A Viewer-Centric Editor for Stereoscopic Cinema

Sanjeev J. Koppal* C. Lawrence Zitnick† Michael F. Cohen† Sing Bing Kang† Bryan Ressler†

Alex Colburn‡

Figure 1: *Sample outputs of our viewer-centric, interactive editing tool for stereo-cinematography. (I) Input stereo pair, (II) digitally altered versions: (a) modified horizontal image translation, (b) modified field-of-view (FOV), and (c) moved (dolly) virtual camera forward while changing the perceived depths of vertical window edges (proscenium arch). Note the parallax changes in all the results.*

Abstract

A digital editor provides the timeline control necessary to tell a story through film. Current technology, although sophisticated, does not easily extend to 3D cinema because stereoscopy is a fundamentally different medium for expression and requires new tools. We formulated a mathematical framework for use in a viewer-centric digital editor for stereoscopic cinema driven by the audience's perception of the scene. Our editing tool implements this framework and allows both shot planning and after-the-fact digital manipulation of the perceived scene shape. The mathematical framework abstracts away the mechanics of converting this interaction into stereo parameters, such as interocular, field of view, and location. We demonstrate cut editing techniques to direct audience attention and ease scene transitions. User studies were performed to examine these effects.

1 Introduction

Stereoscopic cinema is seeing a recent resurgence in popularity. At least 14 mainstream 3D movies are scheduled to be released in 2009 both in specially equipped theaters and in IMAX (see, for example, `http://www.3dmovielist.com/`). Stereo cinema requires providing slightly different points of view to each eye. The differing position of scene points (i.e., disparity) creates the illusion of 3D depth. For many films, the potential to create a visually stunning experience outweighs the extra work needed to overcome the challenges of creating stereo. Stereo requires new tools for planning and post processing that leverage recent advances in stereo vision systems. In addition to regular 2D film editing parameters such as field of field (FOV) and camera position, additional degrees of freedom exist, such as camera vergence and interocular. Each of these operations has perceptual implications.

1.1 Overview of our work

We present a viewer-centric editing interface that provides the editor with new degrees of freedom specific to stereo. Our editing technique is driven by the **audience's experience**, rather than by mere manipulation of camera parameters. This is possible due to a mathematical framework that explains previously recorded perceptual effects and abstracts away the camera-centric calculations usually necessary in 3D

*Harvard University, †Microsoft Research, ‡University of Washington

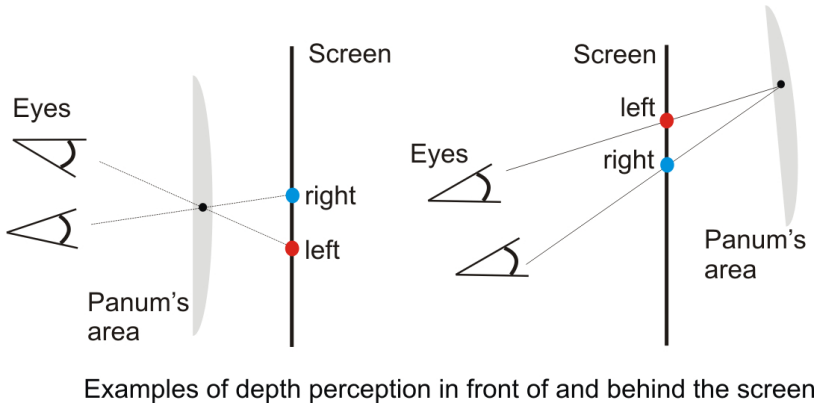


Figure 2: ***Depth perception.*** *During a stereo movie, the eyes verge on a point, which gives absolute depth to the viewer. In an area around this point (called Panum's area) the brain merges the two images to form a single image, thus perceiving relative depth.*

cinema. A stereo-cinematographer can, therefore, concentrate on the desired visual experience while our tool converts the edits, automatically, into camera parameters. These parameters may be used to either render new stereo frames or plan future shots at the same scene.

Before a shot takes place, a director can obtain either rough video or still photographs of the scene. Our interface can then offer a digital "dry run" of the scene by depicting how the rough cuts would be perceived, from the point of view of the audience. If the predicted stereo experience is not what was envisioned, the director can change the shot plan using new camera parameters calculated by our editing tool. This sort of 'stereoscopic preview' ensures that the right configuration of the stereo rig will occur when the real shooting takes place, saving time and money.

After the shooting has finished, our editor can still digitally enhance or remove stereoscopic effects with a variety of tools. These include changing the horizontal image translation, the field-of-view and/or modifying the *proscenium arch* (the perceived depth of the screen's edge). We also allow small changes in the (virtual) camera positions by dollying and varying the camera's interocular, also known as the camera's baseline. Any shift in camera position requires rerendering using pre-computed image disparities.

In stereoscopic film, scene transitions may result in visual discomfort if the shots are not designed carefully. This is because the human visual system requires time to adjust to drastic changes in visual cues [Levonian 1954]. Thus, in addition to edits within a shot, our tool also allows stereoscopic parameters, such as the horizontal image translation, to be cross-dissolved around shot boundaries even when the shots themselves are hard cuts. The cross dissolve of viewing parameters can result in more comfortable shot transitions.

We validate many of the expected perceptual effects of the edits through user studies. In particular, we examine the flattening and depth amplifying effects of changes in field-of-view, comfort of shot transitions relative to disparity changes, and the role the proscenium arch can play. Since rerendering based on stereo can introduce artifacts, we also show that rerendering only a single eye's view can reduce the effects of such artifacts while achieving the desired edit. Finally, we demonstrate the ability of our tool to direct the audience's attention by manipulating image translation and achieve special effects such as a stereoscopic "Hitchcock" zoom.

The major contributions of this work are:

- A mathematical framework to explain perceptual effects on varying stereo parameters.
- A viewer centric editing interface for planning shots, and modifying individual shots and their transitions.
- A user study that examines predicted perceptual effects, validating the usefulness of an editing system.

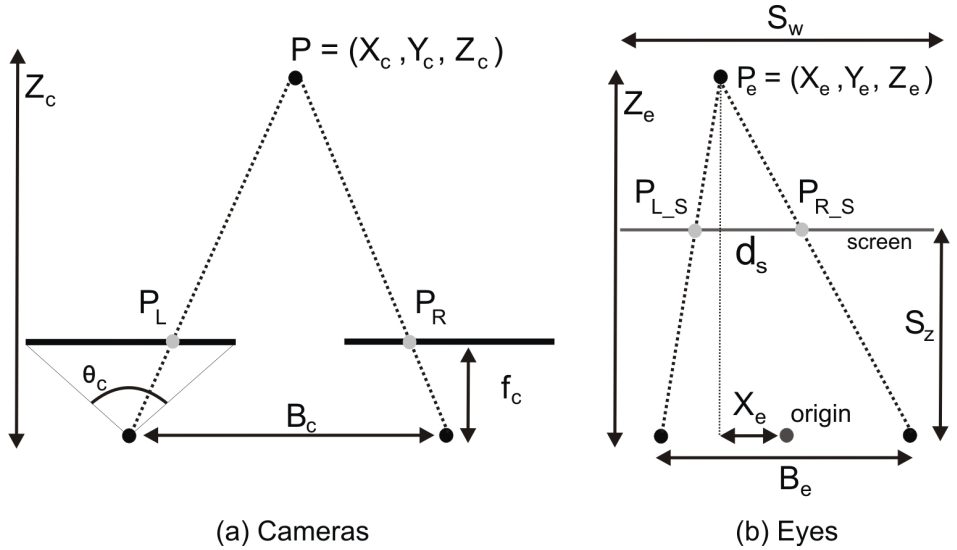


Figure 3: ***Rectified camera and eyes.*** *Here we show the disparities created when a rectified stereo pair views a point P. The point P_e is the perceived location of P when viewed by the eyes on the right.*

As Figure 2 illustrates, perception in a 3D movie first happens when our eyes converge on an object [Valyus 1962; Julesz 1971; Pfautz 2000; Nagata 1991]. Rather than experiencing "double vision", our brain fuses the stereo images to create a perception of scene depth. This occurs in a region known as Panum's area [Buser and Imbert 1992], which results in perceived relative depth about the vergence point. A 3D movie can cause discomfort since there is a discrepancy between accommodation (eyes focus on the screen) and convergence (where the lines of sight cross) which does not exist in the real world. Careful camera setup coupled with digital image rectification (the removal of vertical disparity e.g., [Loop and Zhang 1999]) can help avoid these issues. We do not address this problem in this paper.

