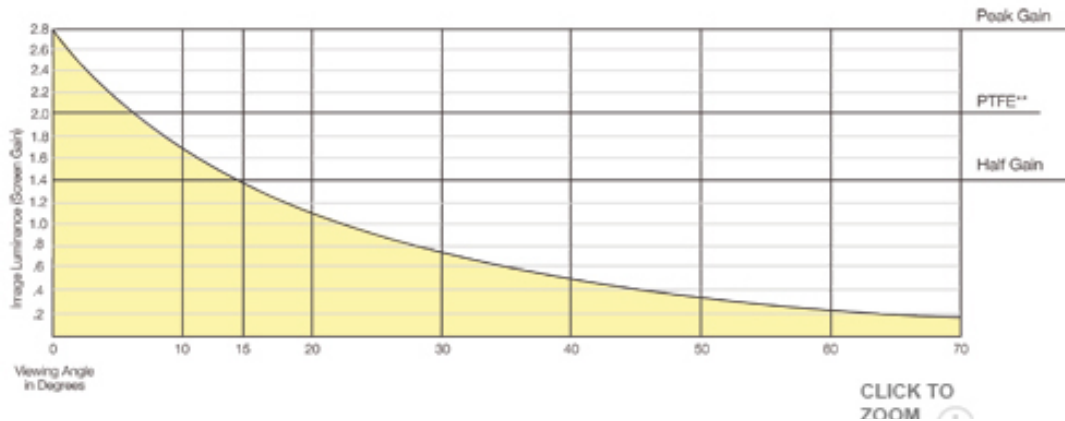


Figure 20. Viewing angle versus gain on a silver screen

3.3.4 Environment

RIT's School of Film and Animation has two different sized silver screens. The largest projected image size achievable on the smaller screen is 36"x65" and the largest image size achievable on the large screen is 70"x125". The smaller screen can be used in any environment but it is recommended that the distance between the projector and the screen is 16'. This ensures that the maximum image size is achieved with maximum brightness. For the larger screen, a screen to projector distance of 20' is recommended.

The larger screen will be permanently mounted in A080 and the size of the room limits the viewing angle of the screen. The smaller screen is portable so it is important to make sure that the viewing angle is limited so that no viewers are too far off axis. Viewers that are off axis will see a dim image so it is important to try to keep as many viewers on axis as possible. Also, the closest seat to the small screen should be no less than 4' so as to reduce excessive image depth. The closer a viewer is to the screen, the more space the images occupy in the retina and the larger the retinal disparity. This can cause eye strain if the retinal disparity is too large. For the large screen, the closest seat should be no closer than 8'.

4. Testing 3D

4.1 Determining Acceptable Depth Range

One goal of the project was to create guidelines to creating stereo images that are comfortable to look at. We ran a qualitative test to determine an acceptable range of depth in an image and we shot several qualitative shots as well. A test was run using several observers to determine when a stereo image was not longer comfortable to look at and when the viewer could no longer fuse an image.

4.1.1. Experiment Procedure

An image of a light bulb was taken and several different images were created from it, at varying depth levels. To create the illusion of different depth levels, the images were shifted so that they had different disparities, ranging from a zero pixel disparity to a 75 pixel disparity. The projector was 4.88m (16') away from the silver screen, which resulted in an image width of 1.64m (36"). We had viewers sit two meters and four meters away from the screen. Table 1

shows the different image disparities in pixels, meters and the angle the image forms in the viewer's eye.

Table 1. Image disparities.
Numbers in green box are divergent and numbers in the red box are convergent.

Disparity in Pixels	Disparity in Meters	Angle at 2m (degrees)	Angle at 4m (degrees)
70	0.059266667	39.6036393	23.10755933
55	0.046566667	28.66379599	17.54050256
50	0.042333333	25.43092397	15.83521529
45	0.0381	22.30489785	14.14517418
40	0.033866667	19.29812821	12.47307251
35	0.029633333	16.41949009	10.82142487
30	0.0254	13.67454676	9.192549128
25	0.021166667	11.06591279	7.588552851
20	0.016933333	8.593695619	6.011324143
15	0.0127	6.255962802	4.462526735
10	0.008466667	4.049195692	2.943599084
5	0.004233333	1.968703184	1.455757091
0	0	0	0
5	0.004233333	1.84496796	1.422881021
10	0.008466667	3.581469532	2.812292126
15	0.0127	5.215874546	4.167824358
20	0.016933333	6.754492974	5.489240025
25	0.021166667	8.203471644	6.776458322
30	0.0254	9.568722513	8.029540579
35	0.029633333	10.8558759	9.248675496
40	0.033866667	12.07025297	10.43416462
45	0.0381	13.21685273	11.58640831
60	0.0508	16.29514804	14.84887725
75	0.0635	18.91323441	17.83338982

The angle in the viewer's eye and the disparity in meters depend on the projector set up and only apply to the specific set up used. However, through some calculations they can be applied to other projection set ups.

