A 3D Visual Comfort Metric Based on Binocular Asymmetry Factor

1st Feng Qi
Key Laboratory of Intelligent Information Processing
Institute of Computing Technology, CAS
Beijing, China
feng.qi@vipl.ict.ac.cn

2nd Tingting Jiang
Institute of Digital Media, EECs
Peking University
Beijing, China
ttjiang@pku.edu.cn

3rd Jian Zhang
ShenZhen Graduate School
Peking University
Shenzhen, China
jianzhang.tech@gmail.com

4th Huizhu Jia
Institute of Digital Media, EECs
Peking University
Beijing, China
hzjia@pku.edu.cn

5th Xilin Chen
Key Laboratory of Intelligent Information Processing
Institute of Computing Technology, CAS
Beijing, China
xlchen@ict.ac.cn

***Abstract*—How to evaluate 3D visual discomfort is a challenging problem in stereoscopic image quality assessment. This is due to an existing gap between human binocular visual perception and current stereoscopic image representation techniques. As an indicator of binocular images' relationship for stereoscopic image, binocular asymmetry has been found that it is one of the most important discomfort-induced factors. Based on the factor, this paper proposes an objective stereoscopic visual comfort assessment (SVCA) model for stereoscopic images. Specifically, binocular asymmetry is interpreted as the two views' image texture features which are represented by the histograms of oriented gradient (HOG) feature and local binary pattern (LBP) feature. The HOG/LBP are integrated into an overall visual comfort score by support vector regression (SVR). Two stereoscopic image databases are chosen to evaluate the applicability of the proposed metric. The experimental results show that the proposed SVCA metric can efficiently predict visual discomfort for stereoscopic image.**

***Index Terms*—visual discomfort, stereoscopic visual comfort assessment, binocular asymmetry, support vector regression, HOG, LBP**

I. Introduction

The prosperity of 3D consumption depends upon the rapid development of stereo/3D technologies, *e.g.* 3D formats, 3D production, 3D video coding, and 3D displays. However, certain physiological symptoms, such as visual discomfort, eyestrain, headache, and dizziness, induce a completely novel problem in stereo/3D technologies. Although no evidence has confirmed permanent damage to eyesight or health from watching 3D images/videos, these physiological discomforts hamper the popularity of 3D applications and therefore have gained much research interest [1]-[24]. Different with the traditional distortion assessment, 3D content assessment focuses on the visual quality of stereoscopic image such as quality of experience, visual discomfort, unnatural experience [2] [3]. Since its primary goal is to minimize or prevent discomfort caused by 3D content, stereoscopic visual comfort assessment (SVCA) is more important in the research of 3D technologies.

Researchers have found that several factors may induce visual discomfort, including excessive screen disparity [4], accommodation-vergence conflict [5], binocular asymmetry [6], range of depth of focus (DOF) [7], vertical disparities caused by geometric distortion [8] and crosstalk artifacts [9]. Based on these findings, several objective SVCA metrics have been proposed. Nojiri *et al.* [10], Jung *et al.* [11], Sohn *et al.* [12], and Kim *et al.* [13] focus on the discomfort induced by disparity factor. For accommodation-vergence conflict factor, Park *et al.* predict the visual discomfort level of S3D images by a 3D accommodation-vergence mismatch predictor algorithm [14] and a 3D visual discomfort predictor model [15] in terms of the physiological optics and neuronal framework. Jones *et al.* proposed an adjusting method for the depth range or the accommodation-convergence relationship based on floating windows, image shifting, and depth control/adjustment [16]. Oh *et al.* developed a Dynamic Accommodation and Vergence Interaction (DAVI) model to predict visual discomfort on S3D images [17]. Chen *et al.* developed a new feature map to accurately predict experienced 3D visual discomfort without disparity [18]. Kim *et al.* developed a unique temporal visual discomfort model (TVDM) to automatically predict the degree of discomfort felt when viewing stereoscopic 3D (S3D) images [19]. Xu *et al.* perdict 3D visual discomfort by sparse representation [20]. For binocular asymmetry factor, Yano *et al.* detected visual discomfort image scenes based on the correlation of left and right images [21]. Choi *et al.* [22] proposed a linear combination SVCA method based on 3D video characteristics. Ide *et al.* also used binocular asymmetry (*e.g.*, color mismatch) postprocess to lessen visual discomfort. In this paper, we attempt to predict stereo visual discomfort by objective measure [23].

In the human binocular visual perception, an intricate process of neural interaction has been occurred to respond

Identify applicable funding agency here. If none, delete this.

