3D VISUAL DISCOMFORT PREDICTION BASED ON PHYSIOLOGICAL OPTICS OF BINOCULAR VISION AND FOVEATION

Taewan Kim, Jincheol Park, Sanghoon Lee, and Alan Conrad Bovik ∗

Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea.
{enoughrice21, slee}@yonsei.ac.kr
* Department of Electrical & Computer Engineering, The University of Texas at Austin, USA
{bovik@ece.utexas.edu}

ABSTRACT

For clear binocular vision in the brain, the neural interaction for accommodation and vergence is performed via inter-operation of two cross-links: accommodative-vergence (AV) and vergence-accommodation (VA). However, when people watch stereo images on stereoscopic display, the neural operation is interrupted due to alternations of the cross-link gains, which are attributed as the main reason of visual discomfort on stereo images in terms of Accommodation-Vergence Mismatch (AVM). In this paper, we present a novel visual discomfort prediction algorithm dubbed 3D-AVM Predictor to quantify the visual discomfort by including neural responses to sensory stimuli of both accommodation and vergence. In particular, we define the 3D local bandwidth (BW) based on physiological optics of binocular vision and foveation for optical activity of accommodation from the perspective of the VA cross-link. Since the 3D-AVM Predictor encompasses the anomalous motor responses of both accommodation and vergence, it is demonstrated that the performance is statistically superior to those of conventional works which rely on the factor of disparity distribution only.

I. Introduction

Stereoscopic three-dimensional (3D) multimedia service provides a realistic immersive experience, so that it is gaining an increasing amount of attention. Nevertheless, the stereoscopic immersion could accompany visual discomfort since the 3D percept of stereo images is rendered on the planar type of stereoscopic displays and hence the accommodation-vergence mismatch (AVM) occurs [1]-[7]. Thus, the stereo images could induce anomalous binocular vision, resulting the visual discomfort. To guarantee human safety, it is necessary to predict the visual discomfort of stereo images by quantifying the anomalous binocular vision.

The vergence and accommodation processes interact with each other in the brain to obtain clearer single binocular vision. A change in retinal blur elicits the vergence eye movement through the accommodative-vergence (AV) crosslink, although the vergence eye movement is primarily driven by retinal disparity [1]. Likewise, retinal disparity also changes accommodative position through the vergence-accommodation (VA) cross-link, although retinal blur induces accommodation as a reflex.

Unlike natural viewing condition, there is a limitation to realize the natural depth attaining various focus planes of natural scene to a stereoscopic display having only one focus plane. It leads to inducing four kinds anomalous effects of binocular vision: alternation of AV/A ratio, alternation of VA/V ratio, absence of de-focus blur and absence of differential blur, as shown in the stereoscopic display viewing. Therefore, the alternations of AV/A and VA/V ratios have negative consequences for natural binocular vision, resulting in visual discomfort of stereo images. The absences of de-focus and differential blurs induce accommodation cue conflicts at peripheral point causing visual discomfort. In addition, the alternation of AV/V ratio influences on the activity of vergence, while the alternation of VA/V ratio and the absences of de-focus and differential blurs influence on the activity of accommodation.

We propose the 3D-AVM Predictor utilizing three features of alternation of VA/V ratio, absence of defocus blur and absence of differential blur from the 3D local BW map. In addition, one more feature of alternation of AV/A ratio is extracted from the disparity map. The combination of the four features is then used to create a quality model by using a support vector regression (SVR) [8]. Finally, the visual discomfort of stereo images is predicted by using the quality model.

II. Human Perception of 3D Space

A. Physiological Optics of Accommodation

When a viewer focuses on a fixation point in depth, the eyes accommodate to the fixation point by a variety of

S. Lee is corresponding author. (Tel: +82-2-2123-2767, Fax: +82-2-313-2879, E-mail: slee@yonsei.ac.kr)
This research was funded by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2014.