The images were presented to the viewer in a random order and each image was on the screen for five seconds. After each test image, a white dot was projected on the screen for five seconds to allow the user to reset their accommodation and vergence to the screen plane. After viewing each test image, the subject was asked if they were able to fuse the image and if they saw a double image. If the subject reported seeing a hint of a double image, they may have been able to fuse the image to see the 3D depth effect, but not completely and so it was uncomfortable to look at and they were able to see a faint outline of the left and right eye images. In this experiment, a total of nine subjects were used and they reported that they were able to fuse all images, but that a double image was noticeable in some cases.

4.1.2 Results

To determine the comfort threshold, we only looked at cases where at least 75% of the subjects said they did see a double image or when at least 75% said they did not see a double image. Table 2 shows the results found.

Table 2. Results.

Boxes shaded in gray are disparities where viewers saw a double image.
Boxes in purple are disparities where viewers saw no double image.

Disparity in Pixels	Angle at 2m (degrees)	Angle at 4m (degrees)
70	39.6036393	23.10755933
55	28.66379599	17.54050256
50	25.43092397	15.83521529
45	22.30489785	14.14517418
40	19.29812821	12.47307251
35	16.41949009	10.82142487
30	13.67454676	9.192549128
25	11.06591279	7.588552851
20	8.593695619	6.011324143
15	6.255962802	4.462526735
10	4.049195692	2.943599084
5	1.968703184	1.455757091
0	0	0
5	1.84496796	1.422881021
10	3.581469532	2.812292126
15	5.215874546	4.167824358
20	6.754492974	5.489240025
25	8.203471644	6.776458322
30	9.568722513	8.029540579
35	10.8558759	9.248675496
40	12.07025297	10.43416462
45	13.21685273	11.58640831
60	16.29514804	14.84887725
75	18.91323441	17.83338982

As can be seen by the results, there is no definite threshold where all the subjects started to see a double image. Every person has a slightly different threshold and some of the viewers had a difficult time determining if they saw a double image or not. But this experiment gives a rough outline of how far of a disparity can be used before viewer confusion and discomfort begins, which is around 9-10 degrees for divergent images and about 5 degrees for convergent images.

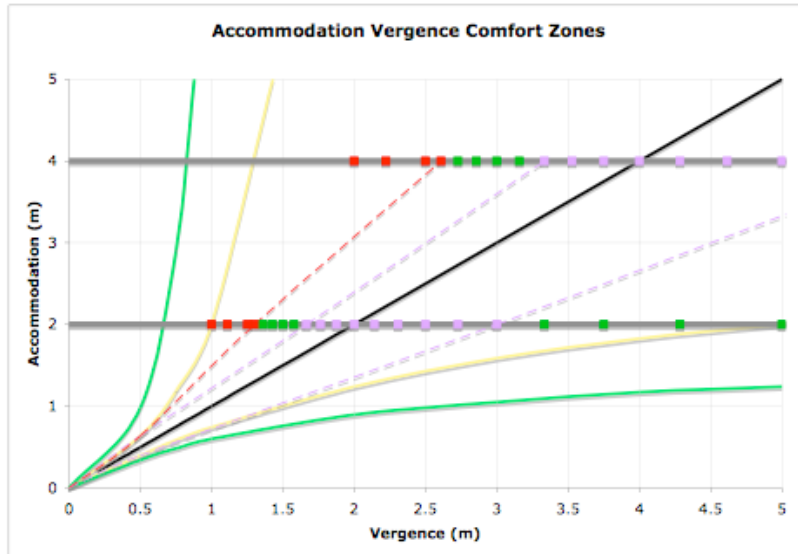


Figure 21. Test results mapped against a vergence accommodation graph¹³.

Figure 21 shows how these figures fit into an accommodation vergence comfort zone chart. The black line is single vision, where accommodation and vergence are the same. The yellow solid line represents the viewing comfort zone and the green line represents the limits of binocular vision. The two gray lines are the two viewer-screen distances used in the experiment. The purple dashed line represents the comfort zone from our results and the red dashed line represents the area where fusion problems start to be introduced. The green dots are the disparities that straddle the line between a single and double image.

Our results differ from the actual zones of comfort and binocular vision. We did not test for any points outside the comfort but our results suggest we have data points that should lie outside of the comfort zone. One reason this may be is that what one person determines is a comfortable image to look at may be very uncomfortable for another person to look at. We based our results off of the presence of a faint double image as the threshold between comfortable and uncomfortable, but maybe a faint double image is still considered comfortable to look at by the creators of this graph. The reason we used the presence of a double image as the threshold is because we are looking for recommendations to make about how disparity film students at RIT can put in their images. If there is a double image, then the image is not acceptable to look at, regardless of how comfortable the image may or may not be.