978-1-5386-5321-0/18/$31.00 ©2018 IEEE

the stimulations from stereoscopic image or video. Neural interaction generates a multitude of depth cues which are used by human visual system (HVS) to understand a 3D scene. However, there exits a gap between natural viewing and stereoscopic viewing. In natural viewing, when the eyes fixate on something, that point can be converged on the same point in two eyes' retinas, while in stereoscopic viewing, the retinas' points are misalignments. Binocular retinas have extremely low tolerance for the misalignments, thus causing viewing discomfort. Unnatural stereoscopic image may lead texture statistical inconsistency which induces binocular asymmetry. Therefore, we represent binocular asymmetry by two views' image texture features (*e.g.* histograms of oriented gradient (HOG) and local binary pattern (LBP)). Also, we attempt to predict stereo visual discomfort induced by the binocular asymmetry factors.

The remaining parts of the paper are organized as follows. Section II describes the framework of the proposed SVCA metric, including image feature extraction and visual comfort calculation. Section III provides the experimental results and Section IV concludes the paper.

II. The Proposed SVCA Metric

The framework of the proposed SVCA metric is shown in Fig. 1. For a given stereoscopic image, to compute the binocular asymmetry-induced discomfort, HOG and LBP features are extracted from each view's image of the given stereoscopic image, respectively. Then, two views' HOG and LBP features are combined to form the image's feature vector. Finally, the comfort score is calculated by the prediction function.

The proposed SVCA metric will be described in detail in the following:

A. Image Feature Extraction

Although the two views' images are highly similar in natural viewing, our eyes can perceive the slight difference and fuse into a single 'cyclopean' 3D image. To assess the difference between the two views' image, exploiting the image features of visual perception, we want to quantify the difference between the two views image features (*e.g.* HOG [24], LBP [25]).

For an input image, the gradient amplitude is firstly calculated both in the horizontal and vertical directions with a 1-D mask template ([-1 0 1]):

$$G_x(x,y) = I(x+1,y) - I(x+1,y), \tag{1}$$

$$G_y(x,y) = I(x,y+1) - I(x,y-1), \tag{2}$$

where $I(x,y)$ is pixel value of the point (x,y), $G_x(x,y)$ and $G_y(x,y)$ denote horizontal gradient amplitude and vertical gradient amplitude respectively.

Then, the gradient amplitude of the pixel (x,y) is:

$$G(x,y) = \sqrt{G_x^2(x,y) + G_y^2(x,y)}, \tag{3}$$

and the gradient direction of the pixel (x,y) is:

$$\theta(x,y) = tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right). \tag{4}$$

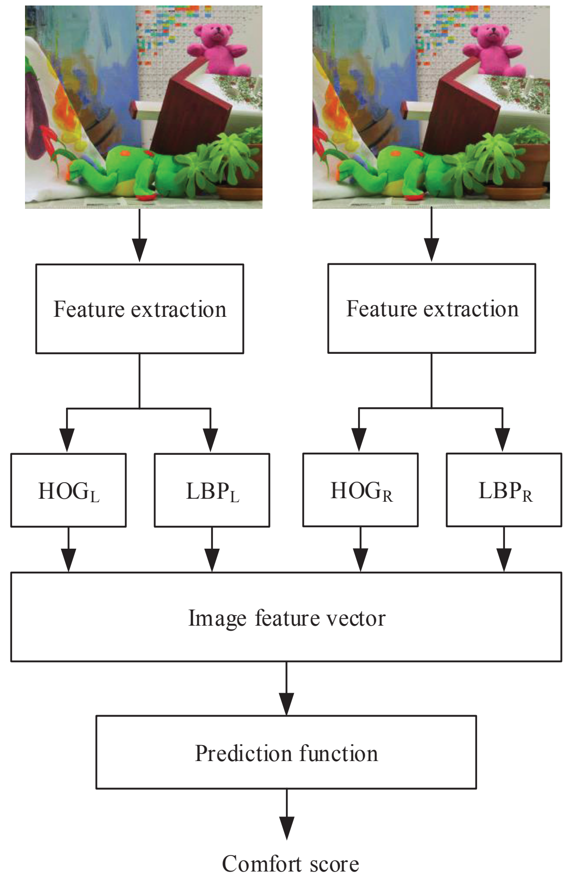

Fig. 1. Framework of the proposed SVCA metric.

