OPTIMIZING 3D IMAGE DISPLAY USING THE STEREOACUITY FUNCTION

Ming-Jun Chen[1], Do-Kyoung Kwon[3], Lawrence K. Cormack[2], Alan C. Bovik[1]

[1]Laboratory for Image and Video Engineering (LIVE), Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX, USA
[2]Department of Psychology, The University of Texas at Austin, Austin, TX, USA
[3]Systems and Applications R&D Center, Texas Instruments
12500 TI Blvd., Dallas, TX 75243

ABSTRACT

We develop an algorithm that predicts the best presentation of a stereo 3D image in the sense of viewers' preference. The algorithm operates in three steps. First, the 3D image is classified as either a "foreground dominant" or "background dominant" image. Next, for "foreground dominant" images, a model of the stereoacuity function is used to optimize the perceptual 3D resolution; for "background dominant" images, the nearest surface is placed in the 3D plane of the display screen. A human study was conducted to assess the algorithm and showed that the proposed model produced 3D images which had the best 3D quality scores among several candidate algorithms.

Index Terms— 3D image presentation, auto-convergence, 3D quality, quality of depth, stereo images,

1. INTRODUCTION

Many current cinematic producers include 3D effects as an extra incentive to attract larger audiences. In fact, the number of 3D films released in 2011 tripled compared to the number in 2008; more than forty 3D films were released in 2011 [2]. However, viewing a 3D film is not a pleasant experience for everyone and numerous complaints are reported. Uncomfortable 3D viewing experiences may be caused by a number of factors, including the method of 3D production, the viewing environment (3D display and viewing distance), and individual differences [3]. In this paper, the discussion will be focused on the 3D production part.

Most 3D content currently available is captured using a dual-camera configuration. There are two kinds of dual-camera settings: the parallel camera configuration and the toe-in camera configuration. Both have their own strengths and weaknesses. The toe-in camera configuration requires more knowledge to successfully shoot stereo videos, since the producer needs to decide the convergence point in depth during the shooting and post-processing is more difficult due to the keystone distortion. Therefore, the parallel camera configuration is still often used to capture videos especially in low-cost consumer cameras or smart phones. When using a parallel configuration, however, post-processing is required to enable binocular fusion of the stereo content and to avoid visual discomfort when viewing the captured 3D content.

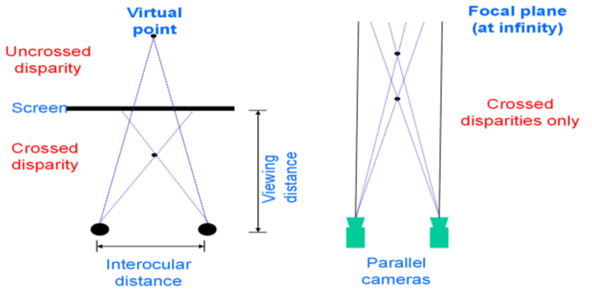

Fig. 1. Left: Illustration of cross and uncross disparity. Right: The parallel camera configuration.

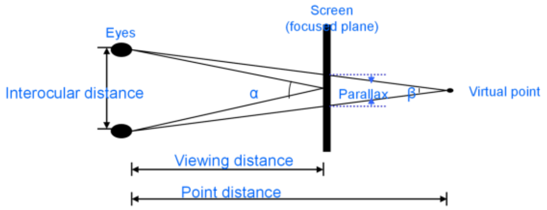

Fig. 2. Zone of comfortable stereo viewing.

The 3D content captured by a parallel camera configuration only creates crossed disparity values (shown in Fig. 1) which limits the amount of depth that can be presented in a stereoscopic display environment. Hence, post-processing is necessary to create uncrossed disparity values and to shift all disparity values within a range called as "zone of comfortable viewing. [4]" A number of studies have been conducted in this area, but the zone of comfortable viewing is not tightly defined; different numbers have been suggested by different authors. For example, Wopking [5] claimed that human subjects will not experience any discomfort when viewing a stereo 3D image if $|\alpha - \beta| \leq 1^{\circ}$ in Fig. 2, while smaller tolerances, such as $|\alpha - \beta| \leq 0.5^{\circ}$, have also been reported [3].

To improve the stereo viewing experience, we propose post-processing techniques that not only cause the disparities in a 3D presentation to fall within the zone of comfortable viewing but also deliver an optimal 3D viewing experience. Specifically, we seek to compute a best 3D presentation that delivers the most pleasant 3D viewing experience.

