Stereoscopic Vision Comfort

Skanda Shridhar*
Stanford University

Raj Shah†
Stanford University

Stanford EE 267, Virtual Reality, Course Report, Instructors: Gordon Wetzstein and Robert Konrad

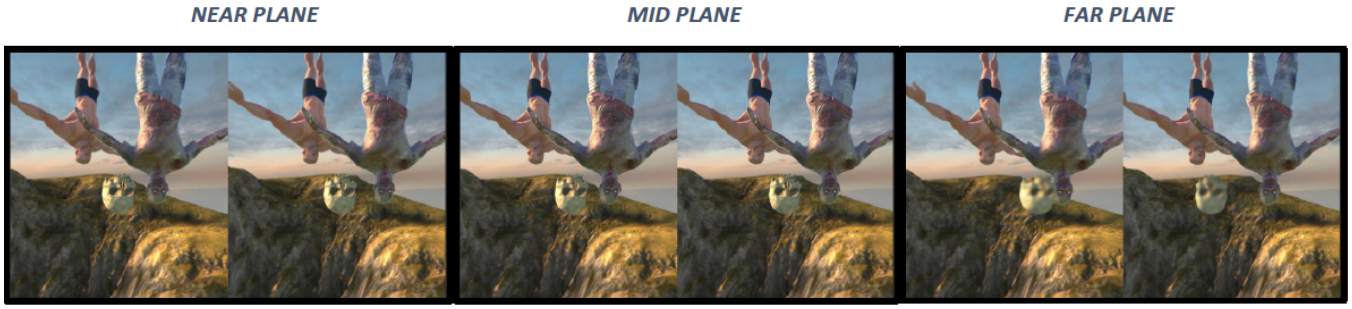

Figure 1: *Three pair of stereo images rendered using different stereoscopic parameters, namely, inter-axial separation and camera convergence based on the object of attention's distancefrom the camera. The stereoscopic controller algorithm can adjust these parameters, so that a specific depth range around the object the user is viewing is mapped into the zone of comfort.*

Abstract

In this work, we describe a method for mapping a virtual scene into a comfortable viewing range while being mindful of the depth of field effect in optics.

Instead of mapping an entire scene depth range into the comfort zone, the subset of the scene comprising the depth of field range is rendered into a target depth range. In certain circumstances, this has the effect of increasing interaxial separation, leading to more of a sense of 3D perception.

In our subjective user tests, our proposed method was preferred to a naive alternative. In cases where the object of attention was in the midground of the scene, a greater sense of 3D perception was reported.

Keywords: stereo, vision, depth of field, comfort zone

1 Introduction

Sterescopic content processing has advanced rapidly in recent years. With greater demand, there is growing concern about the adverse effects of prolonged viewing of stereoscopic content, especially visual fatigue and discomfort. Excessive disparities tend to be a cause of eye strain and sometimes nausea [Shibata et al. 2011]. In order to mitigate these effects, the stereoscopic parameters can be adjusted to control the disparity according to the viewing configurations and scene.

Oskam et. al [2011] propose a stereoscopic controller to control the interaxial separation between the cameras and their convergence settings to adjust the disparities and provide a mapping of the scene to a depth range around the physical screen. Manipulating stereoscopic parameters helps in controlling the perceived depth. The *perceived depth* is the geometric depth reconstructed by the user given a particular disparity. In the course of this paper, we review geometric models to find the perceived depth in terms of the scene depth and stereoscopic parameters. The volume around the screen to which the scene is mapped is referred to as *target depth range*.To make the viewing experience comfortable, the target depth range should be in the *comfort zone*, which is the range of the depths around the screen comfortable for perception.

Shibata et. al [2011] present methods to estimate the zone of comfort based on properties of the setup such as screen dimension. The method proposed by Oskam et. al [2011] attempts to map an entire scene depth range into the comfort zone, with the goal of improving the experience of interactive stereoscopic applications, such as games. As we will discuss, under certain circumstances, this approach produces small interaxial separations which can diminish the stereoscopic effect.

