Enhanced Depth Discrimination Using Dynamic Stereoscopic 3D Parameters

Arun Kulshreshth
Department of EECS
University of Central Florida
Orlando FL 32816
arunkul@knights.ucf.edu

Joseph J. LaViola Jr.
Department of EECS
University of Central Florida
Orlando FL 32816
jjl@eecs.ucf.edu

Abstract

Most modern stereoscopic 3D applications (e.g. video games) use optimal (but fixed) stereoscopic 3D parameters (separation and convergence) to render the scene on a 3D display. However, keeping these parameters fixed does not provide the best possible experience since it can reduce depth discrimination. We present two scenarios where the depth discrimination could be enhanced using dynamic adjustments to the separation and the convergence parameters based on the user's look direction obtained from head tracking data.

Author Keywords

Stereoscopic 3D; Dynamic Stereo Parameters; Head Tracking; Visual Fatigue; User Study; User Experience.

ACM Classification Keywords

H.5.m [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous.; K.8.0 [Personal Computing]: Games

Introduction

Stereoscopic 3D displays present two images offset to the left and right eye of the user and these images are then fused by the brain to give the perception of 3D depth. The generation of these two images uses two stereo parameters: separation and convergence. Separation is defined as the interaxial distance between the centers of
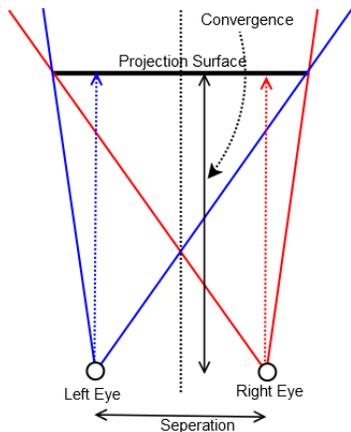
Figure 1: Off-Axis stereo projection.

the two virtual eye camera lenses in the scene and the convergence is defined as the distance of the plane where left and right eye camera frustums intersect (see Figure 1). Currently, most stereoscopic 3D applications fix convergence and separation values for optimal viewing during usage time. However, this approach reduces stereo depth in certain scenarios. Two examples are when the depth range has a large variability between different scenes (e.g. transition from inside a room to an outdoor scene) and when a large object (e.g. a gun in FPS games, the cockpit in air-combat games, etc.) is present in front of the camera. The fact that these parameters are optimized to minimize visual discomfort uniformly during usage usually limits the convergence and separation values. Depth discrimination (the ability to judge relative depths of objects in the scene) in a stereo 3D application could potentially be improved if the stereo parameters are dynamically adjusted based on the scene.

We designed two scenarios: a scene with large depth range variability across different directions and a scene with a large object in front of the camera. Each of these scenarios had situations that could be augmented by using dynamic stereo parameters to enhance depth discrimination. We conducted a within subjects experiment to evaluate the effectiveness of dynamic stereo parameters (separation and convergence). We examined qualitative data on a user's perception of depth, immersion, presence, and visual discomfort.

Related Work

Recent work on stereoscopic 3D found it useful for games depending upon the task involved [6, 7]. While stereoscopic 3D has shown some positive benefits depending on the task, it also has shown to cause negative symptoms as well, such as eyestrain, headache, dizziness, and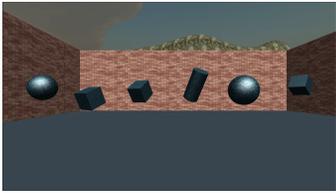 nausea [5]. Stereo comfort could be increased by either changing stereo parameters or using depth of field (DOF) blurring.

Several researchers [1, 3] have explored gaze-based depth of field (DOF) effects to minimize visual fatigue. However, people generally disliked the DOF effect with temporal lag of the gaze-contingent effect being a possible reason. Ware [8] proposed dynamic adjustment of the stereo separation parameter to minimize visual discomfort and optimize stereo depth. Our work adjusts both separation and convergence parameters for a better visual experience with enhanced depth discrimination. Furthermore, their results revealed that the separation must be changed gradually over a few seconds to allow users to adjust without noticing any visual distortion of the scene. Bernhard et al. [1] explored dynamic adjustment of stereo parameters using gaze data and found that it reduces stereo fusion time and provides a more comfortable viewing experience.
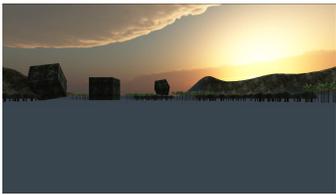
The past work on dynamic stereo mentioned above used simple static scenes (e.g. random-dot stereograms, a picture, etc.) to evaluate their work. None of the work explored the benefits of dynamic stereo in complex scenes like in modern video games. To the best of our knowledge, our work is the first to systematically explore dynamic stereo for more complex dynamic scenes.