The gradient magnitude and orient of each pixel in the 16 × 16 block is voted into 9 bins with the tri-linear interpolation. Tri-linear interpolation makes the image block of the border region better allocated to enhance the stability of the characteristics. The histograms of the block concatenated on a $45-D$ feature vector $\{HOG\}$ that is normalized by $L1-Hys$ to reduce the influence of the local variation in illumination and foreground-background contrast. Then, for the stereoscopic image, two views' image HOG features are $\{HOG_L, HOG_R\}$.

LBP is a texture descriptor for its invariance of gray-scale and rotation and is defined as:

$$LBP_{P,R} = \sum_{i=0}^{P-1} s \cdot (g_i - g_c) \cdot 2^P, \tag{5}$$

where

$$s(\cdot) = \begin{cases} 1, & s \geq 0 \\ 0, & s < 0 \end{cases}, \tag{6}$$

g_i is the gray value of a sampling point in an evenly spaced circular neighborhood of P sampling points and radius R around point i. In practice, Eq.(5) means that the signs of the differences in a neighborhood are interpreted as a P-bit binary number, resulting in 2^P distinct values for the LBP code. Then, for the stereoscopic image, two views' image LBP features are $\{LBP_L, LBP_R\}$.

B. Visual comfort calculation

As an important discomfort-inducing factor, binocular asymmetry is constructed by an stereoscopic image features vector $v : \{HOG_L, HOG_R, LBP_L, LBP_R\}$, which consists of the two views' HOG and LBP features. As the basic method of feature extraction, HOG and LBP can efficiently describe the gradient and texture features of the image. Then, two views' HOG and LBP features are used to calculate the binocular asymmetry-inducing discomfort.

The overall discomfort score of a stereoscopic image is computed using the perceptual vector v and a prediction function $f(\cdot)$. That is, the final quality score is given by:

$$Q = f(v). \tag{7}$$

Since support vector regressor (*SVR*) has achieved good performance on high-dimensional regression problems and successfully exploited in SVCA algorithms [11] [14], f is trained in advance using *SVR* regression (ε -SVR) [26]. Here, $f : R^2 \rightarrow R$ takes the three dimensional perceptual loss vector as input and produces output as a corresponding quality score.

In the ε -SVR, the unknown function f is constructed by linearly combining the results of a nonlinear transformation of the input samples.

$$f(x) = \sum_{i=1}^{l} (a_i - a_i^*) \cdot K(x_i, x) + b, \tag{8}$$

where a_i and a_i^* are the Lagrange multipliers, and $K(x_i, x)$ is the kernel function to perform the nonlinear transformation.

Here, we choose the radial basis function (RBF) kernel as following:

$$K(x_i, x) = e^{-\gamma \|x_i - x_j\|^2}, \tag{9}$$

where γ is a positive number, which represents the variance of the kernel function. In solving the *SVR*, the ε -insensitive loss function is used, which ignores errors that are smaller than a certain threshold $\varepsilon > 0$. Also, the penalty parameter C is used to control the complexity of prediction function f.

III. Experimental Results

In this section, we firstly present the experiment setting, then evaluate the performance for the proposed SVCA metric. Finally, we discuss the other discomfort-inducing factors.

A. Experiment setting

We performed the experiments on IEEE-SA [27] and EPFL [28] databases. The IEEE-SA database consists of 800 stereoscopic images with associated *MOS*, with resolution of 1920 $\times$ 1080 pixels. The EPFL database contains 9 scenes and each scene has 6 different camera distances, totally 54 stereoscopic images with associated *MOS*. Each view image' s resolution is close to 1920 $\times$ 1080 pixels. The IEEE-SA database was randomly divided into test and training subsets. The training and testing subsets did not overlap in content. We use 100-times 10-fold cross validation when measuring the performance [29]. Four popular evaluation criteria are chosen to compare Q with *MOS*, including PLCC, SROCC, KRCC and RMSE. A good objective method should have high PLCC, SROCC and KRCC values but low RMSE value.

In our experiment, one view image's HOG feature vector is a 45-D vector, and LBP feature vector is a 256-D vector. Then, the $v : \{HOG_L, HOG_R, LBP_L, LBP_R\}$ vector is a 602-D vector. The regression of the prediction function is performed using the LIBSVM [30]. Three parameters are fixed by a grid search, *e.g.* $\gamma = 1$, $C = 128$, $\varepsilon = 0.25$. The entire IEEE-SA database is used to train the prediction function, then the prediction function with the same parameters is tested on the EPFL database.