In this paper, we assume knowledge of the user's eye gaze and present a method to map the depth of field around the object of attention into the comfort zone. It results in larger interaxial separation and subjective user opinions seem to suggest that it provides more of a sense of depth perception. This work assumes prior knowledge of the objects the viewer is focusing on in a scene and hence, the major application of this work would be in stereoscopic system with integrated eye-tracking technology. Additionally, we blur regions outside the depth of field to add realism.

2 Related Work

In this section, we discuss related works on stereoscopic geometry, zone of comfort and dynamic camera control.

Stereoscopic Geometry: [Woods et al. 1993] present the geometry of stereoscopic camera and display systems. They also explore how system parameters effect image distortions such as depth plane curvature, shear distortion, etc. They also compare toed-in and parallel camera configurations and conclude that parallel camera configuration is preferred to the toed-in camera configuration. They also contend that the parallax range is limited by the association between vergence and accommodation.

Perception and Zone of Comfort: The issue of vergence-accommodation is studied through a series of experiments by [Shibata et al. 2011]. They find that negative conflicts are less comfortable at far distances and positive conflicts (caused when the ver-

*e-mail:skanda.shridhar@gmail.com
†e-mail:shahraj@stanford.edu

gence distance is greater than focal distance) are more comfortable at near distances (when the focal distance is greater than vergence distance). Their work also specifies different means of specifying the vergence distance. They also propose a method to estimate the comfort zone and thus find a range of comfortable disparities.

Dynamic Camera Control: [Oskam et al. 2011] proposed a stereoscopic controller for camera convergence and inter-axial separation for real time gaming. They present an optimization scheme for mapping the scene depth range to the comfort zone. In their work, the nearest and farthest scene depths are determined and constraints for the stereoscopic controller are determined. Their stereoscopic controller can thus render any scene into a target depth range for any display and viewing setup. Their goal is to achieve dynamic control of stereo parameters to assist with vision comfort in real-time application like games.

3 Geometry of Stereoscopic Cameras

In this section we will review the relations between the perceived depth and scene depth, given the interaxial separation and camera convergence. These relations will later assist with deriving the constraints needed to map the scene depth range into the target depth range. The relations given below have also been presented in the work by [Oskam et al. 2011]. The interaxial separation denoted by b is the distance between the two cameras and the camera convergence c_{cvg} is the distance from the convergence point to the midpoint of the line joining the two cameras as shown in Fig.(2). The distance of the viewer from the screen is denoted by d_v, whereas w_s is the screen width. The inter-pupillary distance d_e is usually considered to be 64 mm. When the user reconstructs a point at depth z, it results in an on-screen parallax p. The perceived depth z can then be formulated as

$$z(b, c_{cvg}, c) = \frac{d_v p(b, c_{cvg}, c)}{d_e - p(b, c_{cvg}, c)} \tag{1}$$

The disparity d on the image plane is proportional to the parallax p and can be found by scaling p by a factor of w_s/w_i, where w_s is the width of the screen and w_i is the width of the image plane. Thus, we can rewrite Eq. 1 as

$$z(b, c_{cvg}, c) = \frac{d_v d(b, c_{cvg}, c)}{\frac{w_i d_e}{w_s} - d(b, c_{cvg}, c)} \tag{2}$$

The triangle formed by the image shift h and focal length f is similar to the triangle formed by $b/2$ and c_{cvg}. Thus,h can be defined as $fb/(2c_{cvg})$. Also, the triangle formed by half of the scene distance $c/2$ and $b/2$ and the triangle formed by f and $h - d/2$ are similar. This can be used to reformulate $d(b, c_{cvg}, c)$ as $fb(c - c_{cvg})/(cc_{cvg})$. Using the above relation for d, we can write Eq. 2 as

$$z(b, c_{cvg}, c) = \frac{d_v(c - c_{cvg})}{\frac{w_i d_e c c_{cvg}}{w_s fb} - (c - c_{cvg})} \tag{3}$$

