Stereo Rendering: An Overview

(Technical Report CS-2005-02)

Jamie Brettle
Natural Phenomena Simulation Group
School of Computer Science
University of Waterloo

Gladimir V. G. Baranoski*
Natural Phenomena Simulation Group
School of Computer Science
University of Waterloo

Abstract

Humans visually perceive the world using a variety of depth cues. Stereopsis is an example of such a depth cue, and it is directly associated with the human binocular visual system. Different computer graphics approaches can be used to take advantage of stereopsis in order to enable a true 3D perception of virtual environments. In this report, we briefly review computer graphics hardware and software tools designed for this purpose as well as optimization techniques that can be applied to speed up the synthesis of stereo images.

1 Introduction

An important aspect of computer graphics is the ability to create the illusion of image depth to users, despite the typical limitations of projecting images on a flat computer screen. Humans use a variety of depth cues from their environments that allow for 3D scenes to be interpreted and processed by the brain. Depth cues individually provide information about a given environment, and are typically added to provide humans an account of a 3D space. These depth cues can be subdivided into physiological and psychological depending on how the information is processed by the viewer, either by specialized biological systems or inferences derived from the brain respectively. Computer graphics can capitalize on both physiological and psychological depth information. This report will be concerned with physiological depth cues, primarily stereopsis, and how this can be leveraged both in computer software and hardware to present viewers images that appear to be in 3D. Stereopsis is the term that refers to the combination of images presented to both eyes to achieve a 3D perception of the world. According to Steinman *et al.* [2000], this is the only depth cue that the human brain can take direct advantage of in order to collect spatial information about the environment [Steinman et al. 2000].

The goal of this report is to investigate how humans use stereopsis to perceive 3D space, and how computers can be used to generate graphics that take advantage of this physiological response in viewing objects with depth. The remainder of this report is organized as follows. Section 2 discusses how humans utilize stereopsis to gain depth information about their environment. Section 3 investigates hardware and software tools that can be used to generate images that take advantage of stereopsis to give the impression of depth to viewers. Section 4 addresses practical issues related to stereo computer graphics applications. Finally, the report concludes with a summary of the main aspects of this introductory study on stereo rendering.

2 Physiological Background

The scientific information presented in this section was originally provided by Steinman *et al.* [2000], and it is inserted in this report for completeness.

2.1 Depth Perception

Stereoscopic vision allows the estimation of the depth and distance that an object is relative to the observer. Having a vision system that has two eyes (binocular) provides many advantages over a monocular vision system despite the increased overhead in coordinating the images captured by two eyes into a single vision [Steinman et al. 2000]. Binocular vision provides greater accuracy in determining depth information, allowing objects to be picked and manipulated in a 3D space, and the navigation through complex environments. This accuracy is important not only for real physical environments, but also for computer generated ones [Steinman et al. 2000].

When a human focuses his vision on a particular object, the object is placed identically onto each retina. The object projected onto each retina is said to be located in corresponding retinal regions. Retinal data is then passed to the visual cortex via the optic nerve. The visual cortex can then process these two images and combine (fuse) them into a single image [Steinman et al. 2000].

Figure 1 illustrates important regions when describing how environments are perceived using stereopsis. The object that the viewer is focusing his eyes upon is referred to as object **a** in Figure 1. The plane that object **a** resides in is called the horopter or stereo plane. Object **a** is projected onto identical regions of the retina of the left and right eyes, resulting in a single combined (fused) image. This projection is labeled as $\mathbf{a}_l$ and $\mathbf{a}_r$ for the left and right eyes respectively.