Dynamic Stereoscopic 3D

Stereo parameters (separation and convergence) could be optimized based on the type of scene. Ideal application candidates for these optimizations could be classified in two broad categories. The first category is an application where there is a large variation in depth range across scenes and the second category is an application which always has a large object in front of the camera.

(a) Depth is limited by wall in this direction

(b) Unlimited depth in this direction

Figure 2: Scene 1: A scene with variable depth range across different directions.

Figure 3: Scene 2: A scene with a large object in front of the camera.

Type 1: Large depth range variation

The separation value is dependent on the depth range of the scene. For better depth discrimination, the separation is directly proportional to the maximum depth in the scene. Similarly, the convergence distance is also limited by the depth in the scene for a comfortable viewing experience. When there is a large depth variation across scenes, the separation and convergence values have to be set based on the scene with least depth range. If the separation and the convergence values are set based on a scene with large depth then they will make another scene with less depth uncomfortable to look at. Therefore, these parameters must be changed dynamically from scene to scene for enhanced depth discrimination in all the scenes.

We implemented a scene which has a limited depth in one direction and a large depth range in the opposite direction (see Figure 2). Head tracking is used to control the head of a first person controller (FPC) and a mouse is used to rotate the body of the FPC. The convergence value is dynamically changed based on the object being looked at and the separation is changed based on the depth range of the scene in front of the camera (see Algorithm 1 for details). The convergence and the separation values are changed gradually, as proposed by Ware [8], to allow enough time for the user's eyes to adjust.

Type 2: Large object in front of camera

When a large object (e.g. a gun in FPS games, the cockpit in air-combat, etc.) is present in front of the camera, the stereo parameters have to be optimized to keep that large object always in focus thereby limiting the depth discrimination ability. However, when the player's head is rotated/translated, that nearby object may not be in the player's view and stereo depth could be increased leading to enhanced depth discrimination.

We implemented an air combat game scene (see Figure 3) as a representative of this category of applications. In the

game, the player has to control an aircraft, using a joystick, in a first person view controlled using head tracking. In addition, the user can move his/her head closer to the screen to zoom into the scene for iron-sighting distant enemies. We optimized stereo parameters under two conditions. First, when the user is looking sideways (left/right) and second, when the user is zoomed into the scene (see Algorithm 2 for details). In both of these cases, the user is not looking at the cockpit. When the player's head is rotated sideways (left/right), the separation is increased with linear scaling proportional to the head's rotation and the convergence is not changed. When a user zooms in the scene, the separation is increased with linear scaling proportional to the head's displacement. At the same time, the convergence is linearly decreased with the head's displacement to keep both the crosshair and background in focus. These dynamic parameters ensured a comfortable stereoscopic 3D experience and provided better depth discrimination for this air-combat game.

Implementation Details

We used Nvidia's 3D vision for our implementation and thus used the NVAPI library to change the convergence and the separation. According to the NVAPI library, the normalized eye separation is defined as the ratio of the interocular distance (between the eyes) and the display screen width. The separation value used in the driver is a percentage of this normalized eye separation and hence is a value between 1 and 100. Convergence is defined as the distance (in meters) of the plane of intersection of the left and right eye camera frustums with off-axis (or parallel) projection (see Figure 1). Projection matrices were calculated automatically by the driver.

Scene 1. For static stereo, the convergence was set to 1.0 and the separation was set to 20.0. In the case of dynamic stereo, the algorithm is described in Algorithm 1.

Figure 4: The experiment setup consisted of a 27" BenQ XL2720Z 3D monitor, Nvidia 3D Vision kit, a TrackIR 5 with Pro Clip (mounted on a headphone), a Logitech Extreme 3D Pro joystick, and a PC (Core i7 4770K CPU, GTX 780 graphics card, 8 GB RAM).

Algorithm 1 Calculate stereo parameter for scene 1

1: $S_1 \leftarrow$ separation for lower depth range
2: $S_2 \leftarrow$ separation for higher depth range
3: $C_1 \leftarrow$ convergence for higher depth range
4: $SF \leftarrow$ smothing factor
5: $threshold \leftarrow$ depth threshold
6: $\Delta t \leftarrow$ time between frames rendered on screen
7: $t \leftarrow SF \times \Delta t$
8: $C \leftarrow 1.0$
9: $S \leftarrow S_1$
10: Use raycast to find object Obj in front of camera
11: $d \leftarrow$ distance of Obj
12: **if** $d < threshold$ **then**
13: $\quad C \leftarrow C + (d - C) \times t$
14: $\quad S \leftarrow S + (S_1 - S) \times t$
15: **else**
16: $\quad C \leftarrow C + (C_1 - C) \times t.$
17: $\quad S \leftarrow S + (S_2 - S) \times t.$
18: $convergence \leftarrow C$
19: $separation \leftarrow S$

We set $SF = 3, threshold = 50, C_1 = 30, S_1 = 20$ and $S_2 = 50$ in the implementation. These values were obtained based on several pilot studies for scene 1.