We will now find the scene depth in term of the perceived depth z, inter-axial separation b and camera convergence c_{cvg} using the screen-centric model in Fig. 2. Using similarity of triangles, the scene depth c is given by

$$c(b, c_{cvg}, z) = \frac{fb}{2h - d} \tag{4}$$

The image disparity d can be substituted by the on-screen parallax p according to the screen width w_s and image plane width w_i. Also, using the relation $h = fb/(2c_{cvg})$ found in the previous derivation,

$$c(b, c_{cvg}, z) = \frac{fb}{\frac{fb}{c_{cvg}} - \frac{w_i p}{w_s}} \tag{5}$$

Finally, the on-screen parallax p can be found in terms of the perceived depth z and distance of the viewer from the screen d_v using the fact that the triangle defined by p and z is similar to the triangle formed by d_e and $d_v + z$.

$$c(b, c_{cvg}, z) = \frac{fb}{\frac{fb}{c_{cvg}} - \frac{w_i}{w_s}\frac{d_e z}{d_v + z}} \tag{6}$$

4 OSCAM

4.1 Summary

In this section, we summarize and discuss the procedure for controlling the camera convergence c_{cvg} and interaxial separation b such that changing scene content can be optimally mapped to a target depth range [Oskam et al. 2011].

First, given a desired target depth z_i, we transform it into a corresponding d_i value using Equation (2). This is allowable since the transformation is independent of c_{cvg} and b. The next step is to formulate the mapping problem by using Equation (4) with the computed disparities d_i, the scene points c_i and the relation $h = fb/(2c_{cvg})$ to obtain the following system of equations.

$$fbc_i - fbc_{cvg} - c_i d_i c_{cvg} = 0 \quad for \quad i = 1...n, \tag{7}$$

In [Oskam et al. 2011] the special case is considered, in which two real-world depth values, which we will refer to as the near target depth z_1 and the far target depth z_2 are mapped to two scene points, referred to as the near scene point c_1 and the far scene point c_2. We will refer to the range $z_2 - z_1$ as the "target depth range" and $c_2 - c_1$ as the "scene depth range". The following solutions result from this procedure.

$$c_{cvg} = \frac{c_1 c_2 (d_1 - d_2)}{c_1 d_1 - c_2 d_2} \tag{8}$$

$$b = \frac{c_1 c_2 (d_1 - d_2)}{f(c_1 - c_2)} \tag{9}$$

4.2 Discussion

As stated by the authors, the goal of the OSCAM approach is to optimally map dynamically changing scene content, as during games, into the target depth range. In such a context, information about user gaze is unavailable. The heuristic of mapping the entire scene depth range is used, which is reasonable in the circumstances.

However, with this approach, certain commonly arising situations can result in very small inter-axial separation. An example of such a situation is when the near scene point is at 1.5m and the far scene point is at 1000m. Figure 3 plots the calculated interaxial separation against target depth range, for this scenario. The green dotted line shows the estimated inter-pupillary distance of eyes, 0.064m. Relative to that, the computed inter-axial separation is very small. While this may be comfortable to view in the sense that a flat image is comfortable to view, we believe that it has an adverse subjective effect, since it diminishes stereopsis.

Table 1 shows the inter-axial separation we obtained in our implementation of the OSCAM approach for four scenes.

