Guidelines for Capturing High Quality Stereoscopic Content Based on a Systematic Subjective Evaluation

Di Xu, Lino Coria, Panos Nasiopoulos
Department of Electrical and Computer Engineering
The University of British Columbia
Vancouver, BC, Canada
{dixu, linoc, panos}@ece.ubc.ca

Abstract— **Although 3D TVs have already been introduced to the consumer market, the availability of stereoscopic (3D) content remains a serious challenge. Another challenge is that while some manufacturers are introducing 3D cameras and Hollywood is using proprietary solutions, there are no guidelines for consistently capturing high quality stereoscopic content. In this paper, we present a comprehensive stereoscopic image database with content captured at various distances from the camera lenses. We conducted subjective tests to assess the perceived 3D quality of these images which were shown on displays of different sizes. Finally, we provide guidelines of acquisition distances between the cameras and the real scene.**

Keywords- Stereoscopic, 3D, quality of experience, subjective test, capturing guidelines, mean opinion scores

I. INTRODUCTION

Many versions of 3D TVs and other devices that are capable of displaying stereoscopic content are presently available in the consumer market. Even though 3D movies are becoming extremely popular in theaters, the production of 3D content is technically challenging, in many cases leading to questionable quality and causing headaches and nausea to viewers. This is one of the main reasons that hold back the wide acceptance of 3D displays among consumers. In order to increase the quality of stereoscopic content, it is necessary to gain a better understanding of the technical and artistic challenges of this medium. There are a few empirically-obtained rules of thumb for capturing stereoscopic content that have been widely used by the 3D community [1]. However, it is necessary to perform a systematic evaluation to find which elements and parameters need to be considered in order to capture high-quality 3D content.

In order to identify design considerations for 3D-capable hand-held devices, subjective tests were performed in [2]. A test scene was developed in order to introduce controlled image points. This test scene provided accurate positioning of objects up to 5 meters from the camera and involved several small objects on top of a group of tables.

Subjective tests were also performed in [3] with the purpose of studying the influence of the acquisition parameters on the perceived quality of stereoscopic images. A 46-inch polarized stereoscopic display was employed for these tests.

Although the previous experiments provide useful recommendations for the creators of 3D content, it is still necessary to perform other tests that take into account other factors from both the capturing process and the displaying of stereoscopic content. For instance, it is desirable to consider in the testing process that the same 3D content will be seen on devices of many different display sizes that use different technologies for displaying 3D content (from 2.5'' autostereoscopic displays used on mobile phones to 80'' 3D TVs that employ active shutter glasses). The reason for this is illustrated in Fig. 1 where the same content is shown on a large display (Fig. 1a) and on a smaller one that is half the size of the first display (Fig. 1b). In Fig. 1, the blue squares (A_L, B_L) and red squares (A_R, B_R) indicate the pixels from the left view and right view, respectively, and the gray squares (A, B) denote the object position in 3D. Because the second display is smaller, the content occupies less space. The viewer's interpupillary distance remains constant while, generally, she gets closer to the screen. Because of this, the perceived depth is different for each display. In addition, more meaningful results could be obtained if the images employed for these tests resembled content that will actually be shown on 3D TVs (i.e., featuring people and objects in ordinary surroundings instead of an artificial lab setting).

In this study, we tested the effect that different distances (measured from the 3D camera setup to the photographed objects) have on the quality of the stereoscopically captured images. The test system employed allows us to identify the essential elements that need to be considered in order to obtain high-quality 3D images. We have created our own image database that is comprised of scenes depicting people and landscapes. Several viewers of different ages watched these images on displays of different sizes (a 2.8" 3D hand-held device, a 22" 3D computer monitor, and a 55" 3D TV) and rated them. We used the outcome of these subjective tests to produce a set of guidelines for capturing high-quality 3D content that will be watched on different size displays.

The rest of the paper is organized as follows. Section 2 describes the 3D content acquisition process. The subjective evaluation environment and parameters are specified in Section 3. In Section 4, we present the statistical analysis of the subjective test scores and discuss the findings of the tests. We conclude the paper in Section 5.

This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

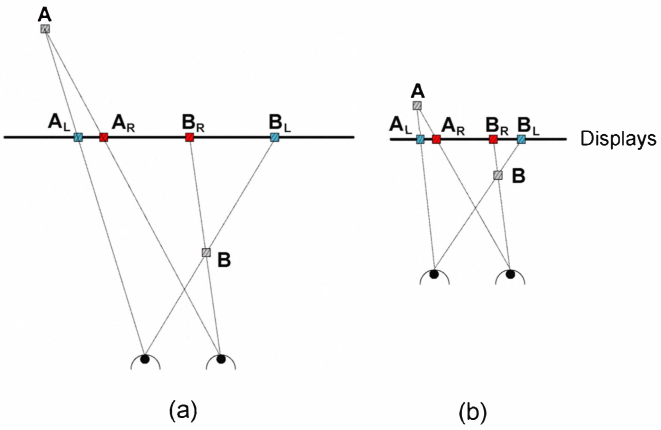


Figure 1. The same content is shown on (a) a large display and (b) a smaller. The perceived depth changes due to the size of the screen and the distance between the viewer and the screen.