---

*gvgbaran@curumin.math.uwaterloo.ca

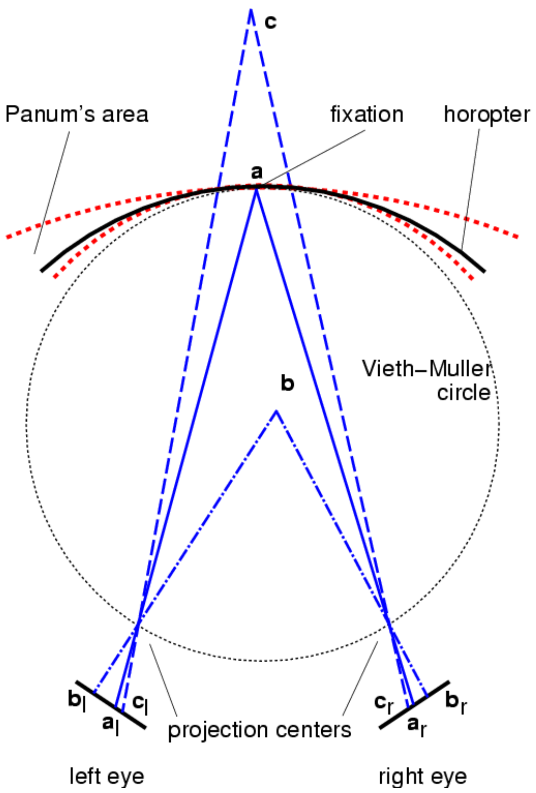

Figure 1: Object positions involved with stereopsis (redrawn from [Steinman et al. 2000]).The Vieth-Muller circle encapsulates the theoretical set of points in space that stimulate corresponding retinal points.

When focusing on an object, if another object is within the viewer's sight, the latter will be projected onto non-corresponding regions of the retina. If the non-corresponding regions are located close enough[1] together, the brain will be able to combine the images into a single image, thus allowing the user to make inferences with respect to depth information based on how far apart the non-corresponding regions are located.

In Figure 1, a viewer is focusing on object **a**, and object **b** is still in the viewer's sight. Object **b** is projected on points $\mathbf{b}_l$ and $\mathbf{b}_r$ on the retina of the left and right eyes respectively. These points are non-corresponding regions on the retina. How the viewer will perceive the object will be dependent on the proximity of the non-corresponding regions.

This region between the inner and outer limits of the zone of a single binocular vision (in front and behind the focused object) is referred to as Panum's (fusional) area. The exact area that formulates this region is directly proportional to the distance that the object lies from the viewer. Objects that lie in front or behind the Panum's area will be projected onto non corresponding retinal regions. However the non-corresponding regions will be too distant on each retina, and therefore cannot be combined together. The end result will be the image being seen twice by the viewer.

The direction of the depth relative to the object that is being focused upon is determined by the disparity of the images projected onto the eyes. When an object is in front of the viewer's fixation point (object that the viewer is focusing his eyes upon), the difference in retinal locations is called crossed disparity. Object **b** in Figure 1 is said to have crossed disparity to the viewer. Alternatively, displacement behind the fixation point is called uncrossed disparity, which is represent by object **c** in Figure 1. As long as objects are within the Panum's area of the fixated object, the brain will combine the images together, and it will be able to determine the relative distance information, thus providing true environmental depth information.

Another key aspect of stereo vision is parallax. Parallax is the apparent shifting of an object when viewed at different angles. This pertains to stereo vision as the left and right eye view the same object at different angles, and thus present the object shifted from each perspective. Parallax gives humans the capability to perceive depth and determine the relative distance between objects. Since eyes are only separated horizontally, images presented to the brain will only have horizontal parallax and not vertical parallax.

---

[1] When referring to close enough with respect to the object that is not located in corresponding retinal locations, the object that is being focused upon determines this distance [Steinman et al. 2000].

2.2 Binocular Summation

The visual cortex has cells specific to process binocular images called binocular cells. Binocular cells will increase firing when a stimulus is placed into the visual field of one of the eyes. A particular binocular cell will fire if the object is placed into either eye's visual field. However, if it is placed in both visual fields , then the binocular cell will fire more vigorously. This increase in firing is what results in a binocular summation and thus stereopsis [Steinman et al. 2000].

In addition to binocular cells, there are cells in the visual cortex which are designed to detect disparity in images. As mentioned previously, images that are not being focused on will have either crossed or uncrossed disparity. These cells can detect these disparities, and fire specifically according to the disparity type [Steinman et al. 2000].

3 Software and Hardware Methodologies

When two images of a scene are created, one for each eye, they are referred to as a stereo pair. In order to view stereo pairs that have been rendered in software, certain hardware requirements need to be achieved. There are two methods that allow stereo pairs to be presented to viewers, namely time-multiplexed and time parallel stereo rendering.