1.2 A brief history

Wheatstone [1838] is arguably the first to discover stereopsis and defined "disparity" in terms of differences in subtended angles. Following this work, Brewster, in 1844, built the first viewing device, called the stereoscope. Helmholtz and Rollman brought about the anaglyph (red-cyan) format while the polarized method was introduced by Norling in the United States and Spottiswoode in Britain in the 1940s [Lipton 1982]. These and other inventions fueled a boom in 3D movies in the 1950s, which was followed by a chequered run of popularity to the current day [Sammons 1992].

Along with the rise of stereo cinema, there has been the development of many related areas of research and engineering. Much work has been done on human perception in stereo films, such as improving the 3D movie experience [Held and Banks 2008; Kim et al. 2008; Siegel and Nagata 2000] and understanding dizziness and other physiological effects [Lambooij et al. 2007]. 3D displays (including autostereoscopic displays) have been introduced, allowing color stereo movies to be displayed on the desktop [Beaton 1990; Tessman 1990; Schwerdtner and Heidrich 1998]. Another area is display technologies which use high-speed mirrors and projectors [Traub 1967; Jones et al. 2007]. Virtual reality [Burdea and Coffet 2003; Rheingold 1992; Robinett and Rolland 1991] and human-computer interaction applications exist. Finally, there has been considerable amount of work on 3D television to find protocols for transmission, encoding and portable display solutions [Matusik and Pfister 2004; Isono et al. 1992].

1.3 Current Stereoscopic Editing Technology

Since stereoscopic movies have begun to make significant impact on studio revenues, a variety of editing software have been proposed. Our editor is broadly different from these since it has been designed solely with the viewer's experience in mind. For example, many editors, such

Known effect	Heuristic used or commonly held belief	Geometric explanation
Cardboarding	Keep object "roundness" over 20% [Empey and Neuman 2008]	Camera focal length (f_c) > Eye focal length (f_e)
Pinching	Match eye-camera field-of-view [Clark 2007]	Camera focal length (f_c) < Eye focal length (f_e)
Gigantism	Caused by narrow camera baseline [Streather 2007]	Camera baseline (B_c) < Eye baseline (B_e)
Miniaturization	Avoid hyperstereoscopy [Lipton 1982]	Camera baseline (B_c) > Eye baseline (B_e)

Table 1: *To avoid inadvertent stereoscopic effects, film-makers commonly use heuristics. Of the many experienced artists, we select quotes from stereo-cinematographer Lenny Lipton, documentary producer Barry Clark, IMAX film-maker Phil Streather and Disney animators Mark Empey and Robert Neuman. These heuristics have geometric explanations, which can be exploited to enhance or remove stereoscopic effects.*

as [Wang and Sawchuk 2008; Suto 2006; Kawai et al. 2002], provide significant control, but do not model any viewer characteristics, such as eye position/parameters, and only directly manipulate the disparity map or the raw images. Others are designed simply for shot-planning on location rather than for post-production. These include such the Stereoscopic Calculator by Florian Maier and Mueller et al.'s ([Mueller et al. 2008]) system for easier shooting of live action 3D movies. Masaoka et. al ([Masaoka et al. 2006]) also use a bird's eye view of the scene: however, they do not allow user interaction with the reconstructed point cloud for re-rendering of images. This characteristic is what allows our editing software to enable the creation of new effects, such as the Hitchcock effect for stereoscopy. Finally, there are a large number of commercially available editing tools. The inner workings of these tools are proprietary and are not easily compared with our interface. However, our editor is different since it allows control of the camera position in 3D (unlike Tweak's RV) including dollying the camera forward (unlike the Foundry's Ocula plugin for Nuke). Quantel's editing software provides a broad swathe of controls for stereoscopic content. However, the fundamental primitive for these tools are to adjust image and camera parameters to achieve the desired 3D view. In contrast, our editor is built around user interaction of a point cloud that correctly depicts the viewer's 3D experience: the rendered images follow as a by-product of this. In this sense, our viewer-centric editor complements other camera-centric tools. Finally, our editing framework allows blending of all the different stereoscopic parameters over the transitions between cuts, whereas other softwares are limited to a few of these parameters.

Format and Glasses: Many formats exist to simultaneously display the two stereo images to our eyes. The stereo pair is multiplexed either in time (using fast projectors and displays), in space (with alternate rows or columns belonging to different images) or in wavelength (through the red, green and blue color channels). As we later explain in our experimental setup, we used a polarized color display for our results. While the need to distribute our content requires us to downgrade to anaglyph, we wish to remind the reader that this format may contain compression artifacts, such as bleeding between colors, which were not present either during development or in the user studies. The reader must obtain these glasses from a store such as [Direct 2009]. The red-cyan glasses must have red for the left eye and cyan for the right eye.

2 A geometric framework for stereoscopic effects

A stereo movie experience is the result of a complex combination of factors such as camera parameters, viewing location, projector-screen configuration, and psychological factors [McKay 1953]. The viewing experience can range from pleasant to distracting or even cause eye strain (e.g., due to long exposures to large disparities). The communities of 3D film-makers and photographers have learnt over the years the various heuristics for avoiding or deliberately enhancing well-known stereoscopic effects. Table 1 lists the major effects and their representative heuristics.

A fair amount of work has been done on modeling these distortions (e.g., [Grinberg et al. 1994; Woods et al. 1993; Masaoka et al. 2006]). We use a different framework which abstracts the camera-projector-screen-viewer geometry as ratios, allowing easy manipulation by a user. This editing setup also suggests a geometric interpretation of the major stereoscopic effects. We investigate a rectified stereo setup (Figure 3) and assume that the eyes can be represented as pin-hole cameras with parallel optical axes (as validated by [Wald 1950]). This approach is

Variable name	Geometric meaning
(X_c, Y_c, Z_c)	Real world coordinates of point P
(X_e, Y_e, Z_e)	Perceived coordinates of P
$\mathbf{p}_L = (c_L, r_L)$	Left image coordinates of P (similarly for $\mathbf{p}_R$)
$\mathbf{p}_{L_S} = (c_{L_S}, r_{L_S})$	Left screen coordinates of P (similarly for $\mathbf{p}_{R_S}$)
B_e	Eye baseline
B_c	Camera baseline
d	Image disparity
d_S	Screen disparity
f_c	Camera focal length
S_z	Viewer screen distance
S_w	Screen width
V_c	Horizontal shift (vergence)
V_{c0}	Original vergence position
θ_c	Camera field-of-view
θ_{c0}	Original camera field-of-view
α_θ	Ratio change for FOV
B_{c0}	Original baseline
α_B	Ratio change for baseline
Z_s	Dolly (forward camera shift)
Z_{s0}	Original doll position
α_Z	Ratio change for dolly

Table 2: *Here we list the variable names for all the stereoscopic parameters used in this work.*

most similar to [Held and Banks 2008] who suggest that geometry is a conservative predictor of what humans can actually fuse. However, while they investigate the relationship of the geometric setup to actual perception, we exploit it to model the visual experience of a cinema audience.

We assume several parameters associated with the viewer's experience are known. These include the screen width S_w, the distance from the viewer to the screen S_z, and the distance between the viewer's eyes B_e. We assume that all parameters share the same units, and the world coordinates are centered between the viewer's eyes. Thus, the positions of the left and right eyes are $\{-\frac{B_e}{2}, 0, 0\}$ and $\{\frac{B_e}{2}, 0, 0\}$, respectively. Let the left and right image widths be W. We use the ratio $S_r = \frac{S_w}{W}$ to map pixel locations to physical screen location.

Let a corresponding pair of points across the left and right images be $\mathbf{p}_L = (c_L, r_L)$ and $\mathbf{p}_R = (c_R, r_R)$, respectively. Since we assume both images are rectified, $r_L = r_R$. After projecting both images onto the screen, we have the corresponding screen locations $\mathbf{p}_{L_S} = (c_{L_S}, r_{L_S})$ and $\mathbf{p}_{R_S} = (c_{R_S}, r_{R_S})$ (see Figure 3). It is important to note that $\mathbf{p}_{L_S}$ and $\mathbf{p}_{R_S}$ are specified in **pixels**.