Based on these results, when the projected image size has a width of 1.64m (which is the largest image size using the smaller silver screen), it is recommended that the image does not have a disparity much larger than 30 pixels for divergent objects and 20 for convergent objects. This is assuming the closest viewer is at least 4m away from the screen. For use of the larger screen, it is recommended that divergent images do not have a disparity greater than 15 pixels and convergence images do not have a disparity greater than 10 pixels. The larger screen would have an image width of 3.2m and the closest seat recommended would be 2.4m from the screen.

These results are concerning objects in the scene that the audience will look at directly. If the shot has excessive depth but it's only in the background and the viewer's attention is directed elsewhere, then the image will look alright. In any stereo scene, you want to direct the viewer's

attention to the converged object, away from any areas that may have excessive depth, such as the background of the shots.

4.2 Recommendations

When it comes to shooting a 3D movie much more must be taken into account than just the basic concepts of film capture and cinematography. The introduction of the stereo-illusion demands its own set of rules for correctly maintaining the 3D experience, which must exist in conjunction with the director's own artistic desires. There are some tools that are commonly used in regular filmmaking that must be rethought for 3D movies. We must always remember that we are working with depth cues and that where we set convergence greatly effects the stereo illusion. At the point of convergence, the two images are perfectly aligned on top of one another. This means then that at all other parts of image, the images are disparate and will need to be fused by the visual system to create the illusion of depth.

The most common setup for dialogue involves the over the shoulder shot, allowing the edge of one speaker to exist in the frame, set out of focus near the camera while we look over their shoulder to the other person in the conversation. This common shot is problematic in 3D as it offers us directly conflicting cues to the viewer by "breaking the window" as it is called. If we are looking over someone's shoulder then our convergence and usually the focus is set on the person whose face we are looking at. This puts the speaker at the screen plane in terms of depth and everything in front of them, aka the shoulder of the other person at the edge of the frame, as coming out at the viewer. The cues given by the retinal disparities are telling you that the person at the edge of the frame is coming out at the audience. However the existence of the frame as the "window" cuts off the person whose shoulder we are looking over. This cutting off at the edge of the frame is a cue that is telling us the person must be behind or at screen plane. Theoretically if the person is coming out at the audience, then he should be fully visible as we look over his shoulder. The only way to compensate for this windowing problem is to set convergence so that everything in the scene is moving back into the screen. However, if this is done, the background and the person in focus can become so divergent that the retinal disparity is too great and the images cannot be fused. It is for this reason that over the shoulder shots are not suggested.



Figure 22. Over the shoulder shot convergence options

Another popular artistic trick is the use of rack focus in a scene to draw attention to a specific area of the image. Since two cameras are used to shoot, the cameras must always be in sync. Introducing a rack focus is difficult because the focus must change identically in each camera. The same applies to a zoom in during a shot. If one camera zooms faster than the other, our eyes will not be able to fuse the images and the stereo effect is lost.

We must always remember that shooting stereo movies mirrors human vision but by no means has the same capabilities. If an object on the screen comes out too far towards the audience, this can cause a problem. The gag of an axe flying toward the screen or a hand coming out all the way into the audience must be controlled and used in moderation. If the image comes out too far and the two images may have too large of a disparity and they may not be able fused. This results in a double image. Similarly if an object changes depth too quickly, the viewers eyes will not have a chance to fuse the images. Even the human visual system takes a moment to adjust the convergence of the eyes and focusing of the object. To overcome this problem depth must not be overdone especially when dealing with objects coming out of the screen at the audience, and the changes must be slow enough for the visual system to fuse. An option is to have the object fly out in slow motion to achieve the gag without the viewer discomfort.

At all times on set the edges of the frames as well as lens flares and reflections must always be checked to ensure that the images for each eye are the same. Since the two cameras will be capturing light at slightly offset distances, a lens flare in a scene or reflections off of objects or mirrors may only appear in one eye. Similarly the different images mean that if a small object is on the edge of the frame of one camera, it may not be in the frame of the other. When this happens, the two eyes fight for dominance and the image becomes more challenging to fuse, and a flickering back and forth can occur.



Figure 23. Examples of objects on the edge of screen.
Half of a ketchup bottle is on the right, but only seen by the left eye.
Half of a box is on the left but only seen by the right eye.

When panning with a 3D camera rig for stereo display, the speed of the pan must be carefully monitored. With objects moving into and out of the frame across the scene, the pan must give viewers enough time to adjust to the changing depth within the scene. Either the pan must be very slow to allow the viewer to see the space and have the time to properly fuse the images, or it must be a snap pan. With a snap pan that is fast enough, objects are not discernible so the viewer is not trying to fuse the images to achieve the perception of depth.