B. Performance evaluation

We developed two SVCA metrics for comparison. The metrics proposed by Sohn *et al.* [12] require disparity/depth map to compute visual discomfort, in our implementation, the disparity/depth information is estimated by GC [31]. The metrics proposed by Kim *et al.* [13] estimates disparity/parallax and use it to predict visual discomfort.

Table I and Table II provide the performance comparison results on the IEEE-SA and EPFL databases, respectively. It can be seen from Table I, on the IEEE-SA database, the proposed metric achieves the best performance in PLCC and SROCC, while the metric by Kim *et al.* achieves the best performance in KRCC and RMSE. Table II shows the performance comparison on the EPFL database, the metric by Kim *et al.* achieves the best performance in SROCC and KROCC, while the proposed metric achieves the best performance in PLCC and RMSE. Although the stereo image contents, the display systems, the subjects, and the subjective experiments are different in the two databases, the proposed SVCA metric can still work well.

TABLE I
PERFORMANCE COMPARISON ON THE IEEE-SA DATABASE

Metric	PLCC	SROCC	KRCC	RMSE
Sohn *et al.* [12]	0.4354	0.3798	0.2788	0.3527
Kim *et al.* [13]	0.7088	0.6183	**0.5313**	**0.3056**
The proposed metric	**0.7253**	**0.6306**	0.4640	0.5318

TABLE II
PERFORMANCE COMPARISON ON THE EPFL DATABASE

Metric	PLCC	SROCC	KRCC	RMSE
Sohn *et al.* [12]	0.7122	0.6254	0.4815	15.2494
Kim *et al.* [13]	0.8519	**0.8546**	**0.7446**	10.6943
The proposed metric	**0.8763**	0.8176	0.7045	**7.0306**

C. Discussion

Since the different discomfort-induced factors are exploited in the three metrics, the performances of these metrics are quite difference. Both Sohn *et al.* and Kim *et al.* used disparity to predict visual discomfort. The performance of Kim *et al.* metric is much better than Sohn *et al.*'s. There are two reasons. The first is that Sohn *et al.* only considered the horizontal disparities, while Kim *et al.* took both horizontal and vertical disparities into account. The second is that the poor quality of the disparity/depth map estimated by GC influences the evaluation performance of Sohn *et al.* metric. Since the proposed metric is a binocular asymmetry based measure method, it is sensitive to the slight change of both views' image. Although the proposed metric does not achieve an overwhelming advantage in performance comparison, it need not compute the disparity directly and can predict visual discomfort more efficiently.

IV. CONCLUSION

To solve the binocular asymmetry problem in stereoscopic image quality assessment, this paper proposes an objective SVCA metric. In the proposed metric, HOG and LBP features of two views' image are firstly extracted. Then, the two kinds of features are integrated into an overall quality score by SVR. The experimental results show that the proposed SVCA metric can efficiently predict the visual discomfort in two stereoscopic image databases.

It should be noted that this study only focuses on the visual discomfort inducing by the binocular asymmetry. Many other visual discomfort factors, such as accommodation-vergence conflict, disparity, range of depth of focus and other discomfort-inducing factors, need to be further investigate in the future work.

ACKNOWLEDGMENT

This work was supported by the China Postdoctoral Science Foundation under Grant No. 2017M610115 and National Postdoctoral Program for Innovative Talents under Grant No. BX201600006. This work was partially supported by the Natural Science Foundation of China under contracts 61527804.

REFERENCES

[1] M. Lambooij, W. IJsselsteijn, D. G. Bouwhuis and I. Heynderickx, "Evaluation of stereoscopic images: Beyond 2D quality," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 432–444, 2011.

[2] S. Winkler and D. Min, "Stereo/multiview picture quality: Overview and recent advances," *Signal Processing: Image Communication*, vol. 28, no. 10, pp. 1358–1373, 2013.

[3] A. K. Moorthy and A. C. Bovik, "A survey on 3D quality of experience and 3D quality assessment," *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics*, vol. 8651, pp. 86510M-186510M-11, 2013.

[4] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 2–4, 2009.

[5] T. Shibata, J. Kim, DM. Hoffman and MS. Banks, "The zone of comfort: predicting visual discomfort with stereo displays," *Journal of Vision*, vol. 11, no. 8, pp. 1–29, 2011.

[6] Y. J. Jung, H. G. Kim and Y. M. Ro, "Critical Binocular Asymmetry Measure for the Perceptual Quality Assessment of Synthesized Stereo 3D Images in View Synthesis," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 26, no. 7, pp. 1201–1214, 2016.