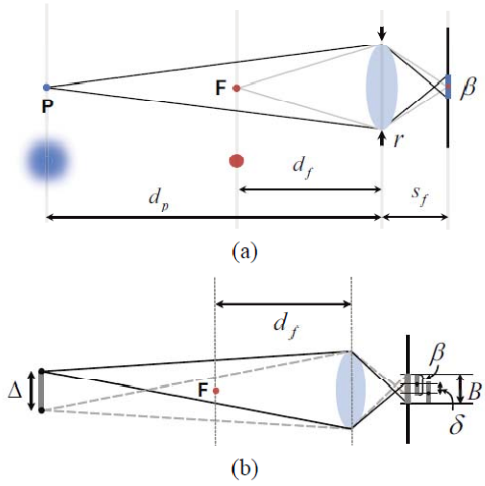

Fig. 1 Geometry of blurred pixel width. (a) Blur circle of diameter β caused by de-focus blur at a peripheral point P when a human eye is fixated at point F. (b) Blurred pixel width when focusing at another point F.

anatomical changes. Once the shape of the crystalline lens is formed by the accommodation process, the optics of the eye can be modeled by the thin-lens equation:

$$\frac{1}{s_f}+\frac{1}{d_f}=\frac{1}{f} \tag{1}$$

where f is the focal length, s_f is the posterior nodal distance, and d_f is the distance in depth from the nodal point to the fixation point.

As depicted in Fig. 1 (a), when the eye accommodates to a fixation point F at distance d_f, the fixation point is projected on the retina in focus, where f varies with accommodation and r is the pupil size. However a peripheral point P at a distance d_p in depth from the fixation point projects to an out of focus, blurred image on the retina. The amount of blur is often expressed as a blur circle diameter, induced from (1) by ray tracing [9]:

$$\beta = r \cdot s_f \left|\frac{1}{d_f}-\frac{1}{d_p}\right|. \tag{2}$$

As shown in Fig. 1 (b), when the eye focuses on a pixel of width Δ at distance d_p, then it projects onto the retina with width δ. Since a ray of light which passes through the optical center of the lens proceeds without any refraction, the projected pixel width on the retina is $\delta = \Delta\frac{s_f}{d_p}$. However if the eye focuses at point F as depicted in Fig. 1 (b), the projected pixel width increases owing to the de-focus blur depicted in Fig. 1 (a). The total width of the blurred pixel is obtained as the sum of the blur circle diameter β and the projected pixel width δ, $B = \beta + \delta$.

A visual angle of 1° subtends $\lambda = s_f\tan(\frac{\pi}{180})$, on the retina. Then define the effective pixel resolution (EPR) to be the number of blurred pixels that fall within a visual angle of 1° on the retina, $p = \lambda/B$ (pixels / degree). This calculation is equivalent to the display visual resolution given in [10]. According to the Nyquist sampling theorem,

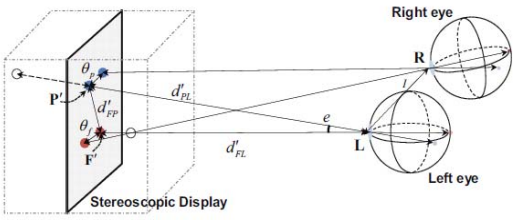

Fig. 2 3D space rendered on a stereoscopic display; the fixation point F and peripheral point P are necessarily projected onto the retinae from the same frontal display plane.

the maximum frequency that can be represented without aliasing is half the sampling frequency. Then, regarding the EPR as sampling frequency, the local accommodation BW can be expressed in cycles/degree:

$$f_B = \frac{p}{2} = \frac{1}{2}\tan(\frac{\pi}{180})\cdot(\frac{\Delta}{d_p}+r\left|\frac{1}{d_f}-\frac{1}{d_p}\right|)^{-1}. \tag{3}$$

B. Foveation in 3D space

The density of photopic retinal photo-receptors (cones) rapidly decreases with distance away from the fovea [11]. The authors of [11] define a critical frequency beyond which visibility deteriorates as a function of eccentricity, e, in cycles/ degree:

$$f_F(e) = \frac{e_2 \ln(\frac{1}{CT_0})}{\alpha[e_2+e]} \tag{4}$$

where $CT_0 = 1/64$ is a minimum contrast threshold, $\alpha = 0.106$ is a spatial frequency decay constant, and $e_2 = 2.3$ is the half-resolution eccentricity. In order to apply this foveation model to perceived 3D space and stereoscopic display viewing, we calculate eccentricity e using the 3D geometry shown in Fig. 2.