When placing the images on the screen two approaches may be taken: a vergent configuration or parallel configuration. Small screens typically use a vergent configuration in which the image centers are placed at the center of the screen. Larger screens commonly use a parallel configuration in which the image centers are offset by the assumed eye interocular. The equations below are the same for both, except where noted. The image disparity is given by $d = (c_R - c_L)$. The screen disparity $d_S = (c_{R_S} - c_{L_S})$ is either equal to d for the vergent configuration or equal to $d_S = d + B_e/S_r$ for the parallel configuration. In both cases, the *perceived* depth Z_e found from the triangle in Figure 3 whose vertices include the point P_e and the viewer's eyes. Equating the base ratios $\frac{d_S}{B_e}$ and the height ratios $\frac{Z_e - S_z}{Z_e}$, we get:

$$Z_e = \frac{B_e S_z}{B_e - d_S S_r}. \tag{1}$$

Similarly perceived X coordinate from the viewer's perspective, X_e, is computed from the right triangle created by dropping the normal from P_e, as depicted in the figure. We equate the the base ratios $\frac{S_r\left(c_{L_S} - \frac{W}{2}\right) - X_e}{\frac{B_e}{2} - X_e}$ and the height ratios $\frac{Z_e - S_z}{Z_e}$, to get:

$$X_e = \frac{Z_e}{S_z}\left[S_r\left(c_{L_S} - \frac{W}{2}\right) - \frac{B_e}{2}\right] + \frac{B_e}{2}. \quad (2)$$

The perceived Y coordinate is computed in a similar manner.

To explain the stereoscopic effects in Table 1, we assume an initial configuration where the camera and eyes have the same field of view and interocular. In this situation, the eyes see exactly what the cameras "see" and there is no distortion, as illustrated in Figure 4(a). Let us call this state the "initial case".

Cardboarding and pinching (field-of-view effects). Changing the field of view stretches the world in the X and Y directions, changing all the parameters in the above equations, making it difficult to depict the perceived behavior simply from the formulae. Instead, to illustrate the effect, we show a simpler example in Figure 4(b). Here, the effect of the projector is ignored and the image created by the camera is directly seen by the viewer's eyes. This causes a flattening effect known as *cardboarding* for narrower field of views, and pinching for wider field of views.

Gigantism and miniaturization (interocular effects). Let us again start with the "initial case", where $B_e = B_c$. Without loss of generality, let us assume we can change the eye baseline instead of the camera baseline, since the change between the two is relative. If we 'increase' the eye baseline B_e, then, in Equation 1, the denominator is increased, reducing the depth Z_e. A more direct relationship decreases X_e in Equation 2. This holds both in the depicted scenario when the ratio $\frac{Z_e}{S_z}$ is greater than 1, as well as when the perceived image is in theater space ($\frac{Z_e}{S_z} \leq 1$) which causes the signs in Equation 1 to reverse. This results in the scene being perceived as more miniaturized or "toy-like", and is termed miniaturization (Figure 4(d)). The opposite effect, called gigantism, occurs when the camera interocular is smaller than the eye interocular, and is shown in Figure 4(e).

Horizontal and vertical viewer motion. The above math can be easily extended for vertical (forward-backward) motion of the viewer, since that implies a new value for S_z. Horizontal (sideways) viewer motion does not change the perceived depth Z_e since the motion is parallel to the screen. It does, however, result in a skew-like distortion of the scene shape due to a change in the x-coordinate X_e. If K_x is the horizontal shift of the viewer, the corrective term $\frac{-K_x(Z_e - S_z)}{S_z}$ is added to X_e in Equation (2).

Perspective distortion. When the viewer rotates their head as they move around the theater space, perspective distortion causes a 'keystone' effect. Although largely ignored in conventional cinema, this would add vertical disparity to stereoscopic content. Previous work has noted that, despite eye-strain, binocular fusion in stereoscopic cinema is typically robust to such changes ([Held and Banks 2008], [Lipton 1982]). We allow the user to decide the amount of tolerable strain by providing the vertical disparity d_V between two corresponding points $\mathbf{p}_{L_S}$ and $\mathbf{p}_{R_S}$, $d_V = (h_L - h_R)$. Assuming the viewer's height is equal to the center of the screen, h_L is given by $(r_{L_S} - \frac{H}{2})\frac{f_e}{Z_{p_L}}$ where H is the image height in pixels. Z_{p_L} is the euclidian distance along the X and Z dimensions between the left eye and $\mathbf{p}_{L_S}$, and is given by $\sqrt{\left(\left(K_x - \frac{B_e \cos\theta}{2}\right) + S_r\left(c_{L_S} - \frac{W}{2}\right)\right)^2 + \left(S_z + \frac{B_e \sin\theta}{2}\right)^2}$. Similar equations exist for h_R and the derivation of these is provided in the appendix.

2.1 User controlled parameters

In our editing interface, we allow the user to change the viewer's perception of the scene by varying four parameters: camera FOV θ_c, camera interocular B_c, horizontal image translation V_c, and dolly Z_s. The horizontal image translation is similar to changing the cameras' angle of vergence. Assuming the cameras are rotated along the y-axis and they are rectified in a specific manner a change in vergence will result in exactly a horizontal image shift.

Changes in the FOV and horizontal image translation require resizing and shifting of the images respectively. However, manipulating the interocular and dollying the camera require the scene to be rerendered. This is because changing the interocular and dollying result in camera translation, which has to account for scene parallax. We now show how the new pixel positions are computed based on the user-specified edited parameters θ_c, V_c, B_c and Z_s. We apply changes in these values in the order corresponding to a cameraman performing the same changes at video capture time: dolly Z_s, interocular B_c, FOV θ_c, and finally the image translation V_c.

While V_c is directly manipulated, the other three parameters are manipulated as ratios of the original camera parameters θ_{c0}, B_{c0}, and Z_{s0}:

$$\tan\left(\frac{\theta_c}{2}\right) = \alpha_\theta \tan\left(\frac{\theta_{c0}}{2}\right),\ B_c = \alpha_B B_{c0},\ \text{ and } Z_s = \alpha_Z Z_{s0}. \tag{3}$$

By definition $V_{c0} = 0$. From Equation (3), α_θ scales the image about its center, α_B is the relative change in camera baseline, and α_Z is the "normalized" dolly using the unit distance Z_{s0}. Z_{s0} is computed as a function of the viewer to screen depth as reprojected in camera space. This is done simply by scaling screen width S_w in the eye diagram on the right of Figure 3 by the ratio $\frac{B_c}{B_e}$. The screen is then aligned with the image plane, on the left of Figure 3. The new 'focal length' then becomes the quantity we desire, and we can write the following relationship:

$$Z_{s0} = \frac{B_{c0} S_w}{2 B_e \tan(\frac{\theta_{c0}}{2})}. \tag{4}$$

Casting user controlled quantities as ratios is useful in scenarios in which camera parameters are hard to quantify or are unknown. If only post-production effects are desired, the camera parameters are not needed. However, to plan a shot the original camera parameters need to be known. Our key assumption in using ratios to represent camera parameters is that by directly manipulating the stereoscopic effect we are indirectly changing the camera parameters that caused it. This is supported by the linearity of Equations 2 and 1 in the four editable parameters. For example, we will scale the scene in a manner inversely proportional to the camera interocular ratio α_B. Therefore, we are addressing gigantism and miniaturization by changing the scene shape, which is equivalent to changing the camera baseline.