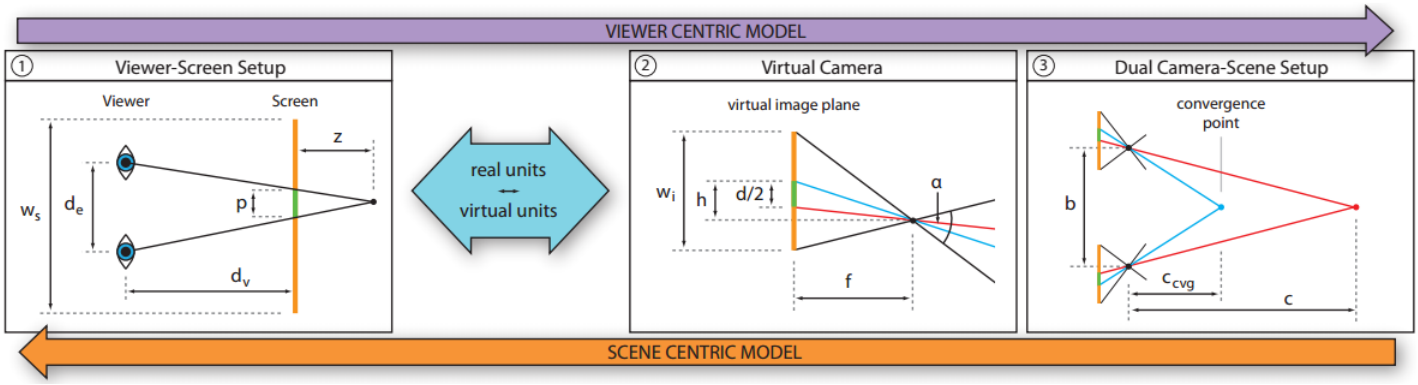

Figure 2: *Geometric Stereoscopic Model from [Oskam et al. 2011] (1) Viewer-Screen Setup where the viewer at a distance d_v from the scene reconstructs a point at depth z leading to an on-screen parallax of p (2) The parallax p corresponds to image disparity d. Here, f is focal length and h is the image shift (3) The cameras with inter-axial separation b and image shift h converge at c_{cvg}. The distance c corresponds to the image disparity d*

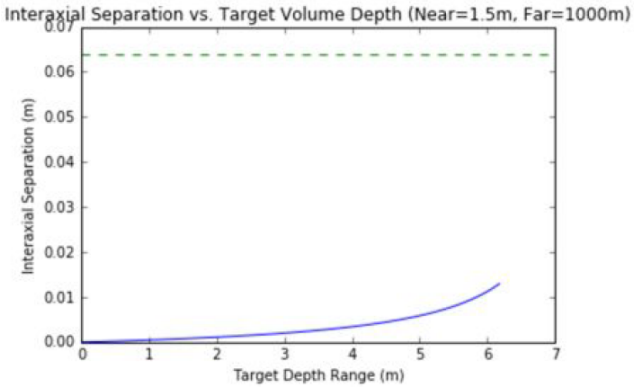

Figure 3: *Inter-axial separation obtained for increasing target depth volumes, when field being mapped is wide (1.5m to 1000m)*

	Inter-axial Separation (b) (in mm)	Min. Scene Depth (c_{min}) (in m)	Max. Scene Depth (c_{max}) (in m)
Scene 1	5.87	0.35	150
Scene 2	8.42	0.5	85
Scene 3	5.88	0.35	89
Scene 4	5.9	0.35	50

Table 1: *Disparities obtained in our implementation of Oskam's approach for 4 scenes.*

In our method, we allow ourselves the benefit of knowing where in our scene the user is looking, and we attempt to leverage that information to reduce the "flattening" effect of low inter-axial separation observed in the cases discussed above.

5 Methodology

As mentioned above, in our method we make use of knowledge of the user's gaze and attempt to explore whether we can leverage this to control camera convergence c_{cvg} and b in order to render virtual content in a fashion that is both comfortable and realistic, in the sense that it preserves stereopsis where possible.

5.1 Depth of Field

Before discussing our method, we briefly review the concept of *depth of field*. The depth of field range describes the region in a scene that appears acceptably sharp. Perfectly sharp focus is only possible for point objects at a single distance from the lens which we refer to as the *distance to focal plane* s. Point objects farther or closer than s produce blur spots on the retina, the radii of which increase with absolute distance from s. Our eyes can resolve these blur spots without confusion upto a certain diameter limit, which is referred to as the *circle of confusion*, c.

$$r = \frac{f}{s_1 - f} A|s - s_1| \tag{10}$$

In which s_1 is the distance to which the eye is accomodated, f is the focal length and A is the eye aperture.

Having determined the circle of confusion, we can then determine the near distance c_{near} and the far distance c_{far}, that form the boundary of the depth of field range.

$$c_{near} = \frac{hs}{h + (s - f)} \tag{11}$$

$$c_{far} = \frac{hs}{h - (s - f)} \tag{12}$$

Where h is the hyperfocal distance, given by,

$$h = \frac{f^2}{Ar} \tag{13}$$

5.2 Our Method

We assume knowledge of the user's gaze, so s is known. Using Equations (11) and (12), we can determine c_{near} and c_{far}. We then use Equations (8) and (9) to determine values for camera convergence c_{cvg} and interaxial separation b in order to map the depth of field region into a $subinterval$ of the comfort zone. The width of the size interval can either be hardcoded from empirical experiments or determined using a technique such as that described in the next section.