Scene 2. For static stereo , the convergence was set to 4.0 and the separation was set to 5.0. The dynamic stereo algorithm is described in Algorithm 2. We set $C_0 = 4.0, C_1 = 0.001, S_0 = 5.0, S_1 = 60.0, roty_1 = 10$ and $roty_2 = 60.0$ in our implementation. These values were obtained based on several pilot studies for scene 2.

User Evaluations

We conducted an experiment to evaluate the effectiveness of dynamic stereo parameters. We recruited 12 participants (10 males and 2 females ranging in age from 18 to 33 with a mean age 27.83) from the university population. The experiment duration ranged from 20 to

---

Algorithm 2 Calculate stereo parameter for scene 2

1: $ihp \leftarrow$ initial head position
2: $mhp \leftarrow$ head position when completely zoomed in
3: $chp \leftarrow$ current head position
4: $roty \leftarrow$ current head rotation along y-axis
5: $roty_1 \leftarrow$ min head rotation along y-axis
6: $roty_2 \leftarrow$ max head rotation along y-axis
7: $C_0 \leftarrow$ initial convergence
8: $C_1 \leftarrow$ final convergence after zooming
9: $S_0 \leftarrow$ initial separation
10: $S_1 \leftarrow$ maximum separation
11: $C \leftarrow C_0$
12: $S \leftarrow S_0$
13: nnzoom in case
14: **if** $|ihp - chp| > 0$ **then**
15: $\quad C \leftarrow C_1 + (C_0 - C_1) \times (mhp - chp)/(mhp - ihp)$
16: $\quad S \leftarrow S_1 + (S_0 - S_1) \times (mhp - chp)/(mhp - ihp)$
17: nnlook left/right case
18: **if** $roty > roty_1$ and $roty < roty_2$ **then**
19: $\quad C \leftarrow C_0$
20: $\quad S \leftarrow S_1 + (S_0 - S_1) \times (roty_2 - roty)/(roty_2 - roty_1)$
21: **else if** $roty > roty_2$ **then**
22: $\quad C \leftarrow C_0$
23: $\quad S \leftarrow S_1$
24: $convergence \leftarrow C$
25: $separation \leftarrow S$

---

30 minutes. The experiment setup is shown in Figure 4. We used the Unity3D game engine for implementing the scenes. The TrackIR 5 camera and the Nvidia IR emitter were mounted on the top of monitor. Participants were seated about 2 feet away from the display. To make sure that all our participants were able to see stereoscopic 3D, we used the Nvidia medical test image to test stereo abilities of participants and all our participants passed the test. Note that Nvidia 3D glasses are designed such that they can be easily used over prescription glasses without any interference to the user.

Question	Scene1	Scene2
DD	$Z = -3.084, p < 0.005$	$Z = -3.078, p < 0.005$
JD	$Z = -3.086, p < 0.005$	$Z = -2.971, p < 0.005$
PF	$Z = -2.810, p < 0.05$	$Z = -2.638, p < 0.05$

Table 1: Results of Wilcoxon signed rank test for qualitative questions. DD: Depth Discrimination, JD: Judgment of Depth and PF: Preference

We chose a within-subjects design for our experiments. Each scene was presented to the participants with both static and dynamic stereo parameters. The users were asked to judge the relative depth of objects in both the scenes (like the cubes in the first scene and other objects in the second scene) and based on that they answered questions about depth discrimination. While performing this judgment task, they did not know if the scene used dynamic stereo or static stereo. In addition, they were asked to rotate their head and not their eyes to look around in both scenes. Each condition was presented to the participants in pre-selected counterbalanced order based on a Latin square design. After the experiment, the participant filled out a post-questionnaire about each scene with questions about depth discrimination, user preference, and visual discomfort.

Results

To analyze the Likert scale data, we used Wilcoxon signed rank test with $\alpha = 0.05$. The results for the qualitative questions are summarized in Table 1 and mean values are plotted in Figure 5. Compared to static stereo:

- depth discrimination was significantly improved with presence of dynamic stereo.
- significantly more people felt that they were able to correctly judge the relative depths of objects in scenes when dynamic stereo was present.
- significantly more people preferred using dynamic stereo.

Except for one participant, no one felt any significantly negative symptoms by watching the scenes in stereoscopic 3D (static as well as dynamic). One participant was very sensitive to stereoscopic 3D. He experienced moderate eye strain and discomfort with both static as well as dynamic stereo.

Discussion

Our scenes were designed keeping stereoscopic viewing in mind and used design guidelines from the literature [6, 7]. We chose the separation and the convergence values for each scenario such that the visual discomfort was minimized. During our pilot testing, these values were optimized based on user feedback to ensure that they are comfortable for most users. Most of our user study participants did not experience any visual discomfort with either static or dynamic stereo.