We use Equations (1) and (2) to compute the original X_e and Z_e coordinates before any manipulations using the original screen column location c_{L_S} and screen disparity d_S for pixel $\mathbf{p}_{L_S}$. Applying the changes in camera interocular and dolly, we find a new set of 3D perceived coordinates $\bar{X}_e$ and $\bar{Z}_e$:

$$\bar{Z}_e = \frac{Z_e + S_z \alpha_Z - S_z}{\alpha_B}\ ,\ \bar{X}_e = \frac{X_e}{\alpha_B}. \tag{5}$$

Next, we can project the transformed point onto the movie screen to find a new set of screen coordinates $(\bar{c}_{L_S}, \bar{r}_{L_S})$ and screen disparity $\bar{d}_S$:

$$\bar{c}_{L_S} = \frac{\left(2\bar{X}_e S_z + B_e \bar{Z}_e - B_e S_z\right)}{2\bar{Z}_e S_r} + \frac{W}{2}. \tag{6}$$

We can similarly compute the value of $\bar{c}_{R_S}$, after which we can compute the new disparity $\bar{d}_S = \bar{c}_{R_S} - \bar{c}_{L_S}$. We then apply our FOV and horizontal image translation changes to find the new screen coordinates (c'_{L_S}, r'_{L_S}) and warped screen disparity d'_S:

$$\begin{aligned} c'_{L_S} &= \alpha_\theta(\bar{c}_{L_S} - \frac{W}{2}) + \frac{W}{2} - \frac{V_c}{2}, \\ d'_S &= \alpha_\theta \bar{d}_S + V_c\ ,\ \text{ and } c'_{R_S} = c'_{L_S} + d'_S. \end{aligned} \tag{7}$$

Equation (7) assumes a vergent configuration. If a parallel configuration is used the images would have to be additionally shifted in the X

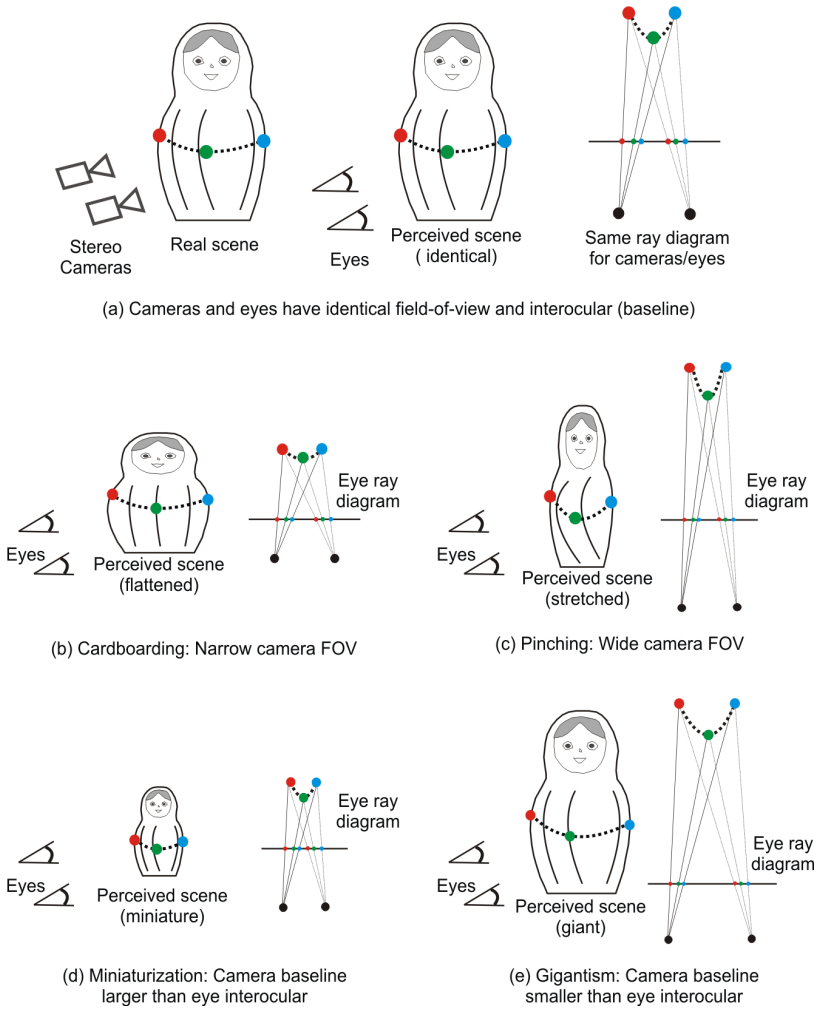


Figure 4: ***Geometric explanations of well-known stereo cinema effects:*** *When the internal parameters of the camera and eyes are the same, as in (a), the perceived scene is identical to the real world. Any difference between cameras and eyes causes distortions. In (b) and (c), changing FOV either flattens the scene ("cardboarding") or elongates it ("pinching"). In (d) and (e), varying baseline scales the scene larger ("gigantism") or smaller ("miniaturization").*

direction (by $\frac{B_e}{2S_r}$) before and after scaling. Given these mathematical preliminaries, we are now ready to describe our interactive stereoscopic editing tool.

3 An interactive stereoscopic editing tool

While a human fuses a stereo pair to perceive depth, a stereo algorithm can reconstruct the scene from the same images. The key contribution of our interface is that the user directly manipulates the shape of the world *as perceived by a viewer*. This is enabled by a top-down, "bird's eye view" of the perceived scene's point cloud, as shown in Figure 6(a). The algorithm of [Zitnick et al. 2004] is used to automatically generate the image disparities and render a new set of stereo images given the edited parameters. (All the stereo examples in this paper were shot using our stereo rig shown in Figure 5. Note that the large rig baseline is only for presentation purposes.)

3.1 Editing operations via box widget

We designed a box widget as part of the interface to allow the user to easily manipulate the perceived shape of the world. As shown in Figure 6(a), the box is overlaid on the perceived scene points. The user manipulates various parts of the box to effect specific changes. *Please*

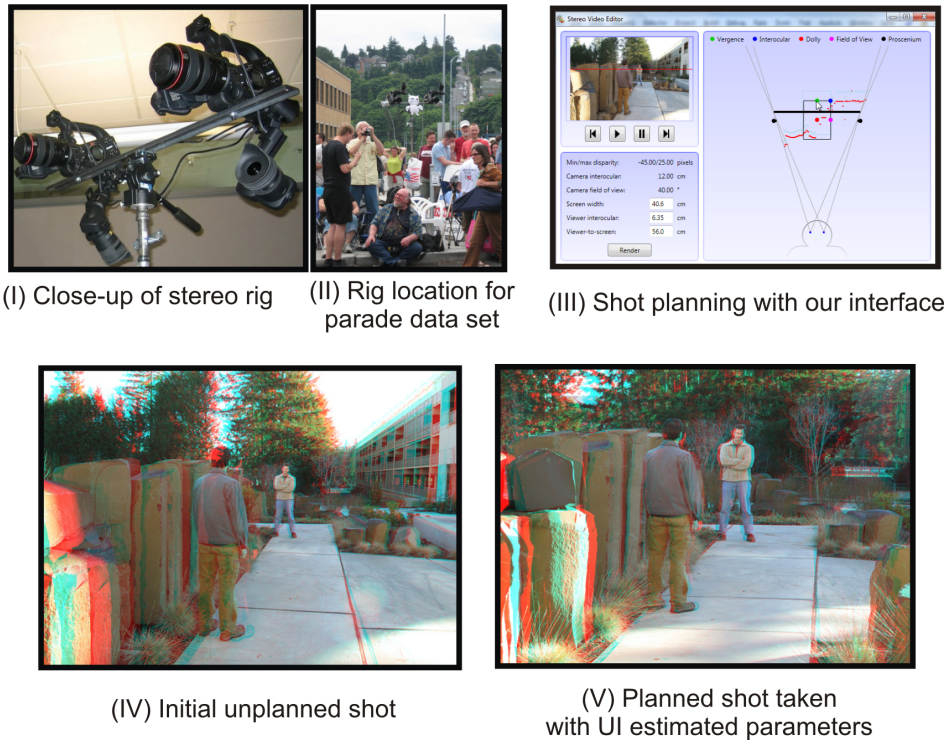


Figure 5: ***Stereo rig setup.*** *(I) The data in our edited movies was collected by two Canon HD XLH1 cameras that were synchronized and placed on a metal stereo rig designed to allow normal pan-tilt motion. (II) Our stereo-rig placed among the audience during the parade. (Note, the large baseline is only for presentation purposes.) (III) UI capture during shot planning. (IV) The shot used for planning. (V) The shot taken given UI parameters.*

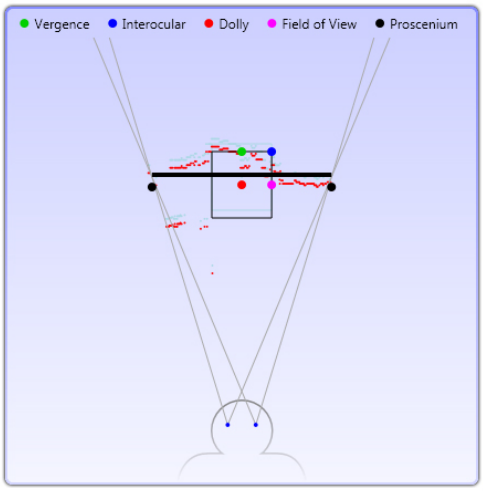
see our accompanying video (StereoEdit.mov) for examples of these effects.