II. ACQUISITION

A. Equipment

In order to capture stereoscopic images we employed two identical HD cameras (Sony HDR-XR500V 1080 60i NTSC) using the same firmware and settings. These cameras were aligned in parallel and attached to a bar that was specifically made for them. Subsequently, the bar was secured to a tripod as shown in Fig. 2. Since zoom lenses may differ [3], only the extreme ends of the zoom range were used. A single remote control was employed to obtain the best possible synchronization.

Figure 2. Stereo camera setup consisting of two identical HD camcorders.

B. Image Capturing

The stereoscopic image capturing process is illustrated in Fig. 3. Both cameras capture a slightly different image of the same event. Each event consists mainly of a person or object standing in front of the camera with a wall or a building as background. There are four important distances that need to be considered for every stereoscopic image pair:

- Distance between the two cameras (d_{cam}). Kept constant throughout the capturing process: 77 mm.
- Minimum distance (d_{min}). The distance between the cameras and the closest point captured in the stereoscopic image pair.
- Distance to main object (d_{obj}). The distance between the camera and the main object (usually a person). In most cases, $d_{obj} = d_{min}$.
- Maximum distance (d_{max}). The distance between the camera and the background. If the sky is visible, then d_{max} is considered to be infinity.

Some of the images were captured indoors and others were captured outdoors. We kept track of this information to see if viewers had any particular preference to lighting conditions.

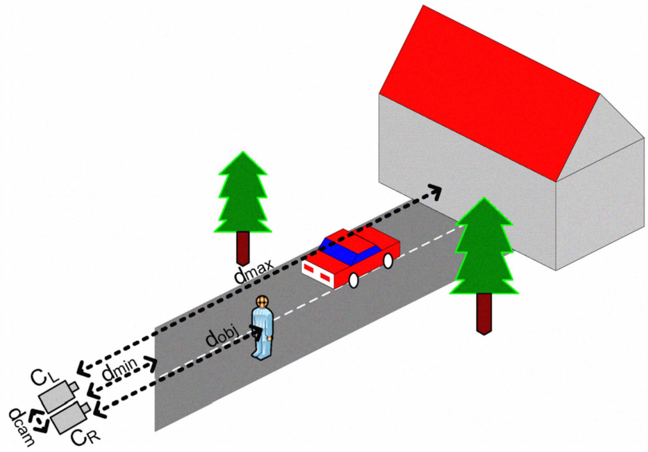


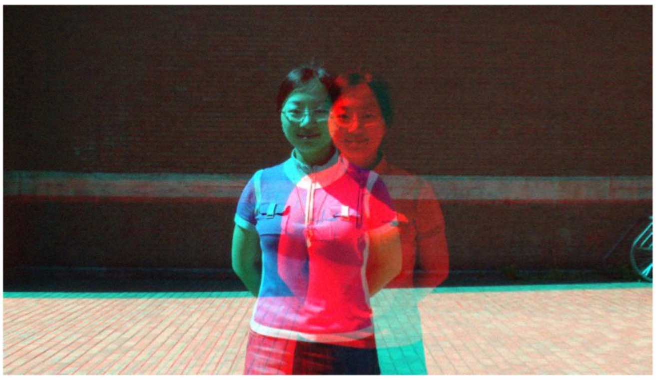
Figure 3. Capturing a live-action event with two parallel cameras CL and CR; dcam is the distance between the cameras, dmin is the distance from the cameras to the closest point, dobj is the distance from the cameras to the main object (usually a person), and dmax is the distance from the cameras to the background.

C. Image Alignment

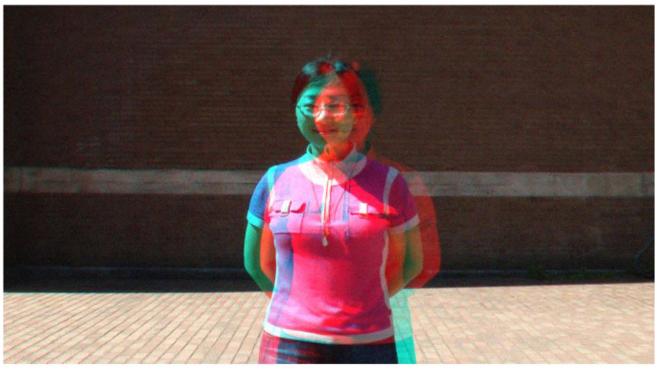
Basic image processing operations were performed on each stereoscopic pair in order to reduce the inaccuracies derived from the capturing process. Even though the cameras were carefully lined up, it is virtually impossible to do this task perfectly. Therefore, it is necessary to digitally align the image pair both vertically and horizontally.