Our study also had some limitations. We used head tracking data to approximate the user's look direction. But, a user may not always be looking straight ahead since the eyes could look in a different direction. We asked our users to rotate their head and not their eyes to look around in the scene. However, this was not natural and could have a minor effect on our results. We expect that using an eye tracker would even further improve our results. We did not consider the variation in interocular distance between the users in our experiments. However, we expect that the results would be similar since our algorithms use (see implementation details) the ratio of interocular distance (between 58mm and 70mm [2]) and display width (27 inch in our experiment) which is minimally affected by this variation in interocular distance. In addition, our small sample size (12 participants) could have a minor affect on our results.

We would like to mention that the use of dynamic stereo would change the geometry of the scene (e.g. an increase in separation makes the world seem smaller and/or the observer feel larger) and may not be a good idea in situations where scale is of critical importance such as in case of industrial design applications. Regardless, our results indicate that dynamic stereo has potential to improve depth discrimination in stereo 3D applications. Future application designers should use dynamic stereo adjustments to provide a better experience to the user.
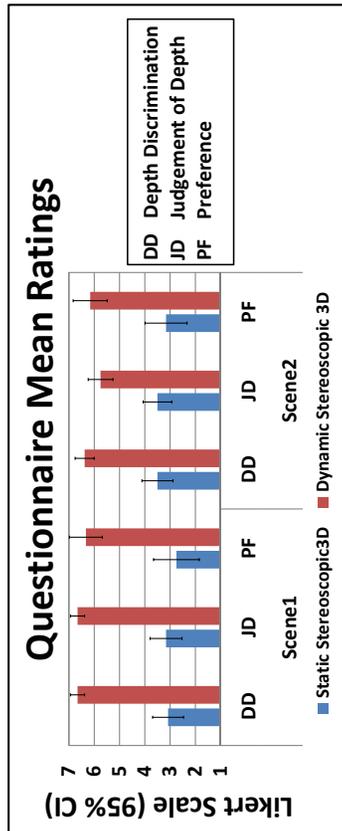
Figure 5: Mean qualitative ratings for both scenes based on type of stereoscopic 3D

However, these parameters should be chosen wisely, based on the scene, to minimize visual discomfort.

Conclusion and Future Work

We presented two scenarios where optimizing the stereo parameters (separation and convergence) could enhance the depth discrimination of the user. Our preliminary results indicate that participants preferred to use dynamic stereo over static stereo since it significantly improved the depth discrimination in the scene. Our study is a preliminary step towards exploring the effectiveness of dynamic stereo in stereoscopic 3D applications and further research with more scenarios is required.

In our experiment, the values of the stereo parameters were determined based on our pilot studies. However, we believe that these values could be expressed in terms of display size, distance of the user from the display, and distance of the object being looked at in the scene. We plan to explore this direction in future work. Furthermore, we did not consider any quantitative measures as part of this work. We plan to include depth judgment tasks (e.g. Howard-Dolman test [4]) in our future experiments to quantify the differences between dynamic and static stereo scenes. In addition, we used head tracking data to approximate the user's look direction. In the future, we would like to use eye tracking to get more accurate gaze direction and it should further improve our dynamic stereo algorithms.

Acknowledgments

References

[1] Bernhard, M., Dell'mour, C., Hecher, M., Stavrakis, E., and Wimmer, M. The effects of fast disparity adjustment in gaze-controlled stereoscopic applications. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, ACM (New York, NY, USA, 2014), 111–118.

[2] Dodgson, N. A. Variation and extrema of human interpupillary distance, 2004.

[3] Duchowski, A. T., House, D. H., Gestring, J., Wang, R. I., Krejtz, K., Krejtz, I., Mantiuk, R., and Bazyluk, B. Reducing visual discomfort of 3d stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '14, ACM (New York, NY, USA, 2014), 39–46.

[4] Howard, H. J. A test for the judgment of distance. *Transactions of the American Ophthalmological Society 17*, 195-235 (1919).

[5] Howarth, P. A. Potential hazards of viewing 3D stereoscopic television, cinema and computer games: a review. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists) 31*, 2 (2011), 111–122.

[6] Kulshreshth, A., Schild, J., and LaViola Jr., J. J. Evaluating user performance in 3D stereo and motion enabled video games. In *Proceedings of the International Conference on the Foundations of Digital Games*, ACM (New York, NY, 2012), 33–40.

[7] Schild, J., and Masuch, M. Fundamentals of stereoscopic 3D game design. In *ICEC, International Federation For Information Processing* (2011), 155–160.

[8] Ware, C. Dynamic stereo displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, ACM Press/Addison-Wesley Publishing Co. (New York, NY, USA, 1995), 310–316.