When this box is exactly a square, it signifies there is zero distortion for the viewer. The shape of the box is meaningful because it summarizes the stereo effects present in the rendered images. For example, cardboarding or pinching correspond to a flattening or elongation (respectively) of this square. The user can change the perceived scene shape (and subsequently rerender new stereo images) by manipulating the box in the following ways:

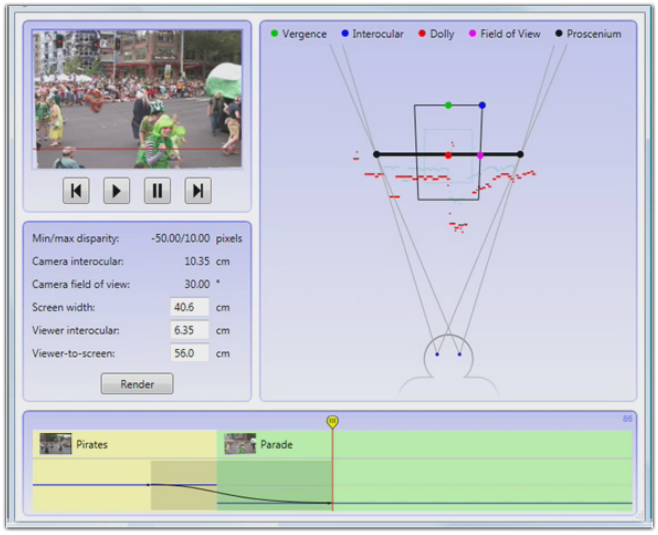
Adding/enhancing cardboarding and pinching effects (parameters affected: FOV). The user can change the field of view (FOV) by dragging the purple dot on the side of the box; this changes the original camera focal length as well. The distortion of the box mirrors the pinching effects that occur with wider fields-of-view. Figure 7(d) shows a scene whose FOV has been digitally altered. Note the foreground tree appears flat in the final image.

Translating images left and right (parameter affected: approximate camera vergence). Recall that the parts of the scene with zero disparity appear to be located on the screen. Changing the horizontal image translation has the effect of changing the parts of the scene that appear to be located at the screen. The user translates the images by moving the green dot at the top of the box up or down. This results in the left and right stereo frames being shifted in the X direction. Note that this action distorts the 3D scene shape non-uniformly. The flag-holder in Figure 1(II)(a) was shifted out of screen space and now appears closer.

Translating the scene forward/backward (parameters affected: dolly). The user dollies (i.e., changes the camera-scene distance) by dragging the red dot in the center of the square. As the scene gets closer to the viewer, the virtual cameras move closer to the scene. The dolly causes no distortions, since it accounts for parallax effects (which are depth dependent). The extent to which we can dolly depends on



(a) User interface for controlling stereo parameters.



(b) User interface for editing scene transitions.

Figure 6: ***The user interface for our stereoscopic editing tool.*** *(a) The key feature of our UI is the "bird's eye" view of the scene. (b) shows the stereo movie timeline, allowing the cross-fade of stereoscopic parameters across scene transitions. User interaction with the point cloud is possible through changes in horizontal image translation, FOV, dolly and interocular. In addition, the perceived screen edge depths can be adjusted using the proscenium arch.*

the quality of the stereo data. Although small shifts are possible, they may result in a large change in the stereo experience, as demonstrated in Figure 1(II)(c).

Scaling the perceived scene size (parameter affected: interocular). By dragging the blue dot on the corner of the box, the user can scale the scene to appear larger or smaller. This effect changes the camera baseline, and is identical to miniaturization and gigantism as described in Section 2. In Figure 7(e), the interocular is decreased, and the figures appear larger than life, relative to the viewer.

Parameter coupling. Our system allows the user to "lock" different camera parameters together to create new stereoscopic effects. One example is the stereoscopic equivalent of the Hitchcock zoom pioneered in the film *Vertigo*. We demonstrate our own form of stereo "Hitchcock" zoom by coupling together the dolly, FOV, and image translation parameters of our UI. If a parallel configuration is used, only dolly and FOV need to be coupled. Figure 7(f) shows the stereo "Hitchcock" zoom applied to a scene in which the size of the foreground girl is stabilized in the image, and her position is stabilized in 3D.

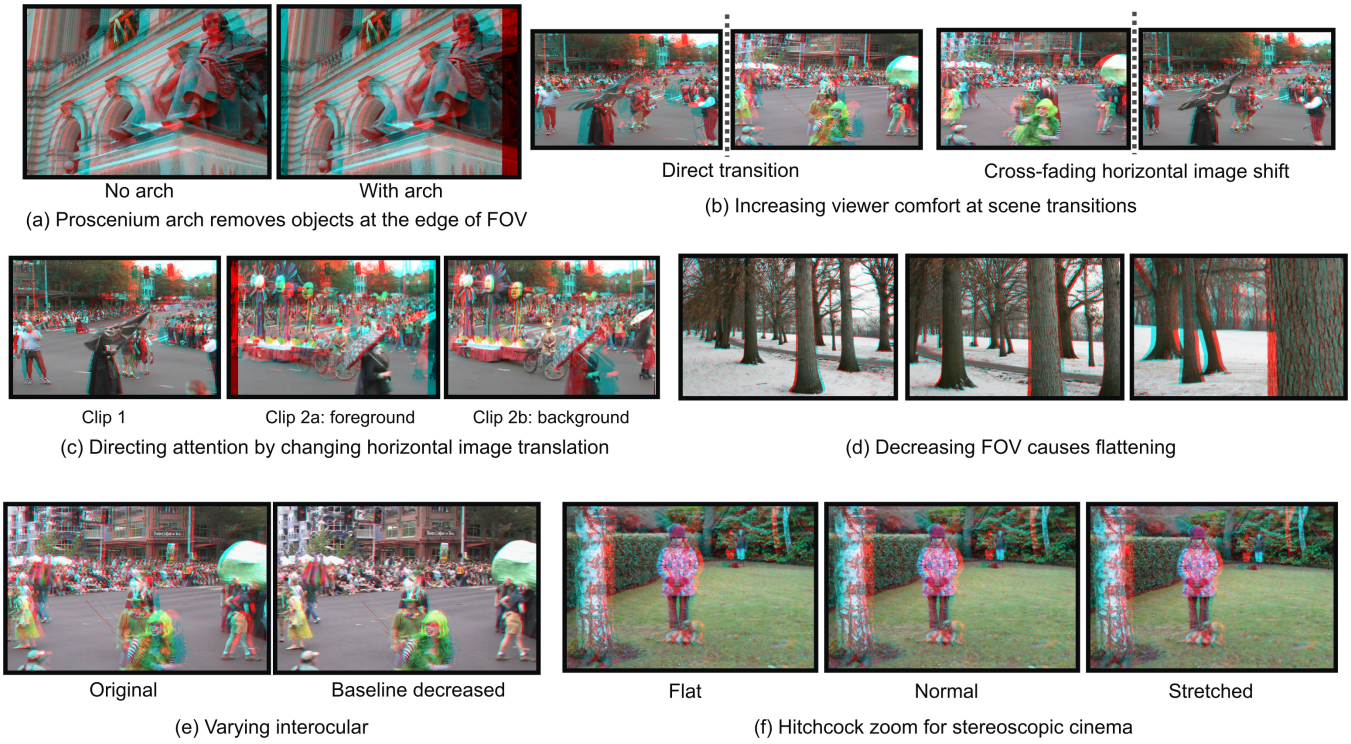


Figure 7: ***Editing effects for stereoscopic movies:*** *In (a) we show a scene without and then with the proscenium arch. In (b) the transition Clip 1 and Clip 2 is eased by crossfading the horizontal image translation to zero at the cut. In (c) we transition between Clip 1 and two versions of Clip 2 with different horizontal image translation. The user's attention is directed to the foreground in Clip 2a and the background in Clip 2b. In (d) we demonstrate scaling a stereo image to change its field-of-view, causing flattening effects. We demonstrate changes in interocular in (e) and our 3D "Hitchcock" effect in (f). Please see the accompanying video (StereoEdit.mov in [Authors 2009]) for more effects.*

Shifting proscenium arch. In many stereoscopic shots with objects appearing in front of the screen, there tend to be regions on the edges of the screen that can be seen by only one eye. These areas appear inconsistent with the scene edges and can cause eye strain. The proscenium arch simply blacks out part of the stereo frame to move the perceived edge of the screen closer to the viewer. Our interface has black markers for both left and right vertical edges of the image. The length of the black markers are adjusted by moving these along the line of sight. In Figure 7(a), we show an image whose edges are aligned. When the proscenium arch is appropriately positioned, it becomes easier to fuse the objects near the image edge, such as the statue.