With all of these problems sometimes the greatest challenge is knowing what is causing a problem. Since all of the cues in a scene work together, if one or two are wrong it may not be

clear why the image causes discomfort but it is clear that the viewer cannot enjoy the film. This is why it is extremely important to carefully monitor your footage while shooting and even seek the advice of a stereographer.

It is also important to use 3D as a tool and not continuously have extreme depth in every shot. Use extreme depth as a tool to get a point across to the viewer. A director must also plan depth accordingly as any jumps in depth between cuts can be confusing for the viewer. Just as plot is thought out and planned with art direction, so must 3D shooting and depth mapping must be planned before a film is made.

5. Conclusion

The goal at the beginning of this project was to create a complete stereoscopic 3D workflow from capture to exhibition to fit into the existing Rochester Institute of Technology School of Film and Animation workflow and facilities. The final product consists of a rig for dual camera capture with a live preview for 3D viewing on set, guidelines for shooting proper 3D for a comfortable and effective stereo-illusion, a post production workflow for editing and export, and a dual projection linear polarized system for exhibition. Through test shoots and research it was determined what cinematography principles do not translate from 2D to stereo capture. Shooting over the shoulder shots causes problems with breaking the window as does any unforeseen objects that should be popping out in front of the screen but are cut off by the edge of the frame. If the eyes are not given proper time to fuse an image as it changes depth, a double image will occur and cause viewer discomfort. This ties into objects that come too far out of the screen at the audience. If the images are too disparate, the viewer cannot fuse them due to the large retinal disparities. When panning across a scene, the speed of the pan must be very slow to allow the changing depth information to be comprehended and processed, or so fast that nothing is discernible such as in a snap pan. Someone on set must also always keep an eye on the edges of the frame as well as any lens flares and reflections that will only appear in one eye. The final product was a success and has been exhibited. Capture, editing, exporting and projection all worked successfully. While the goals of this project were met, there are still many future adjustments and improvements to be made, to tighten up the workflow and perhaps find faster methods of editing, projection, or exporting. There are also many tests to be run to learn more about 3D and the thresholds of the human visual system for comfortable fusing of stereo images. Further experiments to determine differences in color, focus, and exposure may be further examined to gain a greater understanding of how to create a quality 3D movie.

References

1. Spiro, Lisa. "A Brief History of Stereographs and Stereoscopes." *Connexions*. Oct. 30, 2006. <<http://cnx.org/content/m13784/1.5/>>.
2. Zone, Ray. Stereoscopic Cinema & the Origins of 3D Film 1838-1952. University Press of Kentucky, Kentucky, 2007.
3. Rownd, Henry. "Early 3D: A Case Study." *Bright Lights Film Journal*, Issue 68, May 2010. <<http://www.brightlightsfilm.com/68/683D.php>>.
4. Hayes, R.M. 3D Movies: A History and Filmography of Stereoscopic Cinema. McFarland & Company, Inc., Publishers, Jefferson, NC, 1989.
5. Barber, Laura. "New SMPTE 3D Home Content Master Requirements Set Stage For New Market Growth - SMPTE.org." *Home - SMPTE.org*. SMPTE, 22 Apr. 2009. Web. 08 May 2011.
6. HUBEL, D. H. & WIESEL, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106-154.
7. M. Lambooi, W. IJsselsteijn, M. Fortuin, I. Heynderickx, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review", *Journal of Imaging Science and Technology* 53, 030201(2009).
8. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue", *Journal of Vision* 8, 1-30 (2008).
9. McAllister, David F., Zhou, Ya and Sullivan, Sophia. "Methods for Computing Color Anaglyphs." *Electronic Imaging, '10*, San Jose, CA, Jan, 2010, *Stereoscopic Displays and Applications XXI*, Vol. 7524
10. Wimmer, Peter. "Anaglyph Method Comparison." 2005. <http://3dtv.at/Knowhow/AnaglyphComparison_en.aspx>
11. A. J. Woods, C. R. Harris. "Comparing levels of crosstalk with red/cyan, blue/yellow, and green/magenta anaglyph 3D glasses." *Proceedings of SPIE Stereoscopic Displays and Applications XXI*, vol. 7253, pp. 0Q1-0Q12, January 2010.
12. DSLR Remote Pro for Mac, Breese Systems, 2010. <<http://www.breezesys.com/DSLRRemotePro4Mac/index.htm>>
13. Banks, Marty. "Stereoscopic Vision: How Do We See in 3D?". 2009 <http://www.etcenter.org/files/publications/Marty_Banks_NAB09.pdf>