Quality of Experience for the Horizontal Pixel Parallax Adjustment of Stereoscopic 3D Videos

Di Xu, *Student Member, IEEE*, Lino E. Coria, *Senior Member, IEEE*, and Panos Nasiopoulos, *Member, IEEE*

***Abstract*--When capturing stereoscopic 3D video, video distortions can be avoided by setting up the cameras in parallel. Unfortunately, the resulting 3D scene appears to be in front of the screen, which is uncomfortable for the viewers. This problem can be solved by modifying the horizontal pixel parallax, which brings the picture on and behind the screen. We employ an algorithm that extracts and matches features from both left and right video sequences to automatically shift the video frames so that the closest object to the screen has zero pixel parallax. We then perform subjective tests to show that this parallax-adjusting process significantly increases the 3D quality of experience by 19.86%.**

I. Introduction

Despite the recent availability of 3D-enabled theaters and displays, many viewers complain about the headaches and nausea caused by low-quality stereoscopic content. Over the years, stereographers have empirically obtained a few rules of thumb for capturing fine 3D content [1]. In addition, systematic subjective evaluations have allowed researchers to find and quantify the elements and parameters that need to be considered in order to capture high-quality 3D content [2]–[4].

One of the most important factors for capturing high-quality stereoscopic content is the proper setup of the two cameras, since it allows content creators to control the 3D effect [1]. There are basically two options for setting up the cameras. The first option is having the cameras converge. This configuration, however, has shown to have some side-effects, such as the keystone effect [5], which produces undesirable distortions to the content [6]. The second option consists of setting up the two cameras so that their axes are parallel. In this case, the cameras converge at infinity and the resulting 3D scene appears to be entirely in front of the screen. Each photographed object in this case is known to have a negative horizontal pixel parallax, which occurs when the left view of an object is located further to the right than the right-view version of the same object [7].

Experimented stereographers are aware that viewers feel uncomfortable when the objects constantly appear in front of the screen. In order to reallocate the 3D effect behind the display, it is necessary to modify the depth information (i.e., horizontal pixel parallax) that is produced when the 3D content is captured with parallel-aligned cameras.

In this paper, we performed subjective evaluations to verify and quantify the increase in 3D quality of experience caused by the adjustment of the horizontal pixel parallax. In order to do these tests, we employed an algorithm that extracts and matches features from both left and right video sequences to automatically shift the video frames so that the closest object has zero parallax and thus appears on the screen.

This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Institute for Computing, Information and Cognitive Systems (ICICS), University of British Columbia (UBC).

II. Acquisition and Alignment

We aligned two identical HD cameras in parallel and attached them to a rig that was specifically made for them. Since zoom lenses may differ [2], only the extreme ends of the zoom range were used.

A video database was created, following the capturing procedure proposed in [4]. The resulting database contains 25 ten-second stereoscopic videos depicting people, objects, buildings, and landscapes. The videos have a resolution of 1920 pixels × 1080 pixels and a frame rate of 30 fps. There are two important distances considered for every video: (i) the distance between the cameras and the closest photographed point, d_{min}, with values of 0.5 m, 1 m, 2m, and 3 m; and (ii) the distance between the cameras and the background, d_{max}, with values of 5 m, 10 m, 50 m, and infinity.

Once the content has been captured, we perform temporal synchronization and video rectification if needed. In addition, we shift the frames horizontally in order to eliminate the negative horizontal pixel parallax. As a result of this action, the photographed objects do not appear to pop out of the screen. In addition, we avoid stereoscopic window violations [1] (i.e., objects that pop out of the screen and are partially blocked by the screen frame provide the brain with conflicting depth cues). However, eliminating negative disparities through horizontal shifting may result in large positive disparity regions in the background of a scene, which cause eye divergence. Since the background is usually not the point of interest, especially in the case of video, such divergence does not have a strong negative impact on the quality of the content. This is later verified by the results of our subjective tests.

Both video rectification and horizontal parallax adjustment are performed automatically in our experiment. This process is done by computing features using the Scale Invariant Feature Transform (SIFT) algorithm [8], and aligning the features that are common to both left and right videos.

III. Subjective Evaluation

The subjective test was conducted using a 65” 3D Plasma HDTV with a resolution of 1920 × 1080 pixels. Viewers used 3D shutter glasses with a refresh rate of 600 Hz.

We set up the viewing conditions for the subjective assessment according to Section 2.1 of the ITU-R BT.500-11 [9], which is also recommended in ITU-R BT.1438 [10]. We first ran a training session to show the viewers the quality range of stereoscopic videos, without imposing the quality of the videos. The four videos used in the training session were different from the test videos. In the subjective evaluation session, 50 videos were used. These 50 videos were generated from the 25 videos in our database, having two versions for each video sequence. The first version has no horizontal pixel parallax adjustment and the second version has the adjustment.

The order of the 50 videos was randomized in order to reduce the influence of previously viewed videos on the subsequent ones. In order to introduce a grading and relaxation period, each video was followed by a four-second interval of a 2D mid-grey image.

Twenty subjects participated in our test, seven females and thirteen males. The ages of the subjects ranged from 23 to 59, with an average age of 32. All observers had little experience watching 3D content, and they were screened for visual acuity using the Snellen chart and color vision using the Ishihara test. The subjects were seated in line with the center of the display at a distance recommended by the display manufacturer. Two viewers participated in each viewing session. The viewers were asked to rate each video based on the level of naturalness of their experience watching the 3D content.