3.2 Planning a capture session

Shooting a 3D film is difficult precisely because it is challenging to imagine how the audience's experience will differ from the director's vision. Our user-interface can address this problem by providing a way to plan for the shot, given rough takes of the scene and/or still images. In Figure 5, we show a scene consisting of two people in front of a set of trees. We assume this sort of still 'prototype' shot can be taken with little or no effort. We can plug these images into our editor, and, as shown in the figure, we get a bird's eye view of the scene will be perceived. Given this setup, the director may wish to change some aspect of the scene. For example, the director may wish the two people to be perceived, in 3D, closer together in theater space. By adjusting the point cloud in the top-down view of our interface, the user changes the original camera parameters. The desired camera parameters are then output as ratios of the parameters used to generate the rough takes.

The examples we show in this paper are on real imagery, which require computer vision technology to compute depth. A more direct use of our tool would be with the use of synthetic imagery where the depths can be exacted directly from the 3D model. Our stereo editing tool could then be integrated with the geometric modeler so one could make edits to the stereo parameters and changes to the 3D scene in the

same application.

3.3 Creating post-production effects in stereo movies

An important capability in a movie editor is cutting between shots. This is because many times the story is told by switching between contrasting scenes [Arijon 1976]. Recent trends in film and TV have tended towards both multiple cuts a minute, as well as many cuts per second. The former are now common in prologues of crime dramas such as *CSI* and *24*. For stereoscopic content the potential for visual discomfort in these cases is large, since there is a lag time in fusing scenes at differing depths. One method for mitigating this issue is to blend the horizontal image translation during a cut so the objects of interest have the same depth at the cut. The subtle shifting of image translation before and after the cut can be done without being noticed by the audience.

Figure 7(b) summarizes such a cut. In the first scene, the flag is shot with negative disparity and appears to be behind the screen. In the next clip, the green-haired girl appears to be in front of the screen, resulting in a jarring "jump" as the viewer quickly re-adjusts. Using our editor, the user can select an object in the previous and next clip and select a disparity as demonstrated in Figure 6(b). The user can choose to have the global disparities blended before and after the cut, so that at the cut the two objects have identical disparities. This results in a more visually pleasing transition.

Directing attention. The next application of horizontal image translation exploits its two properties, namely: (a) Global disparity changes through image shifts are usually not noticed, and (b) full image fusion occurs after a short time lag. If the scene cuts back and forth faster than this time lag (which may vary from person to person) then, we hypothesize that objects with similar disparities to the area currently fused are fused first. Therefore, it seems possible to direct the audience's attention, as in Figure 7(c). Anecdotal evidence from a small set of users confirmed that by alternatively adjusting the areas of similar disparity using horizontal image translation, we were able to shift the audience attention across the scene.

3.4 A discussion on usability

Our interface is driven by two novel innovations in stereoscopic editing. The first is the bird's eye view, which offers the editor the choice of working without stereoscopic glasses. This is possible since we show what the audience will perceive in theater space. The editor can try out different theater dimensions, and even change the interocular and viewer position to see exactly what a specific viewer would perceive. Complementing the bird's eye view, any of the rendered scenes can be also quickly viewed in stereoscopic format if the editor so wishes. The second innovation is the presence of a unit box that allows an intuitive depiction of stereoscopic distortions, without requiring the user to understand the intricacies of projective geometry. With these two tools, our editing interface allows the creation of new effects, such as the hitchcock effect for stereoscopy. In addition, beyond these innovations, the other building blocks of our interface were created from widely-accepted technologies; for example, the time-line showing the current clip in our UI has been used in many successful editing softwares, such as *Adobe Premiere* and *Apple Final Cut Pro*. Furthermore, our editing tool allows real-time manipulation of the clips. While rendering times depend on the back-end used, rough cuts with low resolution are available within a few seconds. Therefore we believe our editing tool does in fact ease the editing process for stereoscopic film makers. However, we leave a user study on the usability of our interface for future work. Instead, in the next section, we explore the more fundamental question of whether the effects we create are perceived by users and what impact does the rendering of frames have on quality.

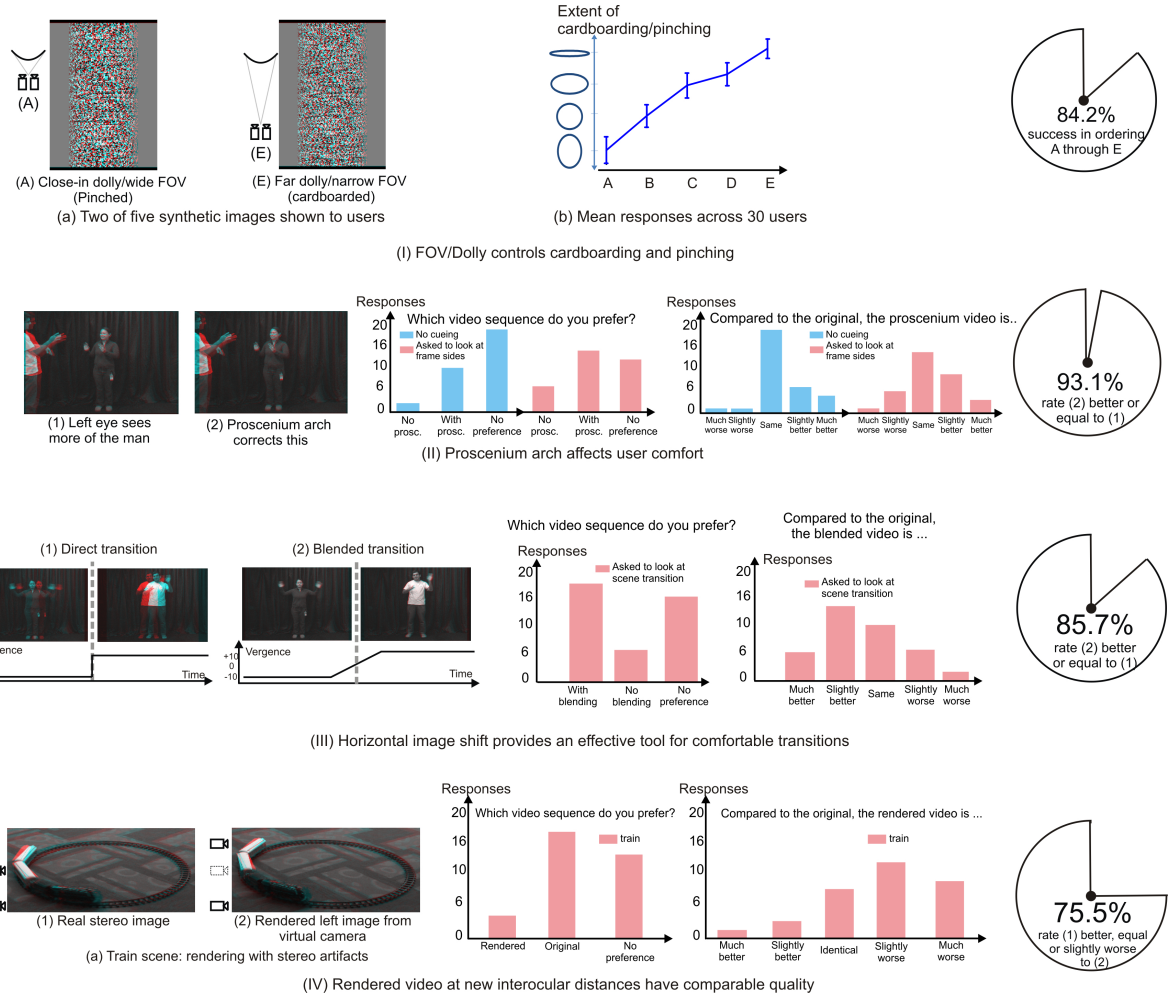


Figure 8: *Overview of and results for stereoscopic user studies associated with (I) Dolly/Field-of-view, (II) Proscenium arch, (III) Horizontal image translation, and (IV) Interocular distance. In (I) we asked users to rank the 'cardboarding/pinching' of five synthetic images by matching icons shown on the Y-axis of I(b). This showed cardboarding/pinching effects occurred when field-of-view and dolly were varied. In (II) we asked users to compare videos with and without proscenium, demonstrating that it positively affected the experience. In (III) users compared a scene cut with and without the image translation cross-fade, showing that it eased the scene transition. In (IV) users rated videos with a rendered left video stream comparably good as or only slightly worse than real videos. This is positive since, in practice, real videos of new viewpoints will not be available for film-makers after shooting has taken place. Please see UserStudies.mov at [Authors 2009] for more details.*