2. PRESENTATION MODEL

Our approach analyzes the content of a given 3D image in order to deliver a more pleasant 3D presentation by avoiding conflicts in 3D viewing and optimizing stereoscopic depth resolution. Human depth perception is affected by both monocular cues and binocular cues [6]. Conflicts between depth cues may create viewing discomfort or ambiguity in perceiving depths. Although it is not yet clear how the brain integrates these cues and produces a final sense of depths, it is rare to experience conflicting depth cues when viewing natural 3D images. Avoiding conflicts of depth cues by post-processing of the disparity values will help produce pleasant 3D viewing experiences.

In addition, when viewing a stereo 3D image on a stereo 3D display, the focused plane (accomodation of our eyes) is fixed on the screen. However, in our daily 3D vision, the accomodation of our eyes constantly changes as the vergence varies while scanning a 3D scene. Since the focal plane is fixed and the vergence plane may vary when viewing stereo 3D images on a display, the disconnect between accomodation and vergence may reduce the quality of depth percept.

2.1. Foreground/ background dominance classification

Naturally, the human eyes have a very wide field of view [7], and thus images that we see in our daily life are likely to be mostly “background dominant”. Hence, to avoid conflicts between depth cues, the composition of a stereo 3D image should be carefully considered as an integral part of post-processing its disparity values. For example, if a stereo image is deemed to be “background dominant”, then it should be disparity shifted so that it appears to be placed farther in the depth when displayed on 3D. Conversely, the presentation of a “foreground dominant” 3D image should be placed closer to the viewer.

To implement this idea, a foreground/background dominant classification process is needed. We have found two factors that can be used to sucessfully classify 3D images in this way: the skew of the disparity distribution, and Relative Dominant Depth (RDD) in the 3D image. The skew is computed as

$$skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})^2\right)^{\frac{3}{2}}}$$

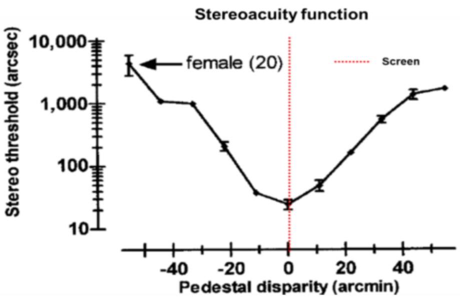

Fig. 3. Stereoacuity function, *s(d)* with 100 ms stimuli from [1].

where d_i is the disparity value of a pixel and $\bar{d}$ is the mean disparity of the 3D image. Then a 3D image is classified being “foreground dominant” if *skewness* > 1 and “background dominant” if *skewness* < −1. Images which have |*skewness*| < 1 either have a non-normal disparity distribution or cannot be classified by skewness. In this case, the RDD is used to force a classification, where

$$RDD = \frac{(dominant\ disparity - minimum\ disparity)}{(maximum\ disparity - minimum\ disparity)} .$$

and the $dominant\ disparity$ is the mode of the given disparity set. A 3D image is classified as “foreground dominant” if $|RDD| < \xi$ and “background dominant” if $|RDD| > \xi$, where $\xi = 0.25$ in this work.

2.2. Maximizing depth resolution

Krekling [8] showed a minimum degree of variation in disparity that is needed for the human vision system to perceive different depths between objects. This is called the stereo threshold. Studies [1, 6] have shown that the lowest thresholds are generally obtained at a zero pedestal disparity, and the threshold increases with increasing crossed or uncrossed pedestal disparity. The function which provides the stereo threshold at different disparities is called the stereoacuity function [1]. Fig. 3 shows the stereoacuity function of a female subject having normal stereo vision and her minimal threshold disparity is 24 arcsec at zero disparity. One can see clearly that human stereoacuity is most sensitive at the focus plane (the viewing screen in the 3D viewing of stereo images) and this observation indicates that human vision system has the highest depth resolution for objects around zero disparity.