3.1 Time-Multiplexed Stereo Rendering

For time-multiplexed rendering, hardware that features fast refresh rates is required. Once the image has been rendered from the perspective of each eye, each image is alternately drawn on each refresh of the monitor. For example, a monitor that refreshes at 100Hz will effectively be refreshed at 50Hz for each eye, with the left eye receiving odd number refreshes, and the right eye receiving even number refreshes [Harrison et al. 1997].

In order to separate the two images displayed on a single screen, additional hardware is required in order for the user to view the correct image for each eye. The most common method of performing this alternating view restriction is for a user to wear shutter glasses. These shutter glasses effectively block one eye, while leaving the other one open to view the image displayed on screen [Harrison et al. 1997]. This process needs to be synchronized with the output from the display to ensure the viewer is seeing the correct disparate image in each eye.

An additional method to view time-multiplexed displays is to use a polarized[2] liquid crystal display (LCD) shutter. This shutter is placed over the monitor displaying the images, and users wear passive polarizing glasses. Although this method is less expensive than using shutter glasses, displays are limited by the size of the polarized LCD.

Devices to display time-multiplexed stereo can include cathode-ray tube (CRT) and LCD displays as well as projector units casting images onto a surface. Using CRT or LCD displays is a cost effective solution, but limits the size of the images viewed and the number of viewers that can comfortably view a given scene. Projectors allow large sized images for multiple viewers, but come at increased cost relative to desktop display technologies.

Currently the most immersive approach in order to display time-multiplexed stereo graphics consists in a system called CAVE, which is a recursive acronym that stands for CAVE Automatic Virtual Environment [Cruz-Neira et al. 1992]. It is a system that was originally developed at the University of Chicago. It takes stereo images produced by SGI (Silicon Graphics Inc.) hardware and software configurations, and uses projectors to display stereo images on walls and floors of a small room where viewers are situated within. Once the glasses have been synced to the output of the projectors, the users are required to wear shutter glasses in order to provide the correct image to each eye. Figure 2 illustrates how walls and projectors are positioned relative to users in order to provide immersive stereopsis.

Motion trackers can also be added to the CAVE to monitor the movements of viewers and allow the users to interact with the environment. The requirements for CAVE technology are projectors, screens, high performance computers, such as SGI Onyx2, and shutter glasses [Cruz-Neira et al. 1992]. Despite the high hardware and software requirements, the CAVE system provides the most credible stereo environment, and it is capable of supporting multiple users. Motivations behind CAVE systems are typically scientific visualizations [K.W. Brodlie and Wood 2003; R.J. Hubbold and Moore 1998].

When designing a CAVE research environment, the same concerns regarding shutter glasses are still valid. Proper synchronization with the projection units, specialized graphics hardware, and shutter glasses are all required in order to produce a comfortable stereo experience to the user.

3.2 Time Parallel Stereo Rendering

An alternative method of stereo output is to use time parallel display technologies. In a time parallel display, two disparate images are displayed simultaneously to the user, much like real environments are viewed. The viewer will see disparate left and right images of the environment, simultaneously to each eye, thus allowing the visual cortex to properly fuse the seen image [Harrison et al. 1997].

---

[2] Polarization of light refers to having the electrical portion of the light waves moving in a single direction rather than in random directions [Overhem and Wagner 1982].

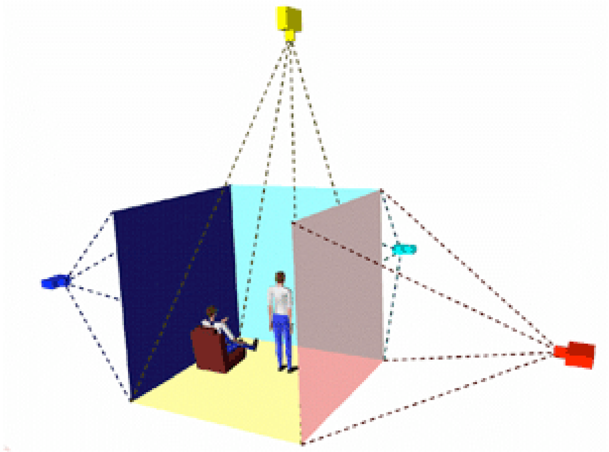

Figure 2: Depiction of a CAVE environment (redrawn from [Cruz-Neira et al. 1992]).