4 Perceptual studies for the stereoscopic geometry model

An editing tool based on geometry can digitally alter stereoscopic film, but predicting the perception effects as a result of these changes is challenging. We performed a series of user studies to validate our ideas of how the audience reacts to certain types of editing. We studied the four important parameters described in Section 3 (dolly/FOV, proscenium arch, horizontal image translation, interocular) under controlled scenarios. We conducted the studies on 32 people, including men (23) and women (9) from a variety of cultural backgrounds.

Our experiments began with a stereo-blindness test and two users who had this condition were not included in the analysis. Our multiple-choice questions contained a mix of binary (yes-no) and rating (best-to-worst) answers. The order of the experiments was randomly permuted to reduce bias. Short stereo movie clips (no audio) were shown in color on a Hyundai P240W 3D monitor to users wearing polarized glasses. (*These clips are shown in UserStudies.mov at [Authors 2009]*) In Figure 8, we display some example frames in anaglyph format. We have performed chi-squared statistical significance tests on our studies. We assumed uniform random distribution (33% each for yes-no-maybe questions and 20% for Likert-scale questions). All our experiments scored with a p-value ≤ 0.01.

The following questions about four stereoscopic parameters were investigated:

Does changing camera FOV cause cardboarding and pinching? Five synthetic stereo images of a cylinder were shown (with varying distances between the camera and cylinder, Figure 8 (I)). We keep the object's image size constant by appropriately adjusting the focal length. The subjects were asked to order these cylinders by cross-section, selecting from icons shown in Figure 8(I)b. 84.2% of the users' rankings agreed with the predictions of the model linking cardboarding and FOV. The mean values of their responses closely matched the actual cross-section order.

Does the proscenium arch improve viewing comfort? Here, the users were shown videos of two people in conversation (Figure 8(II)). In the raw video, the right camera sees more of the man than the left. In the edited video, the proscenium arch was adjusted to make the left and right views more even. The experiment had two stages: first the subjects rated the two videos. 93.1% of the subjects did not prefer the raw video. The subjects were then told to pay attention to the screen edges and again compared the videos. As expected, a similar fraction of the subjects (93.3) preferred the video with the proscenium arch. These effects are shown visually by the significant shift right-wards in both graphs in Figure 8(II).

Does cross-fading image translations ease scene transitions? In this experiment, we show two videos in succession. The first is of a woman waving, with negative disparity (in front of the screen). The second is of a man waving, with positive disparity (behind the screen). In Figure 8(III)(1), the transition is abrupt. In Figure 8(III)(2) we shift the left and right frames before and after the cut so that the depths of the two people are at zero disparity at the "cut". We then asked them to specifically rate the scene transition. From the first graph in the figure, most people either have no preference or prefer the blended video. Supporting this, 85.7% of the subjects preferred the edited transition in the second Likert scale graph.

Does rendering with a new interocular affect quality? A stereoscopic editor may render images by changing the interocular using a stereo reconstruction of the scene. How would the result compare to the original, if it was available? In Figure 8(IV), we compare a pair of such videos for a toy train moving on a circular track. This scene was chosen since it is repeatable and the stereo reconstruction has artifacts. The first graph shows that if the original was available, it would be preferred (although many people have no preference). However, a finer granular questioning in the Likert scale shows that most subjects (75.5%) chose the middle three ratings (slightly better/worse, identical). We should make clear here that the reason we included the "slightly worse" category for this question is because, in practice, shooting the same scene at many different interocular distances is expensive and inconvenient. Therefore, this is a positive result since, in post-production, the actual footage with edited baselines will not be available. We attempted to remove stereo artifacts through blurring the rendered view and found qualitative evidence that some users could not detect any differences, supporting similar previous work [Stelmach et al. 2000].

Although there are strong trends exhibited in our user studies, we must provide some caveats. For example, the subjects may not represent the general stereo movie audience and therefore the results might not accurately reflect how people experience stereo movies. This is because the duration of each clip is short, and in some cases, the subjects were cued (e.g., in the latter part of the proscenium arch experiment, subjects are asked to notice the sides). However, these studies suggest certain perceived effects can be reasonably predicted, and support our goals of building a stereo editor.

5 Concluding remarks

In this paper, we described a foundation for creating a comprehensive digital editing tool for stereoscopic cinema. Our main contribution is an intuitive easy-to-use editing interface that is *viewer-centric*. This allows the editor to specify how the scene is *intended* to be experienced by the audience, as opposed to manipulating camera parameters and visualizing camera-centric outputs. Our mathematical framework computes

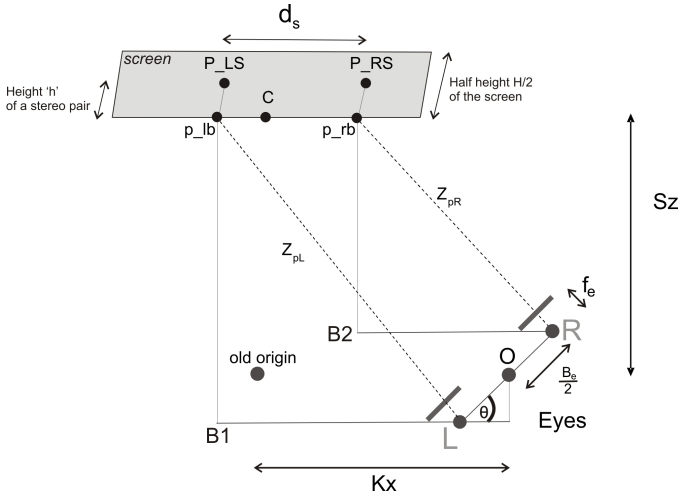


Figure 9: ***Perspective distortion:*** *Except for points P_{LS} and P_{RS} all other points and lines are on the 'middle' plane parallel to the ground. The viewer has moved horizontally by K_x and tilted her head by θ to the x-axis. We project the height h of P_{LS} onto the left eye's retina and the height h of P_{RS} onto the right eye's retina. The difference of the two projected heights will give us the maximum vertical disparity.*

new camera parameters for use in shot planning, and allows for post-production manipulations even if camera parameters are not known. Finally, we answer several questions about 3D scene perception through human studies.

A Appendix: Viewer's Perspective Distortion

First let use show that if we knew Z_{pL} and Z_{pR} then we could get the maximum vertical disparity. The height h of P_{LS} is given by $S_r(r_{L_S} - \frac{H}{2})$ and is identical (due to rectification) for P_{RS}. The projected height of these screen images of P (P_{LS} and P_{RS}) on the retina of the left and right eye respectively are given by h_{left} and h_{right}. We scale these by S_r:

$$h_{left} = \frac{1}{S_r}\left(S_r(r_{L_S} - \frac{H}{2})\right)\frac{f_e}{Z_{pL}} = \left(r_{L_S} - \frac{H}{2}\right)\frac{f_e}{Z_{pL}} \tag{8}$$

and similarly,

$$h_{right} = \left(r_{R_S} - \frac{H}{2}\right)\frac{f_e}{Z_{pR}} \tag{9}$$

We denote the distortion d_v as the difference between the two heights, which should be 0 in the case of no distortion:

$$d_V = h_{left} - h_{right} \tag{10}$$

A.1 Obtaining Z_{pL} and Z_{pR}

Recall from the paper that the x-coordinates of P_{LS} and P_{RS} are given by c_{L_S} and c_{R_S} respectively. Now first consider from Figure 9 the right angle triangle $(p_{lb}, L, B1)$. The base of this triangle is given by $(B1, L) = \left(K_x - \frac{B_e}{2}\cos\theta\right) + S_r\left(c_{L_S} - \frac{W}{2}\right)$. The height of this triangle $(B1, p_{lb}) = S_z + \frac{B_e}{2}\sin\theta$.