Consider the case in which two objects have a relative disparity of 120 arcsec, but an average disparity of zero. (i.e., they have disparityties of +1 and -1 arcmin, respectively) The subject, who has the stereoacuity function shown in Fig. 3, should see these two objects as laying in different depth planes. Now consider the case in which they have the same relative disparity, but one has a pedestal disparity of +40 arcmin and thus the other has a pesdestal disparity of +42 arcmin. In this case, the dispairty between them is below threshod, and no relative depth will be perceived. Hence, we claim that the subject can see depth with better resolution when the two objects are arranged around the zero pedestal disparity. To quantify the ability to resolve depth, we approximate the negative of the stereoacuity function, i.e. 1 – *s(d)*, in terms of the pixel disparity with a Gaussian function, and call it the "depth resolution function" in this work. Then we solve the problem of optimizing the 3D presentation of a 3D image by maximizing the perceived depth resolution. This operation can be expressed by

$$opt\,shift = \underset{-255<i255}{\operatorname{argmax}}\ DRF \cdot Hist(i)$$

where the *DRF* is the depth resolution function using a σ = 20 arcmin, which is chosen to give the best fit to the stereoacuity function. *Hist(0)* is the histogram of the disparity of a 3D image without being post-processed and *Hist(i)* is the histogram of the disparity of a 3D image shifted by *i*. Based on this operation, the shift value that yields the maximum product (i.e. *opt shift*) is deemed to provide a best 3D viewing experience in depth.

2.3 Optimizing the presentation

The proposed algorithm is processed by the following steps:

1. Classify an input 3D image into either foreground or background dominant image as described in Sec. 2.1.
2. For the foreground dominant image, as described in Sec. 2.2, find the shift value that yields the maximum product of the depth resolution function and the disparity histogram.
3. For the background dominant image, find the shift value that places the closest surface on the screen.
4. Shift the left and right images so that the resulting 3D image has the desired depth according to the shift value found in Step 3 or Step 4. Then crop the undefined pixels on the boundary. For example, after an image is shifted to left by 3 columns, there are three undefined columns on the right side of the image.

The overall algorithm is described in Fig. 4.

3. HUMAN STUDY

A human study was conducted to assess the above algorithm. The study is described next.

3.1. Study design

A double stimulus continuous quality scale (DSCQS) protocol [9] was adopted to obtain subjective 3D quality ratings on all of the stimuli in the work. During a single trial, a subject compared two 3D images with different depth relative to the screen and gave both of them a subjective 3D quality score based on his/her preference. The question given to subjects is "Give a 3D viewing quality rating" and it is a forced-choice procedure so that the subjects could not rate both images equal 3D quality scores. A training session was also given to each subject at the beginning of the study to familiarize them with the Graphic User Interface (GUI) of our study program. The training content was different from the images used in the study. Repeated viewing of the same 3D image was allowed before the subject gave a rating.

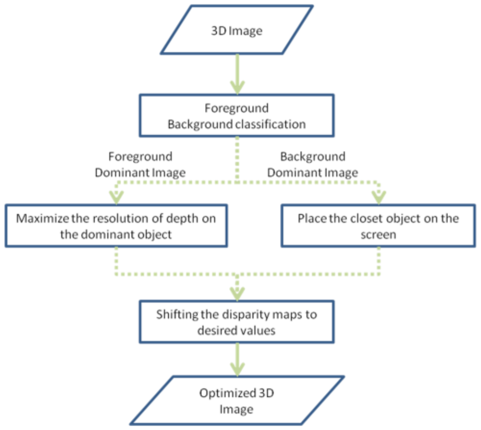
3D Image
Foreground
Background classification
Foreground
Dominant Image
Background
Dominant Image
Maximize the resolution of depth on the dominant object
Place the closet object on the screen
Shifting the disparity maps to desired values
Optimized 3D
Image

Fig. 4. Flowchart of the proposed algorithm.

3.2. Display

An nVidia active 3D kit plus an Alienware OptX AW2310 full HD 3D monitor were used to display the 3D images. The viewing distance from subjects to screen is five times the screen height to minimize potential visual discomfort caused by the accommodation-vergence conflict.

3.3. Observers

Seventeen naïve observers (seven females and ten males) were recruited for the study. The subjects were pre-screened to ensure normal stereovision by asking them to distinguish the depth of three colored rectangles separated from each other by 6 arcmin in depth.

3.4. Stimuli

Twelve stereo images with ground truth disparity were chosen as source images. Seven of these stereo images were captured by the parallel camera configuration from the Middlebury stereo database [10], and five of them were artificial 3D stereo images (three were from MPEG 3D coding test videos, two were created by the authors). We used an approximately equal number of "foreground dominant" and "background dominant" image. The original resolution of the images was equal to or larger than full HD size and they were resized to full HD resolution by cropping the extra part.