Using the coordinate transformation on the stereoscopic display coordinates in Fig. 2, the distances of F and P from left eye are $d'_{FL} = \|F' - L\|$ and $d'_{PL} = \|P' - L\|$ where $\|\ \|$ denotes Euclidean distance. The distance between F0 and P0 on the stereoscopic display is $d'_{FP} = \|F' - P'\|$.

If F' and P' have pixel disparities θ_f and θ_p on the stereoscopic display, respectively, then the distances from L to F and P in 3D space are

$$d_{FL} = \begin{cases} \frac{I}{I+\theta_f}\cdot d'_{FL}, & \theta_f \le 0 \\ \frac{\theta_f}{I-\theta_f}\cdot d'_{FL}, & \theta_f > 0 \end{cases} \tag{5}$$

and

$$d_{PL} = \begin{cases} \frac{I}{I+\theta_p}\cdot d'_{PL}, & \theta_p \le 0 \\ \frac{\theta_p}{I-\theta_p}\cdot d'_{PL}, & \theta_p > 0 \end{cases} \tag{6}$$

where negative disparity ($\theta_f < 0$ and $\theta_p < 0$) denotes crossed disparity and positive disparity ($\theta_f > 0$ and $\theta_p > 0$) denotes un-crossed disparity.

Finally, the eccentricity can be expressed:

$$e(P'|F') = COS^{-1}(\frac{{d'_{FL}}^2+{d'_{PL}}^2-{d'_{FP}}^2}{2d'_{FL}d'_{PL}}). \tag{7}$$

C. 3D Local Bandwidth

The local accommodation BW (3) presents the perceptual Nyquist frequency on the retina, which is determined by a width of projected pixels based on de-focus blur. The local BW of foveation in (4) presents the perceptual capacity on the retina, which is determined by a density of photo-receptors. Given a fixation point F, we obtain the local 3D BW at a peripheral point P by taking the smaller value of the local BWs of accommodation (3) and foveation (4):

$$W_{Real}(P|F) = \min(f_B(P|F),\ f_F(P|F)) \quad (8)$$

where $f_B(P|F)$ and $f_F(P|F)$ are the local accommodation BW and foveation BW at P given F, respectively. In other words, $f_B(P|F)$ is the local accommodation BW at the distance d_{PL} given the distance d_{FL}. When viewing through the stereoscopic display, the eye accommodates to F' and the peripheral point is actually projected from P'. Thus, the 3D local BW in the stereoscopic display is

$$W_{Display}(P'|F') = \min(f_B(P'|F'),\ f_F(P'|F')) \quad (9)$$

III. 3D Accommodation and Vergence Mismatch Predictor

We extract features from the conflicts of vergence (D_{av}) and accommodation (D_{va}, D_b and D_d) cues on a stereoscopic display. The features of D_{av} is simply extracted by calculating statistics of a disparity map, since the activity of convergence is well represented in the disparity map. In order to extract the features of D_{va}, D_b and D_d, we define three kinds of deviations of 3D local BWs between real 3D space and the stereoscopic display. Assuming that all the points in the image are candidates of fixation point, we calculate the deviations for each fixation point.

A. Featuring of Alternation of the VA/V Ratio

If a fixated object has a disparity on the stereoscopic display, the 3D local BW of the stereoscopic display is deviated from that of real 3D space. Thus, we calculate D_{va} by the deviation between the 3D local BWs of real 3D space and the stereoscopic display at the fixation points:

$$D_{va} = f_B(F|F) - f_B(F'|F') \quad (10)$$

where $f_B(F|F)$ and $f_B(F'|F')$ are defined by the local BWs of accommodation of real 3D space and the stereoscopic display at F and F', given F and F', respectively.

B. Featuring of Absence of De-focus Blur

In real 3D space, the fixation point is not only imaged on the retina, but the peripheral regions are also imaged with reduced 3D local BW due to de-focus blur and the reduced retinal sampling frequency of photo-receptors. Thus, the local 3D BW of viewed 3D space is affected by both accommodation and foveation. On the other hand, in the 3D local BW of stereoscopic display, there is no effect of depth, but only the 2D spatial foveation. In order to quantify this deviation, we obtain the absolute difference of 3D local BW between real 3D space and stereoscopic display for a given fixation point as:

$$E_b(P'|F') = |W_{Real}(P|F) - W_{Display}(P'|F')|. \quad (11)$$

where $f_B(F|F)$ and $f_B(F'|F')$ are defined by the local BWs of accommodation of real 3D space and the stereoscopic display at F and F', given F and F', respectively.