Therefore

$$Z_{pL} = \sqrt{\left(\left(K_x - \frac{B_e}{2}\cos\theta\right) + S_r(c_{L_S} - \frac{W}{2})\right)^2 + \left(S_z + \frac{B_e}{2}\sin\theta\right)^2}. \quad (11)$$

Now consider from Figure 9 the right angle triangle $(p_{rb}, R, B2)$. The base of this triangle is $(B2, R) = \left(K_x + \frac{B_e}{2}\cos\theta\right) - S_r\left(c_{R_S} - \frac{W}{2}\right)$. The height of this triangle is $(B1, p_{rb}) = S_z - \frac{B_e}{2}\sin\theta$. Therefore

$$Z_{pR} = \sqrt{\left(\left(K_x + \frac{B_e}{2}\cos\theta\right) - S_r\left(c_{R_S} - \frac{W}{2}\right)\right)^2 + (S_z - \frac{B_e}{2}\sin\theta)^2} \quad (12)$$

References

ARIJON, D. 1976. *Grammar of the Film Language*. Focal Press.

AUTHORS, T. 2009. Supplementary videos. *http://www.koppal.com/SupplementaryVideos.zip*.

BEATON, R. J. 1990. Displaying information in depth. In *SID Digest*.

BURDEA, G. C., AND COFFET, P. 2003. *Virtual Reality Technology*. Wiley-IEEE Press.

BUSER, P., AND IMBERT, M. 1992. *Vision*. MIT Press.

CLARK, B. 2007. The 10 commandments of 3D cinematography. *http://forums.digitalcinemasociety.org/showthread.php?t=44*.

DIRECT, D. G. 2009. Suggested store for anaglyph glasses. *http://www.3dglassesdirect.com*.

EMPEY, M., AND NEUMAN, R. 2008. Stereoscopic depth as a storytelling tool. *Disney Animation presentation at FMX*.

GRINBERG, V., PODNAR, G., AND SIEGEL, M. 1994. Geometry of binocular imaging. In *Proc. SPIE (Stereoscopic Displays and Virtual Reality Systems)*, vol. 2177, 56–65.

HELD, R., AND BANKS, M. 2008. Misperceptions in stereoscopic displays: A vision science perspective. *Applied Perception in Graphics and Visualization*.

ISONO, H., YASUDA, M., TAKEMORI, D., KANAYAMA, H., YAMADA, C., AND CHIBA, K. 1992. 50-inch autostereoscopic full-color 3D TV display system. *Proc. SPIE (Stereoscopic Displays and Applications)*.

JONES, A., MCDOWALL, I., YAMADA, H., BOLAS, M., AND DEBEVEC, P. 2007. Rendering for an interactive 360° light field display. *ACM Transactions on Graphics and SIGGRAPH 26*, 3 (July), article 40.

JULESZ, B. 1971. *Foundations of Cyclopean Perception*. University of Chicago Press.

KAWAI, T., SHIBATA, T., INOUE, T., SAKAGUCHI, T., OKABE, K., AND KUNO, Y. 2002. Development of software for editing of stereoscopic 3-D movies. In *Proc. SPIE (Stereoscopic Displays and Virtual Reality Systems IX)*, vol. 4660, 58–65.

KIM, H., CHOI, J., CHANG, A., AND YU, K. 2008. Reconstruction of stereoscopic imagery for visual comfort. In *Proc. SPIE (Stereoscopic displays and applications XIX)*, vol. 6803.

LAMBOOIJ, M., IJSSELSTEIJN, W., AND HEYNDERICKX, I. 2007. Visual discomfort in stereoscopic and autostereoscopic displays: a review of concepts, measurement methods, and empirical results. In *Proc. SPIE (Stereoscopic displays and applications XVIII)*.

LEVONIAN, E. 1954. Stereoscopic cinematography: Its analysis with respect to the transmission of the visual image. *MA thesis, University of Southern California*.

LIPTON, L. 1982. *Foundations of the stereoscopic cinema*. Van Nostrand Reinhold.

LOOP, C., AND ZHANG, Z. 1999. Computing rectifying homographies for stereo vision. *Proc. of Computer Vision and Pattern Recognition*.

MASAOKA, K., HANAZATO, A., EMOTO, M., YAMANOUE, H., NOJIRI, Y., AND OKANO, F. 2006. Spatial distortion prediction system for stereoscopic images. *Journal of Electronic Imaging 15*, 1 (March).

MATUSIK, W., AND PFISTER, H. 2004. 3D TV: A scalable system for real-time acquisition, transmission and autostereoscopic display of dynamic scenes. *ACM Transactions on Graphics and SIGGRAPH 23*, 3 (August), 814–824.

MCKAY, H. C. 1953. *Three-Dimensional Photography - Principles of Stereoscopy*. http://www.3d.curtin.edu.au/cgi-bin/library/mckay.cgi.

MUELLER, R., WARD, C., AND HUSAK, M. 2008. A systematized wysiwyg pipeline for digital stereoscopic 3d filmmaking. *SPIE*.

NAGATA, S. 1991. How to reinforce perception of depth in single two-dimensional pictures. In *Pictorial Communication in Virtual and Real Environments*, 527–545.

PFAUTZ, J. D. 2000. Depth perception in computer graphics. *Ph.D. thesis, University of Cambridge*.

RHEINGOLD, H. 1992. *Virtual Reality*. Simon and Schuster.

ROBINETT, W., AND ROLLAND, J. 1991. A computational model for the stereoscopic optics of a head-mounted display. In *Proc. SPIE (Stereoscopic Display and Applications II)*, vol. 1457, 140–160.

SAMMONS, E. 1992. *The world of 3D movies*. http://www.3d.curtin.edu.au/cgi-bin/library/sammons.cgi.

SCHWERDTNER, A., AND HEIDRICH, H. 1998. The Dresden 3D display. In *Proc. SPIE (Stereoscopic displays and virtual reality systems V)*, vol. 3295, 203–210.

SIEGEL, M., AND NAGATA, S. 2000. Just enough reality: comfortable 3-D viewing via microstereopsis. *IEEE Circuits and Systems for Video Technology 10*, 1 (April), 387–396.

STELMACH, L., TAM, W., MEEGAN, D., VINCENT, A., AND CORRIVEAU, P. 2000. Human perception of mismatched stereoscopic. *International Conference on Image Processing*.

STREATHER, P. 2007. Response: The 10 commandments of 3D cinematography. *http://forums.digitalcinemasociety.org/showthread.php?t=44*.

SUTO, M. 2006. Open source stereo movie maker. *http://stereo.jpn.org/eng/stvmkr/*.

TESSMAN, T. 1990. Perspectives on stereo. In *Proc. SPIE (Stereo Displays and Applications)*, vol. 1256, 22–27.

TRAUB, A. C. 1967. Stereoscopic display using rapid varifocal mirror oscillations. *Applied Optics 6*, 6, 1085–1087.

VALYUS, N. A. 1962. *Stereoscopy*. Focal Press, London.

WALD, G. 1950. Eye and camera. *Scientific American 183* (August), 32–41.

WANG, C., AND SAWCHUK, A. 2008. Disparity manipulation for stereo images and video. *SPIE*.

WHEATSTONE, C. 1838. On some remarkable, and hitherto unobserved, phenomenon of binocular vision. *Philosophical Transactions of the Royal Society of London 128*, 371–394.

WOODS, A., DOCHERTY, T., AND KOCH, R. 1993. Image distortions in stereoscopic video systems. In *Proc. SPIE (Stereoscopic Displays and Applications IV)*, vol. 1915, 36–48.

ZITNICK, C., KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2004. High-quality video view interpolation using a layered representation. *ACM Transactions on Graphics and SIGGRAPH 23*, 3 (August), 600–608.