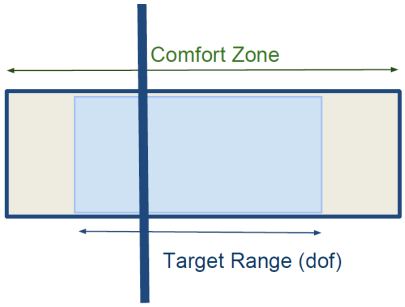

Figure 4: *Illustration showing the comfort zone and target depth range. Target depth range must be aligned around the screen and within the comfort zone.*

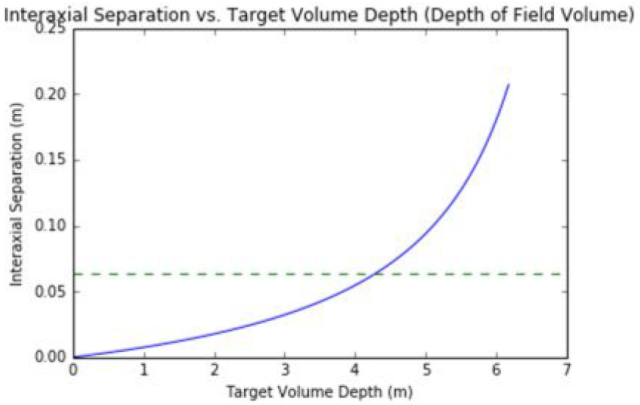

Figure 5: *Inter-axial separation obtained for increasing target depth volumes, when field being mapped is 0.1m wide.*

With the depth of field range having been rendered into the comfort zone, the rest of the scene outside the depth of field range is blurred.

5.3 Determining Target Depth Range

In the previous section we mentioned that the depth of field range is rendered into a sub-interval of the comfort zone. In this section we present a possible method for determining the target depth range.

Consider a scenario where the range being mapped is found to extend between 1.5m and 2.5m. Using Equation (9), we can plot the inter-axial separation against the target depth range. This plot is shown in Figure 5. Comparing this against Figure 3 which mapped a range extending between 1.5m and 1000m, we observe that the calculated inter-axial separations are larger and in the ballpark of the biological inter-pupillary distance (shown as the dotted green line).

A reasonable heuristic for determining the target depth range could then be to choose the value on the x axis that produces an inter-axial distance closes to the biological inter-pupillary distance, without exceeding the width of the comfort zone.

6 Evaluation

We wanted to evaluate our method along two axes, visual comfort and realistic depth perception (stereopsis).

6.1 Visual Comfort

To evaluate visual comfort, we conducted a user study on 5 different subjects.The users were asked to compare our method with the conventional stereoscopic rendering with fixed inter-axial separation and camera convergence on a virtual reality head mounted display. The inter-axial separation for the conventional setting was set to 64mm and the camera convergence was set to 1.7m which was the distance of the screen from the viewer.

In the user study, we rendered 4 different scenes with objects at different depths using two rendering modes, one with the naïve stereoscopic parameters and another in which the parameters were obtained using our method. The users were asked to focus on the object in the foreground, mid-ground and background respectively for each of the scene. The order of the rendering mdoes was randomized. After viewing each scene, the users were asked to chose the rendering mode which they felt was more comfortable and realistic. Table 2 shows the results of the user study. Our method was preferred to the conventional method in 16 out of 20 trials.

	Scene 1	Scene 2	Scene 3	Scene 4
Subject 1	Proposed	Proposed	Naïve	Proposed
Subject 2	Proposed	Naïve	Proposed	Naïve
Subject 3	Proposed	Proposed	Proposed	Proposed
Subject 4	Proposed	Naïve	Proposed	Proposed
Subject 5	Proposed	Proposed	Proposed	Proposed

Table 2: *User preferences. The Proposed method is ours while the Naive approach is a conventional stereo setup*

6.2 Stereopsis

To evaluate this we used two measures; one the one hand we asked users of our study to choose which mode they felt had better 3D depth perception. The feedback overwhelmingly favored our method.