Since the region of large 3D local BW deviation has a influence on the visual discomfort and the worst local deviation has a dominant effect on the overall visual discomfort, D_b is obtained by averaging the upper p^{th}-percentile of the values of $E_p(n)$:

$$D_b = \frac{s}{N_p}\sum_{n \geq N_d \cdot p/100} E'_b(n), \quad (12)$$

where $E'_b(n)$ are the sorted values (order statistics [12]) of $E_b(n)$, $N(d)$ is the total number of $E_b(n)$, and $N(p)$ is the number of values larger than the p^{th}-percentile[2].

C. Featuring of Absence of Differential Blur

In a stereoscopic display, since all pixels have the same focal distances, there is absence of differential blurs between the fixation point and other regions. However, there is still the difference of foveation blur in the stereoscopic display. In order to calculate D_d, we obtain the differential blurs for real 3D space ($\Phi_R(P|F)$) and stereoscopic display ($\Phi_D(P|F)$), respectively, by using 3D local BW:

$$\Phi_R(P|F) = |W_{Real}(F|F) - W_{Real}(P|F)|, \quad (13)$$

$$\Phi_D(P'|F') = |W_{Real}(F'|F') - W_{Real}(P'|F')|, \quad (14)$$

The deviation of differential blurs is then calculated as $E_{d(P')} = |\Phi_R(P|F) - \Phi_D(P'|F')|$. Therefore, D_d is obtained by

$$D_d = \frac{s}{N_p}\sum_{n \geq N_d \cdot p/100} E_d(n), \quad (15)$$

D. Feature used in the 3D-AVM Predictor

In order to estimate visual discomfort of a stereo image, we extract four kinds of features from each map of D_{va}, D_b and D_d. In addition, we also extract the four kinds of features from disparity map for D_{av}. Thus, the total number of features is sixteen. Next, we adopt the SVR to create a quality model capturing the connection between the extracted features and the subjective ratings of visual discomfort.

Since the different signs of disparity have different impacts on the visual discomfort, we separately extract the first and second features from the mean values of positive and negative disparities, respectively:

$$f_1 = \frac{1}{N_{pos}}\sum_{M(n)>0} M(n), \quad (16)$$

$$f_2 = \frac{1}{N_{neg}}\sum_{M(n)\leq 0} M(n), \quad (17)$$

where $M(n)$ is the n^{th} smallest value in a map, and N_{pos} and N_{neg} are the numbers of positive and negative values in a map, respectively. The third and fourth features are lower and upper percentiles, respectively:

[2] 1In our implementation, we used p =95 [13].

$$f_3 = \frac{1}{N_l}\sum_{n \le N_M \cdot p_l/100} M(n), \quad (18)$$

$$f_4 = \frac{1}{N_l}\sum_{n \ge N_M \cdot p_l/100} M(n), \quad (19)$$

where N_M is the total number of values in a map, and p_l and p_h are the lower and upper percentiles, respectively, and N_l and N_h are the number of values lower and higher than p_l and p_h, respectively. In our implementation, we used $p_l = 5$ and $p_h = 95$.

IV. 3D Statistical Performance Evaluation

The 3D-AVM Predictor was tested on the IEEE-SA stereo image database [14], which consists of 800 stereo images, along with associated MOS. Each image has high-definition (HD) resolution (1920x1080 pixels). Twenty five subjects ranged from 24 to 31 years participated in the subjective study. Moreover, all subjects were tested and found to have good or corrected visual acuity and good stereoscopic acuity [6].

Table I Mean LCC over 2000 trials of randomly chosen and test sets on the IEEE-SA database.

	Mean	Median	Std.
Nojiri [2]	0.6854	0.6935	0.0788
Yano [3]	0.3988	0.4045	0.0748
Choi [4]	0.6509	0.6565	0.0783
Kim [5]	0.7018	0.7113	0.0771
3D-AVM	0.8604	0.8672	0.0482

Table II Mean SROCC over 2000 trials of randomly chosen and test sets on the IEEE-SA database.