To create a baseline without ground truth depth, the reference image was created by placing the

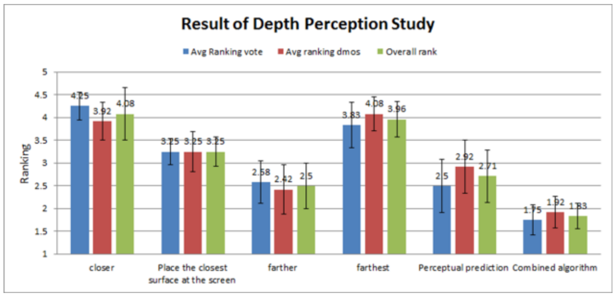

Fig. 5. Performance of different 3D image presentation strategies. Error bars represent the standard errors.

closest surface inside the image at the depth of the screen. Then, four different stimuli were created by either pulling the scene in front of screen or by pushing it deeper into the screen by shifting disparities. The distance of two adjacent stimuli is 13.7 arcmin. In addition, one scene was created by maximizing the depth resolution, as described in Session 2.2. Finally, all stimuli have disparities that satisfy the "zone of comfort viewing" suggested by [3].

4. RESULTS AND ANALYSIS

Differential Mean Opinion Score (DMOS) are usually used as quality scores annotated to the content in image quality database. However, all of the stimuli in this study are pristine images, so comparing depth quality among stimuli which have different content is meaningless. On the contrary, intra-content comparisons can provide insights regarding the best perceptual 3D depth range of a stereo pair. Hence, the average ranking given by human subjects is proposed as a criterion to evaluate the performance of 3D images displayed with different (shifted) depth ranges.

Six different profiles were used for each source image. The ranking of each source image ranges from 1 (the best) to 6 (the worst). The performance achieved by a 3D presentation is represented by the average ranking over twelve source images. Two types of rankings were used. The first is "ranking DMOS (weighted ranking)" which is the ranking sorted by DMOS scores. The second is "ranking vote", which only considers binary decisions (stimulus A gets one vote if one subject prefers stimulus A over stimulus B) and the ranking is sorted by the voting results. The overall ranking is the average of these two rankings.

The experimental results are shown in Fig. 6. The "closer" profile is to set the disparity value of the closest surface at -13.7 arcmin (crossed disparity) and the disparity value of the closest surface for "farther" and "farthest" profiles are 13.7 arcmin and 27.4 arcmin respectively. Four observations can be made from Fig. 6. First, the reference strategy, which places the closest surface on the screen, has a ranking of 3.25. This ranking is slightly better than the expected ranking (3.5) when the nearest surface is placed randomly inside the zone of comfortable viewing. Second, comparing the "closer" and "farther" profiles, we observe when extra computation is not allowed, the better strategy is to push the closest surface deeper into the screen rather than pull it out of the screen. Third, the strategy which maximizes the depth resolution performs better than the reference strategy, but worse than the "farther" profile. A plausible explanation is that this strategy works for "foreground dominant" 3D images, but creates depth cue conflicts for "background dominant" 3D images. Finally, the proposed algorithm which applies both of strategies based on content gives the best performance (overall ranking is 1.83).

5. CONCLUSION AND FUTURE WORKS

We believe that the degree of comfort in viewing a 3D image as a function of the depth range it is assigned is correlated with the stereoacuity function of the human visual system and the content of the 3D image. A human study was conducted which supports our argument. The following points can be further considered. First, there should be a better strategy in post processing "background dominant" images. Our current strategy is simply to place the closest surface at the screen for the background dominant images. Second, other content-related factors such as object contours and the composition of a 3D image may affect the perception of a stereo 3D image.

6. REFERENCES

[1] C. M. Zaroff, M. Knutelska, and T. E. Frumkes, "Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender," *Investigative Ophthalmology & Visual Science*, vol. 44, pp. 891-900, February 1 2003.

[2] *List of 3D movies*. Available: http://en.wikipedia.org/wiki/List_of_3-D_films

[3] M. T. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: a review " *Proc. SPIE*, vol. 6490, 2007.

[4] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *Journal of Vision*, vol. 11, July 21 2011.

[5] M. Wopking, "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus," *J. Soci. Inform. Display*, vol. 3, pp. 101-103, 1995.

[6] I. P. Howard and B. J. Rogers, *Binocular vision and stereopsis*. New York: Oxford University Press, 1995.

[7] Y. Le Grand, *Light, colour and vision*. London: Chapman & Hall, 1968.

[8] S. Krekling, "Stereoscopic threshold within the stereoscopic range in central vision," *Am. J. Optom. Physiol. Optic.*, vol. 51, pp. 626-34, 1974.

[9] I. T. U., *Methodology for the subjective assessment of the quality of television pictures*, 2003.

[10] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Journal of Computer Vision*, vol. 47, pp. 7-42, 2002.