The most basic type of time parallel rendering consists in using anaglyphs. Anaglyphs work by using two complementary colors, red and blue for example. Using the two complementary colors to render an image, two images of a single object are drawn with slight disparity simultaneously. The viewer uses filtered glasses that feature the complementary color on each eye to produce the stereopsis. This method is ideal for quick rendering and low cost needs, but has the disadvantage of distorting the colors of the resulting stereo images.

Stereo pairs can be rendered simultaneously by dividing a monitor into two distinct viewports. In each viewport, left and right disparate stereo images are rendered. Using a mechanical device to ensure that the viewer sees the left and right viewports in correct eyes of the viewer [Harrison et al. 1997]. This is the same principle that View-Master viewers from Fischer-Price use to create stereo images from stereo photography [Harrison et al. 1997]. Dual viewports are advantageous from a hardware cost perspective. However, rendering time would need to be optimized for displaying simultaneous images. Also, this method limits one user per display in order to achieve stereopsis.

Head mounted devices are also a prevalent display method for time parallel stereo rendering. Similar to the previous method of creating two viewports on a single display, this method creates two viewports and renders each to a unique display. Each display is located within a device which is placed directly in front of the viewer's eyes, allowing left and right images to be simultaneous displayed to the left and right eyes respectively.

Head mounted devices can be LCD or CRT based technology. Important features to note when investing in a head mounted device are the resolution of the screens as well as the refresh rates that are available. High refresh rates are important for the quality of the images perceived by the viewer. Additional factors that should be investigated for head mounted devices are weight, and whether or not the display can send data to the image-rendering computer based on the tilt and pitch of the viewers head. This ability is important in interactive realistic applications where the image will need to be updated in real time based on the movement of the viewer's head. Limiting factors to head mounted devices are the cost of each unit, their durability, and the fact that only a single viewer can use at a time.

Dimension Technology[3] has developed a new type of time parallel display technology. The Dimension Technology 3D display is accomplished by using special illumination patterns and optics behind the LCD screen in order to make alternating columns of pixels visible to left and right eyes. The image is rendered into a stereo pair, one image for the left eye, and one for the right eye. The stereo pair is then rendered to the screen simultaneously, with the columns of the images rendered in an alternating fashion on the screen. After the images are drawn on the screen, an illuminating plate located behind the LCD helps to project the alternating columns of pixels to a given eye, left or right.

Using the Dimension Technology 3D screen, the viewer sitting in the correct location will have the stereo left and right images projected on the left and right eyes respectively, giving the appearance of a true 3D image. This method requires specific monitors produced by Dimension Technology and software packages that will properly render stereo pairs into the appropriate screen region. This technology is limited in terms of the number of users that can simultaneously view a stereo image since each display has particular locations that users must view the screen in order to experience the stereopsis.

4 Practical Issues

In this section, we address practical issues related to the rendering of stereo images.

---

[3] The information in this section is available at *www.dti3d.com* which is supported by Dimension Technology Inc.

4.1 How to Generate Stereo Pairs

When creating stereo pairs using software tools, it is important that the pairs follow the same rules that humans use to process real environments. This allows the brain to process the images as if they were representations of real environments.

One of the most important aspects in generating believable stereo pairs is that the images for each eye should have no vertical parallax [Harrison et al. 1997]. Each eye is assumed to be at the exact same vertical level for a viewer, thus there should be no difference in vertical displacements between the same objects in each image. The only thing that should change is horizontal parallax.

In creating scenes, the object of interest should be placed at the stereo plane. Objects that are closer to the viewer than the stereo plane should be placed towards the center of the image, and objects beyond the stereo plane should be place further to the edges of the image.

Keeping color and brightness properties the same across the scene is important, and it is an active area of research in stereo rendering techniques [Harrison et al. 1997]. Colors of objects should be kept similar. However, the disparity caused by the distance between eyes may alter the results of the shading and lighting calculations. Finally, interoccular distance must be realistically set. Generally, 5cm is considered an appropriate value [Harrison et al. 1997].

The general algorithm for generating an image suitable for stereopsis consists in rendering the scene from two vantage points. Each rendered image represents a view for the left and right eye respectively, and takes into account the interoccular distance. One of the fundamental problems with this approach is that if rendering a scene takes T units of time (without optimization), then it will take 2T units of time to render a complete stereo pair. However, optimizations have been devised to reduce the time required to render a stereo pair. Some of these optimizations are discussed in the next sections.