	Mean	Median	Std.
Nojiri [2]	0.6108	0.6155	0.0732
Yano [3]	0.3363	0.3384	0.0732
Choi [4]	0.5851	0.5909	0.0798
Kim [5]	0.6151	0.6195	0.0700
3D-AVM	0.7831	0.7882	0.0451

To implement the SVR, we used the libSVM package [15] with the linear kernel, whose parameters were estimated by cross-validation during the training session. Through the experiment, the database was exclusively separated into 80% and 20% for the training and test sets, respectively, which are completely content-separate. The training and testing subsets did not overlap in content. In order to ensure that the results were not built on a specific train-test separation, we repeated tests and evaluations 2000 times by randomly dividing training and testing sets.

We tested the 3D-AVM Predictor and compared it with prior works [2]-[5]. We used the Spearman rank order correlation coefficient (SROCC) and the Pearson linear correlation coefficient (LCC) between the predicted and subjective scores for the evaluation of the 3D-AVM metric, as tabulated in Tables I and II. As shown in both tables, 3D-AVM delivered the highest LCC and SROCC values and the minimum standard deviation values by a substantial margin.

V. Conclusions

The proposed mechanism named 3D Accommodation Vergence Mismatch (3D-AVM) Predictor extracts two kinds of features representing conflicts of accommodation and convergence cues. The feature of convergence cue is easily extracted from disparity maps. Whereas, in order to extract the features of accommodation cue, we calculated the 3D local bandwidth (BW) based on physiological optics and foveation. Thus, the method presented here, while accurate and valuable, is only one piece of a broader puzzle that confronts 3D vision and video researchers.

5. REFERENCES

[1] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," J. Imaging Sci. Technol., vol. 53, no. 3, pp. 030201.030201-14, May 2009.

[2] Y. Nojiri, H. Yamanoue, A. Hanazato, and F. Okano, "Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV," in Proc. Stereoscopic Displays Virtual Reality Syst. X, vol. 5006, pp. 195-205, 2003.

[3] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," Displays, vol. 23, no. 4, pp. 191-201, June 2002.

[4] J. Choi, D. Kim, S. Choi and K. Sohn, "Visual fatigue modeling and analysis for steroscopic video," Opt. Eng., vol. 51, no. 1, pp. 017206.017206- 11, Jan. 2010.

[5] D. Kim and K. Sohn, "Visual fatigue prediction for steroscopic image," IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 2, pp. 231-236, Feb. 2011.

[6] T. Kim, J. Kang, S. Lee and A. C. Bovik, "Multimodal interactive continuous scoring of subjective 3D video quality of experience," IEEE Trans. Multimedia, vol. 16, no. 2, Feb. 2014.

[7] K. Lee, A. K. Moorthy, S. Lee and A. C. Bovik, "3D Visual activity assessment based on natural scene statistics," IEEE Trans. Image Processing, vol. 23, no. 1, pp. 450-465, Jan. 2014.

[8] B. Scholkopf, A. Smola, R.Williamson, and P. Bartlett, "New support vector algorithms," Neural Computat., vol. 12, no. 5, pp. 1207-1245, 2000.

[9] D. M. Hoffman and M. S. Banks, "Focus information is used to interpret binocular images," Journal of Vision, vol.10, no. 5, May 2010.

[10] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," IEEE Trans. Image Processing, vol. 6, no. 8 pp. 1164-1175, Aug. 1997.

[11] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," Proc. SPIE, vol. 3299, 1998.

[12] H. G. Longbotham and A. C. Bovik, "Theory of order statistic filters and their relationship to linear FIR filters," IEEE Trans. Acoust., Speech, Singal Process., vol. 97, no. 2, pp. 275-287, Feb. 1989.

[13] J. Park, K. Seshadrinathan, S. Lee and A. C. Bovik "VQPooling: Video Quality Pooling Adaptive to Perceptual Distortion Severity," IEEE Trans. Image Process., vol. 22, no. 2, pp. 610-620, Feb. 2013.

[14] Standard for the Quality Assessment of Three Dimensional (3D) Displays, 3D Contents and 3D Devices based on Human Factors, IEEE P3333.1, (http://grouper.ieee.org/groups/3dhf), 2012.

[15] C. Chang and C. Lin, "LIBSVM: A Library for Support Vector Machines, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm/.