Vertical alignment refers to vertically shifting one or both images so that there is no vertical parallax. Although the distance between the cameras, d_{cam}, was always kept at 77 mm, the horizontal parallax changed every time we put the cameras together (our database is the result of four image-capturing sessions). Therefore, we used $d_{min} = 3$ as reference, making sure that the horizontal parallax was always the same for this particular value of d_{min}.

The two larger displays have a 16:9 aspect ratio while the small one has a 4:3 aspect ratio. The original images on our database have an aspect ratio of 16:9. We cropped the left and right sides of each image to achieve a 4:3 aspect ratio for the small display. Although we could have kept the same aspect ratio for this display, this would have introduced black lines on the top and bottom of the display thus reducing the area of an already small screen. Fig. 4 shows an example of a stereoscopic pair before and after the alignment process (shown in anaglyph mode).



(a)



(b)

Figure 4. Two stereoscopic image pairs (presented in anaglyph mode): (a) without any vertical and horizontal shifting, and (b) after vertical and horizontal shifting as well as some cropping to preserve the 16:9 aspect ratio.

III. SUBJECTIVE EVALUATIONS

A. Displays

The subjective test was conducted on three different sizes of stereoscopic displays, namely, a 55” 3D LED TV (Samsung UN55C7000), a 22” 3D LCD display (Samsung 2233RZ), and a 2.8” Fujifilm 3D camera display. The first two displays are paired with different Samsung 3D Active glasses, and the Fujifilm display is an autostereoscopic display that can be viewed without glasses. The detailed specifications of the three stereoscopic displays are listed in Table I.

B. Database

The stereoscopic images that comprise the database include indoor and outdoor scenes, and various combinations of d_{min}, d_{obj}, and d_{max}, where d_{min} is in {0.5, 1, 2, 3} meters, d_{obj} is in {0.5, 1, 2, 3} meters, and d_{max} is in {5, 10, 50, infinity} meters.

C. Procedure

We set up the viewing conditions for the subjective assessment according to Section 2.1 of the ITU-R BT.500-11 [5], which is also recommended in ITU-R BT. 1438 [6]. Before the subjective evaluation of each display, we ran a training session to show to the subjects the quality range of our stereoscopic images, without imposing the quality of the images. Images used in the training session were different from the test images. Thirty test images were used for the subjective test and were shown in a random order. During the test, each stereoscopic image was shown for five seconds followed by a five-second interval of a 2D mid-grey image with the image index as a grading and relaxation period.

TABLE I. PROPERTIES OF THE 3D DISPLAYS USED IN OUR TEST

Size	Type	Resolution	Refresh Rate	Glasses
55”	Samsung TV	1920 x 1080	240Hz	3D shutter glasses
22”	Samsung monitor	1680 x 1050	120Hz	NVIDA GeForce 3D Vision glasses
2.8”	Fujifilm LCD screen	Approx. 230,000 dots	-- --	No glasses needed

IV. ANALYSIS AND RESULTS

A. Screening of the Outliers

Before analyzing the scores provided by the observers, we first screen the outliers according to the subjective scores they gave. The screening process is based on the guidelines provided in section 2.3.1 of annex 2 of ITU-R BT.500-11 recommendation [5].

One out of nineteen observers was detected as the outlier, and all the scores of this observer were eliminated. Therefore, all following data analysis is based on the scores provided by the eighteen valid observers.

B. Score Computation

We take the average score across all valid observers for each capture setting as the mean opinion score. To assess the credibility of the mean opinion score, we use confidence intervals to indicate the reliability of an estimate. The Student's t-tests are used to compute confidence intervals with the significance level being 95%.

C. Influence of Capture Parameters to Image Quality

1) Influence of lighting condition to image quality

For each set of capturing parameters, we included images that were captured indoors and outdoors for comparison. The statistical results show that there is no significant difference in image quality between images taken under different lighting conditions, i.e., indoor lighting and outdoor lighting on a sunny day. Therefore, in the following subsections, the mean opinion score of each capturing parameter set is the average score over the indoor and outdoor scenes.