4.2 Pixel Shifting

Pixel shifting is a method of rendering an image as seen from a point (center of the interoccular distance) between the eyes, and then shifting it to the location where it would be seen from the perspective of the left and right eyes (Figure 3), *i.e.*, placing the previously computed color into the appropriate locations of the pixel buffer. It also uses the Z-buffer [Foley et al. 1990] of graphics hardware to determine which objects within the scene are in the foreground and should be rendered to the pixel buffer. Locations that are not drawn to in the pixel buffer are filled using linear interpolation [Burden and Faires 1993]. This method is significantly faster than rendering the image twice, from left and right perspectives, but lacks proper accuracy.

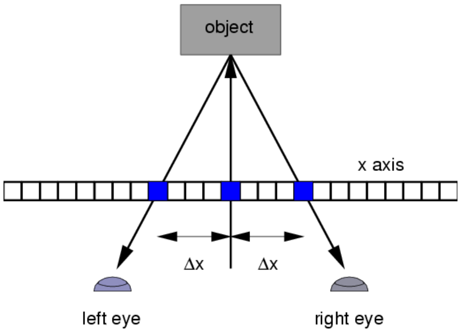

Figure 3: Illustration of pixel shifting algorithm (redrawn from [Harrison et al. 1997]).

4.3 Ray Tracing Optimization

The central idea in the ray tracing algorithm is to cast rays from a single point, located between the eyes, compute intersections with objects in the environment, and perform lighting, shading and other calculations for each pixel [Foley et al. 1990]. In ray tracing for stereo pairs, instead of rays being cast from a single point between the eyes, it is possible to shift the rays to the left and to right by a small amount, and render the scene twice from each displacement.

In order to optimize this algorithm, a method similar to pixel shifting can be used. Figure 4 illustrates graphically how this algorithm works. A ray is traced from the left eye. If it intersects an object, then the desired illumination model is applied. The ray is then back propagated to the right eye. If the ray has a direct path to the right eye, then the appropriate pixel on the screen is set to the same color of the pixel intercepted by the ray cast by the left eye.

If the ray intersects with an object before it can reach the right eye, then the pixel is marked not calculated. If this is the case, then the pixel must be retraced from the viewpoint of the right eye [Harrison et al. 1997]. Figure 5 shows an example of such a situation.

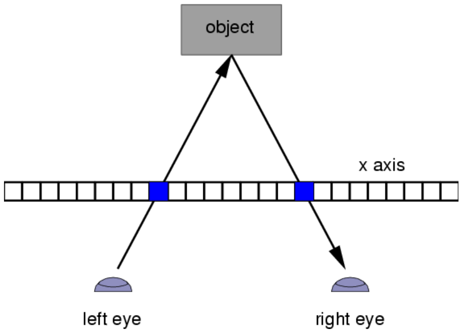

Figure 4: Sketch illustrating ray tracing optimization (redrawn from [Harrison et al. 1997]).

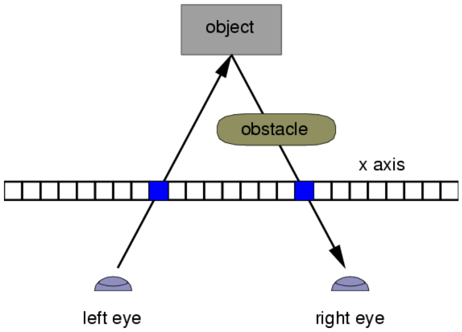

Figure 5: Sketch illustrating a situation where the ray tracing optimization fails due to occlusion (redrawn from [Harrison et al. 1997]).

4.4 Line Clipping Algorithms

Since stereo pairs ideally have no vertical parallax, this can be used to speed up the process of clipping lines. Line clipping for top, bottom, near and rear projection planes of the window can be done once, from the perspective of the eye the image is rendered from first. The same clipping status must be held true for the other eye. However, clipping to the left and right planes must be done for each eye [Harrison et al. 1997].

4.5 Stereo Pairs using OpenGL

The OpenGL standard [Woo et al. 1999] supports stereo pair production natively. The method employed is to render the image from the center of the interoccular distance. Images produced from this perspective are transformed to left and right eye perspective then stored in separate image buffers for each eye [Stoev et al. 2000].