[7] B. Wang, K.J. Ciuffreda, "Depth-of-focus of the human eye: theory and clinical implications," *Survey of Ophthalmology*, vol. 51, no. 1, pp. 75–85, 2006.

[8] F. S. Avarvand, S. Bosse, K.R. Mller, R. Schfer, G. Nolte, T. Wiegand, G. Curio and W. Samek, "Objective quality assessment of stereoscopic images with vertical disparity using EEGA," *Journal of Neural Engineering*, 2017.

[9] L. Xing, J. You, T. Ebrahimi and A. Perkis, "Assessment of stereoscopic crosstalk perception," *IEEE Trans. on Multimedia*, vol. 14, no. 2, pp.326–337, 2012.

[10] 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," *n Proc. of SPIE-IS&T Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems X*, vol. 5006, pp. 195–205, 2003.

[11] Y.J Jung, H. Sohn, S. Lee, H.W. Park, Y.M. Ro, "Predicting visual discomfort of stereoscopic images using human attention model," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 223, no. 12, pp. 2077–2082, 2013.

[12] H. Sohn, Y. J. Jung, S. Lee, H. W. Park and Y. M. Ro, "Attention model-based visual comfort assessment for stereoscopic depth perception," *in Proc. Int. Conf. Digital Signal Processing*, pp. 1–6, 2011.

[13] D. Kim and K. Sohn, "Visual fatigue prediction for stereoscopic image," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 21, no. 2, pp. 231–236, 2011.

[14] J. Park, S. Lee and A.C. Bovik, "3D visual discomfort prediction: vergence, foveation, and the physiological optics of accommodation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 415–427, 2014.

[15] J. Park, H. Oh, S. Lee and A. C. Bovik, "3D visual discomfort predictor: Analysis of disparity and neural activity statistics," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1101–1114, 2015.

[16] G.R. Jones, D. Lee, N.S. Holliman and D. Ezra, "Controlling perceived depth in stereoscopic images," *in Proc. of SPIE-IS&T Electronic Imaging, International Society for Optics and Photonics*, pp. 42–53, 2001.

[17] H. Oh, S. Lee and A. C. Bovik, "Stereoscopic 3D visual discomfort prediction: A dynamic accommodation and vergence interaction model," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 615–629, 2016.

[18] J. Chen, J. Zhou, J. Sun and A. C. Bovik, "Visual discomfort prediction on stereoscopic 3D images without explicit disparities," *Signal Processing: Image Communication*, vol. 51,pp. 50–60, 2017.

[19] T. Kim, S. Lee and A. C. Bovik, "Transfer function model of physiological mechanisms underlying temporal visual discomfort experienced when viewing stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4335–4347, 2015.

[20] H. Xu, G. Jiang, M. Yu, T. Luo, Z. Peng, F. Shao, H. Jiang, "3D visual discomfort predictor based on subjective perceived-constraint sparse representation in 3D display system," *Future Generation Computer Systems*, vol. 83, pp. 85–94, 2018.

[21] 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, 2002.

[22] J. Choi, D. Kim, B. Ham, S. Choi and K. Sohn, "Visual fatigue evaluation and enhancement for 2D-plus-depth video," *in Proc. Int. Conf. Image Proc.*, pp. 2981–2984, 2010.

[23] S. Ide, H. Yamanoue, M. Okui, F. Okano, M. Btou and N. Terashima, "Parallax distribution for ease of viewing in stereoscopic HDTV," *in Proc. of SPIE-IS&T Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems IX*, bol. 4660, 2002.

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *in Proc. of CVPR*, pp. 886–893, 2005.

[25] L. Wang and D. C. He, "Texture Classification Using Texture Spectrum," *Pattern Recognition*, vol. 23, no. 8, pp. 905–910, 1990.

[26] V. N. Vapnik, "Statistical Learning Theory," *New York: Wiley*, 1998.

[27] Standard for the quality assessment of three dimensional (3D) displays, 3D contents and 3D devices based on human factors," *IEEE P3333.1*, 2012. [Online] Available: http://grouper.ieee.org/groups/3dhf.

[28] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," *in Proc. of SPIE-IS&T Electronic Imaging, Stereoscopic Image Processing and Applications*, vol. 7526, 2010.

[29] M. Green and M. Ohlsson, "Comparison of standard resampling methods for performance estimation of artificial neural network ensembles," *in Proc. Computat. Intell. Med. Healthcare*, 2007.

[30] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2001, Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[31] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," *in Proc. ECCV*, vol. 2352,pp. 82–96, 2002.