2) Influence of d_{obj} to image quality

We compare the subjective quality between images taken at different d_{obj} when d_{obj} is the same as d_{min}. Fig. 5 shows the mean opinion scores and confidence intervals versus d_{obj} at different d_{max} distances. The same figure indicates that for the same d_{max}, the image quality increases with d_{obj} and levels off when d_{obj} is beyond two meters. The confidence intervals when d_{obj} is 0.5 meters are smaller than those when d_{obj} is large. In other words, the observers consistently provided low scores when the object of interest was very close to the cameras. Due to limited space, in Fig. 5 we show only the results based on scores from the 22 inches display. It is worth noting that the quality trend affected by d_{obj} is the same for the 2.8-inch and 55-inch displays.

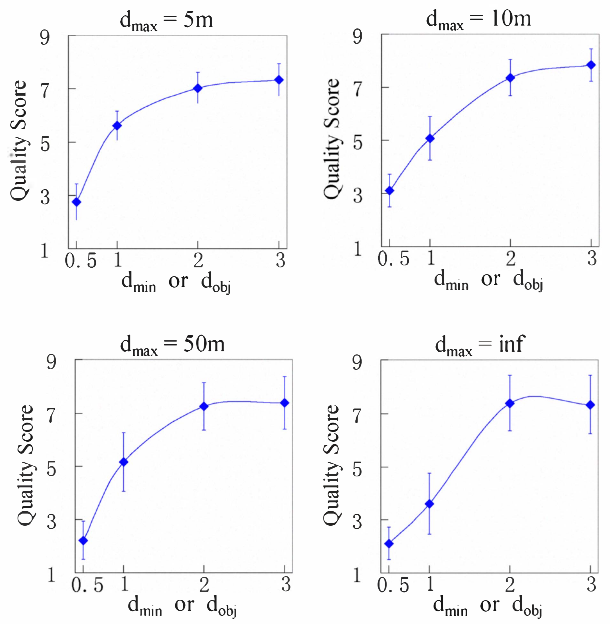


Figure 5. The mean opinion scores and their confidence intervals at various dobj (that is, 0.5m, 1m, 2m, and 3m). In reading order, the four subplots correspond to the cases when dmax are 5m, 10m, 50m, and infinity.

3) Influence of d_{max} to image quality

We compare the quality scores between images with different d_{max} while keeping the same d_{min}. No clear trend can be observed from the four subplots in Fig. 5. Thus d_{max} does not strongly affect the quality of 3D content. In general, the confidence intervals, however, seem to be smaller when d_{max} is small. Again, the same conclusions can be drawn for the 2.8-inch and 55-inch displays.

4) Influence of d_{obj} not being the foreground object to image quality

We tested a few images where the object of interest is not the closest object in the image, that is, when d_{obj} is greater than d_{min}. We compared image sets with the same d_{obj} and the same d_{max}, but various d_{min}. The mean opinion scores (score) and confidence intervals (CI) are listed in Table II (due to space limitations; figures would take much more space). Having compared the four sets of images, we observe that the quality of most images is impacted to certain extent when some background objects, such as floor and ceiling, appear closer to the cameras than the object of interest.

TABLE II. INFLUENCE OF DOBJ NOT BEING THE FOREGROUND OBJECT TO IMAGE QUALITY

d_{max} (m)	d_{obj} (m)	d_{min} (m)	Score	CI
10	2	2	7.36	0.68
		1	7.56	1.12
10	3	3	7.83	0.61
		2	7.22	0.94
		1	7.11	1.02
Inf	2	2	7.39	1.04
		1	5.44	0.86
Inf	3	3	7.33	1.09
		2	6.89	1.01

V. CONCLUSION

A subjective test was conducted to evaluate the influence of the capturing distances and display sizes on perceived 3D image qualities. Through careful analysis of the subjective results, we got a few useful conclusions.

- There is no impact of artificial or natural lighting conditions on 3D image quality. Indoor and outdoor images with the same parameters received very similar ratings.
- The strongest factor on 3D image quality is the distance between the camera setup and the closest photographed object. For our setting, d_{obj} should be at least 2 m.
- There is no apparent connection between the value of d_{max} and image quality.
- Foreground such as floor or grass appearing before the main object does not have a strong impact on the perceived quality of a 3D image.

These conclusions can be treated as guidelines for capturing stereoscopic content in the future.

REFERENCES

[1] Bernard Mendiburu, "3D Movie Making: Stereoscopic Digital Cinema from Script to Screen," Focal Press, 2009.

[2] Lachlan D. Pockett and Marja 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, 680306, (2008).

[3] Goldmann, Lutz ; De Simone, Francesca ; Ebrahimi, Touradj, "Impact of Acquisition Distortion on the Quality of Stereoscopic Images," International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona, USA, January 13-15, 2010.

[4] ITU-R, "Subjective assessment methods for image quality in high-definition television," Tech. Rep. BT.710-4, ITU-R, 1998.

[5] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. BT.500-11, ITU-R, 2002.

[6] ITU-R, "Subjective assessment of stereoscopic television pictures," Tech. Rep. BT.1438, 2000.