Implementing stereo in OpenGL is relatively simple from a developer's point of view provided that the hardware supports stereo output. An OpenGL library call, if supported by graphics hardware drivers, will result in the images produced in stereo. When initializing the OpenGL program the following call should be made: *glutInitDisplayMode* (GLUT_STEREO). Additional initialization parameters must be added using the logic or bit operator. Shutter glasses will then be connected to the video output to produce a stereo output to the user.

It is also possible to use OpenGL to produce anaglyph graphics that can be viewed using glasses with different color lenses. This is done in several steps. The left and right perspectives of each image are generated and stored into separate buffers. Then each pixel can be extracted from the left and right buffers, and then assembled into a single image using the red from the left image and blue/green from the right image.

When the left eye image is rendered, instead of calling *glSwapBuffers*, the image is stored into an array for the left images using *glReadPixels*. Similarly, this is done with the right eye perspective. A third buffer is then used, where the red value is copied from the left buffer, and the green and blue values are copied from the right buffer. Then *glDrawPixels* is called, with the third buffer as a parameter. Using glasses that have red and blue for left and right eyes respectively, the viewer will have the impression of a 3D image.

4.6 Accelerating Stereo Projection

Another technique to reduce the rendering time of stereo images consists in partitioning a scene into background and foreground [Stoev et al. 2000]. Background images are rendered once from the center of the interoccular distance. The image data is stored into the image and depth buffers. A final rendering pass is done for eye specific foregrounds. In this case a camera is placed at each eye, separated by the interoccular distance and the foreground is rendered from each perspective. The corresponding values are written into the appropriate buffer [Stoev et al. 2000].

This method is based on the fact that distant objects appear similarly in both eyes. It has been shown experimentally that rendering time is decreased and credible stereo pairs can be produced [Stoev et al. 2000]. The important factor in implementing this algorithm is where to partition the environment. Research has shown that each scene has an optimal partition location which is based on the distance between the viewer and the object that he is focusing his eyes upon [Stoev et al. 2000].

5 Summary

In this report, we discussed characteristics of the human visual system that should be taken into account by computer graphics researchers interested in creating perceptually realistic 3D images. We also outlined a variety of hardware and software tools that take advantage of the physiological process of stereopsis to give viewers/users the perception of true 3D environments. Finally, we briefly examined optimizations that can be implemented to speed up the rendering of stereo images.

References

BURDEN, R., AND FAIRES, J. 1993. *Numerical Analysis*, fifth ed. PWS-KENT Publishing Company, Boston.

CRUZ-NEIRA, C., SANDIN, D., DEFANTI, T., KENYON, R., AND HART, J. 1992. The cave: audio visual experience automatic virtual environment. *Communications of the ACM 35*, 6, 64–72.

FOLEY, J., VAN DAM, A., FEINER, S., AND HUGHES, J. 1990. *Computer Graphics: Principles and Practice*, second ed. Addison-Wesley Publishing Company, Reading, Massachusetts.

HARRISON, L., MCALLISTER, D., AND M.DULBERG. 1997. *Stereo Computer Graphics for Virtual Reality - Siggraph 97*. Multimedia Lab, Department of Computer Science, North Carolina State Univerisity, Raleigh, North Carolina, USA.

K.W. BRODLIE, D.A. DUCE, J. G. J. W., AND WOOD, J. 2003. Distributed and collaborative visualization. *Eurographics 2003 - State of the Art Reports*, 139–166.

OVERHEM, R., AND WAGNER, D. 1982. *Light and Color*. John Wiley & Sons, New York, N.Y.

R.J. HUBBOLD, D., AND MOORE, C. 1998. Stereoscopic volume rendering. *Visualisation in Scientific Computing 98*, 105–115.

STEINMAN, S., STEINMAN, B., AND GARZIA, R. 2000. *Foundations of Binocular Vision: A Clincial Perspective*, first ed. McGraw-Hill, Appleton&Lange, New York.

STOEV, S., HUTTNER, T., AND STRABER, W. 2000. Accelerated rendering in stereo-based projections. *Collaborative Virtual Environments 3*, 213–214.

WOO, M., NEIDER, J., DAVIS, T., AND SHREINER, D. 1999. *OpenGL Programming Guide*, third ed. Addison-Wesley, Reading, Massachusetts.