IV. Analysis and Results

One out of the twenty observers was detected as an outlier, and all the scores of this observer were eliminated. We take the average score across all valid observers for each capture setting as the mean opinion score. The Student's t-tests [11] are used to compute confidence intervals with the significance level being 95%.

All test videos were divided into four different groups, each with a common value of d_{max}. Fig. 1 shows the mean opinion scores and confidence intervals versus different values of d_{min} for four different d_{max} distances. The blue lines represent data before horizontal parallax adjustment and the pink lines show the data after horizontal parallax adjustment. We observe that for the same d_{max}, the video quality increases with d_{min} when the latter parameter is smaller or equal to 2m. The quality levels off when d_{min} is greater than 2m. This trend in quality applies to videos before and after horizontal parallax adjustment. We also observe that the 3D video quality is greatly improved by the horizontal parallax adjustment except for one video, where the video depth bracket is very small. We computed the average scores of all 25 test videos before and after horizontal parallax adjustment, respectively. The results indicate that viewers perceive horizontally adjusted videos to possess higher quality than non-adjusted ones, with an average quality-score gain of 19.86%.

We can also conclude that although the horizontal parallax adjustment has greatly improved the quality of 3D content, very small d_{min} still leads to unsatisfactory quality (i.e., the mean opinion score is below 50 on a 0 to 100 rating scale in all four subplots in Fig. 1), and hence, needs to be avoided in the capturing process.

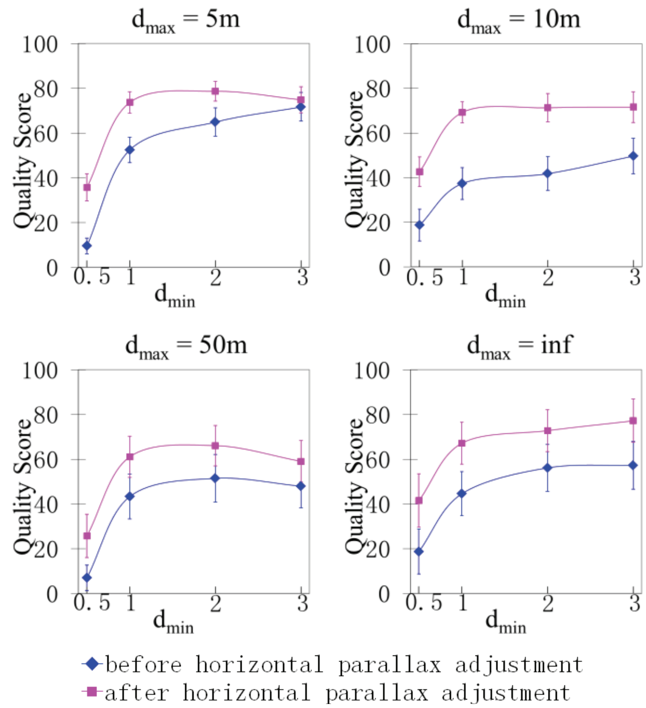

Fig. 1. Mean opinion scores and their confidence intervals for d_{min} = 0.5m, 1m, 2m, and 3m. In reading order, the four subplots correspond to the cases when d_{max} is 5m, 10m, 50m, and infinity. The blue lines represent data before horizontal parallax adjustment and the pink lines show the data after the adjustment.

References

[1] B. Mendiburu, “3D movie making: stereoscopic digital cinema from script to screen.” Elsevier - Focal Press, 2009.
[2] L. Goldmann, F. D. Simone, and T. Ebrahimi, “Impact of acquisition distortion on the quality of stereoscopic images,” in International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona, USA, January 13-15, 2010.
[3] L. D. Pockett and M. P. Salmimaa, “Methods for improving the quality of user created stereoscopic content,” Stereoscopic Displays and Applications XIX, Proc. of SPIE-IS&T Electronic Imaging, SPIE, vol. 6803, no. 680306, 2008.
[4] D. Xu, L. Coria, and P. Nasiopoulos, “Guidelines for capturing high quality stereoscopic content based on a systematic subjective evaluation,” in IEEE International Conference on Electronics, Circuits, and Systems, ICECS 2010, Athens, Greece, 2010.
[5] R. S. Allison, “The camera convergence problem revisited,” in Stereoscopic Displays and Virtual Reality Systems XI, Proceedings of SPIE, vol. 5291, San Jose, CA, USA, 2004, pp. 167-178.
[6] A. J. Woods, T. Docherty, and R. Koch, “Image distortions in stereoscopic video systems,” in Stereoscopic Displays and Applications IV, Proc. SPIE, vol. 1915, 1993, pp. 36-48.
[7] K. Ide and T. Sikora, “Adaptive parallax for 3D television,” in 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Tampere, Finland, 2010, pp. 1-4.
[8] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, November 2004.
[9] ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” Tech. Rep. BT.500-11, ITU-R, 2002.
[10] ITU-R, “Subjective assessment of stereoscopic television pictures,” Tech. Rep. BT.1438, 2000.
[11] S. Boslaugh and P. A. Watters, “Statistics in a nutshell,” O'Reilly Media, Inc., 2008, pp. 151-168.