For the second measure, we examined the inter-axial separations that our approach produced for the various scenes and focal planes. Comparing the interaxial separations in Tables 3, 4 and 5 with those in Table 1 shows that our method yielded configurations with wider separation. The convergences were also appropriately adjusted to render into the comfort zone.

It is observed that for the midground objects, especially, our method produces a high interaxial separation. Based on the feedback regarding visual comfort and depth perception, we believe this may suggest an improvement.

	Near Plane	
	Inter-axial Separation (mm)	Distance to Object of Attention (m)
Scene 1	13.1	0.66
Scene 2	11.73	0.5
Scene 3	20.43	0.35
Scene 4	16.01	0.8

Table 3: *Inter-axial Separations obtained by the proposed method and distances to the near plane for 4 scenes*

	Middle Plane	
	Inter-axial Separation (mm)	Distance to Object of Attention (m)
Scene 2	52.89	2.25
Scene 3	23	1
Scene 4	110.83	1.93

Table 4: *Inter-axial Separations obtained by the proposed method and distances to the middle plane for 3 scenes*

	Far Plane	
	Inter-axial Separation (mm)	Distance to Object of Attention (m)
Scene 2	23	75
Scene 3	125.71	1.9
Scene 4	23	75

Table 5: *Inter-axial Separations obtained by the proposed method and distances to the far plane for 3 scenes*

7 Discussion

Our method attempts to compute the depth of field range and map it into the comfort zone, while blurring out the rest of the scene to produce a realistic effect. Since it relies on mapping the depth of field range, it has the potential to show better results for scene distances that are less than the hyperfocal distance (beyond which the depth of field is at its maximum).

The approach seems to make the most difference when the object of attention is in the midground of the scene. This renders the object in the zone of comfort while nearer objects are rendered with higher disparities and a blurring effect. Based on subjective opinions of subjects in our user study, this effect appears to enhance depth perception. When the object of attention is far off, the blurring of near objects also appears to add value.

But many avenues for improvement exist. Firstly, we should use a custom shader to implement the blurring operation. At the moment we use a Unity library that blurs discrete objects rather following a vertex by vertex approach.

Secondly, we need to investigate improvements to our method of mapping scene depth ranges to target depth ranges. Recently, some nonlinear approaches have been proposed which may be better for the task. [Lang et al. 2010]

Thirdly, time interpolation, as implemented in [Oskam et al. 2011] is a necessary feature of any stereoscopic controller that is used in real time applications. While we did prototype a rudimentary time implementation procedure, it lacked sophistication since we prioritized other aspects of the project. Given more time we would have devoted more attention here.

Finally, since our method involves a mapping from one volume range in the scene to another one in real world space, concerns arise about perceptually jarring "shifts" that may result from odd mappings. While we did not encounter such artefacts in our experiments, this possibility needs to be explored in greater detail and better understood.

Acknowledgments

This project was a part of the EE267 Virtual Reality course at Stanford conducted by Prof. Gordon Wetzstein and Robert Konrad. We would like to thank them for this insightful course, guiding us through this project and providing us with the required resources for the project. We are grateful to Haricharan Lakshman for guiding us through the course of this project.

References

LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3d. In *ACM Transactions on Graphics (TOG)*, vol. 29, ACM, 75.

OSKAM, T., HORNUNG, A., BOWLES, H., MITCHELL, K., AND GROSS, M. H. 2011. Oscam-optimized stereoscopic camera control for interactive 3d. *ACM Trans. Graph. 30*, 6, 189.

SHIBATA, T., KIM, J., HOFFMAN, D. M., AND BANKS, M. S. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of vision 11*, 8, 11–11.

WOODS, A. J., DOCHERTY, T., AND KOCH, R. 1993. Image distortions in stereoscopic video systems. In *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, International Society for Optics and Photonics, 36–48.