AN IMPROVED MODEL OF BINOCULAR ENERGY CALCULATION FOR FULL-REFERENCE STEREOSCOPIC IMAGE QUALITY ASSESSMENT

G.C.V. Perera, V. De Silva, A.M. Kondoz, and S. Dogan

I-Lab Multimedia Communications Research, CVSSP, University of Surrey, Guildford GU2 7XH, UK
c.perera@surrey.ac.uk

ABSTRACT

With the exponential growth of stereoscopic imaging in various applications, it has become very demanding to have a reliable quality assessment technique to measure the human perception of stereoscopic images. Quality assessment of stereoscopic visual content in the presence of artefacts caused by compression and transmission is a key component of end-to-end 3D media delivery systems. Despite a few recent attempts to develop stereoscopic image/video quality metrics, there is still a lack of a robust stereoscopic image quality metric. Towards addressing this issue, this paper proposes a full reference stereoscopic image quality metric, which mimics the human perception while viewing stereoscopic images. A signal processing model that is consistent with physiological literature is developed in the paper to simulate the behaviour of simple and complex cells of the primary visual cortex in the Human Visual System (HVS). The model is trained with two publicly available stereoscopic image databases to match the perceptual judgement of impaired stereoscopic images. The experimental results demonstrate a significant improvement in prediction performance as compared with several state-of-the-art stereoscopic image quality metrics.

Index Terms— *Human Visual System, Stereoscopic quality assessment, Binocular vision*

1. INTRODUCTION

3D visual content has become increasingly popular among modern users, who demand high quality immersive multimedia content. The quality of stereoscopic 3D video content is degraded at different stages of the video delivery life cycle, such as during encoding and transmission. Such quality degradations have an adverse effect on the quality of experience of its users, and thus they need to be quantified accurately to ensure user satisfaction. This has led to the requirement of a reliable quality assessment technique, which mimics human perceptual judgement on 3D visual content. Conventional quality metrics used for 2D content have proven to be inadequate to the purpose at hand, and thus recently there has been a significant interest among research communities to produce novel quality metrics that assess 3D visual content quality. In [1]-[2], metrics were proposed to assess the quality of stereoscopic images under different degradation types, while research effort has been reported on quality assessment of stereoscopic video in [3]. However, there is no quality metric that is robust enough to predict the quality of stereoscopic content, which has undergone different kinds of degradations. Thus, modelling the Human Visual System (HVS) using binocular vision process is envisaged to pave the way for developing a robust stereoscopic visual quality metric [4]. Towards this end, this paper proposes a novel stereoscopic image quality metric that is based on a cellular model of the HVS.

When perceiving stereoscopic images, the binocular visual system combines the left and right eye views to perceive depth and produces a single view known as the cyclopean view. The visual cortex of the brain is responsible for processing the visual information acquired through the eyes. The first part of the visual cortex is the primary visual cortex. There are two main types of cells in the primary visual cortex known as simple cells and complex cells [5].

The work presented in this paper proposes to develop a model to mimic the behaviour of the cells in the primary visual cortex. The research work in [4] addressed mimicking this behaviour, but the performance of the method presented was limited due to the adopted model of the complex cells being too simplistic. Specifically, the complex cells' model did not discriminate between different orientations and sizes of the receptive fields of simple cells, in contrast to the operation of the HVS. Furthermore, the cellular model did not consider the binocular suppression effect that is present in the HVS. To accurately address these issues, the work presented in this paper aims to develop a cellular model that is more consistent with the HVS.

Specifically, this paper proposes a full reference stereoscopic image quality metric utilising the aforementioned model to measure the quality of stereoscopic images objectively, which have gone through various types of degradations. The model is trained using two publicly available stereoscopic image databases. Objective estimations of the final statistical model using images from two registered stereoscopic image databases [6] [1] proved a consistently high correlation with the subjective results.

The rest of the paper is organised as follows. The related work to the proposed research work found in literature is presented in Section 2, and the proposed metric is described in Section 3. Section 4 discusses the results obtained by the proposed metric and subsequently compares its performance against the state-of-the-art metrics. Finally, Section 5 concludes the paper with some insights into future work.

2. RELATED WORK

There are two approaches of assessing contours of depth maps using PSNR [7] and of assessing cyclopean images with disparity maps using SSIM [8] in literature. In [1] and [2], coherence between the metric scores and human perceptual judgement was shown.

There are a few research activities reported to identify real 3D evaluation metrics in literature. A metric was proposed based on

geometrical properties of a 3D object that used Just Noticeable Difference (JND) for redundancy reduction [9]. A no-reference metric for asymmetric JPEG compression based on partitioning of a stereo image-pair in fixed-size blocks to characterise 3D artefacts was proposed in [10]. Binocular vision process was used as an alternative approach for developing a 3D perceptual metric in [4] by modelling binocular fusion.

The requirement of a reliable and robust technique to evaluate human perceptual judgement is still an open research problem. The work described in this paper addresses this problem with an alternative solution to metric approaches using a cascade of statistical and analytical modelling techniques. In the next section, the proposed novel method to assess the quality of stereoscopic images is described.

3. PROPOSED METHOD

A model for stereoscopic quality assessment is proposed using physiological knowledge of the HVS. The binocular view, which results in generating binocular energy, is an indication of the quality of perception in the HVS [4]. According to the physiological studies of the HVS [5], there are mainly two types of cells in the visual cortex responsible for binocular vision. The simple cells are the first to receive the retinal information at the visual cortex. They are responsible for the local spatial frequency. Binocular simple cells work in pairs for left and right eyes, and are connected to binocular complex cells. The complex cells are responsible for the generation of binocular energy.

3.1 Analytical modelling of the primary visual cortex of the HVS

An analytical model, which can calculate required binocular energy as a set of complex cell outputs, is used. This analytical model consists of the individual models that are described below.

3.1.1 Sampling model

There are two main visual characteristics of the HVS modelled using sampling functions in [11] [12]: 1) image decomposition into perceptual channels from low frequencies to high frequencies as in the HVS, and 2) localising image elements in spatial and frequency spaces. Colour antagonism in the HVS is modelled using a colour space conversion to CIE L*a*b* according to [13]. Here, an image is represented with a single channel of luminance L* and two perpendicular channels of chrominance a* and b*. For better performance in latter stages of modelling, the resultant of a* and b* is considered as chrominance C* in the proposed model.

Initially, the stereoscopic (left and right) images are represented as complex wavelets. For this purpose, the Complex Wavelet Transformation (CWT) [12] is applied on the luminance (L*) component, which produces a pair of real and imaginary coefficients, and the Discrete Wavelet Transformation (DWT) [11] on the two chrominance components. The wavelet representation of the chrominance (C*) component is organised as the DWT (a*) as the real part and DWT (b*) as the imaginary part [4].

3.1.2 Simple cell model

The characteristics of simple cells are modelled from the coefficients obtained from the complex representation of the images. The Bandelet Transform (BT) is used to model binocular simple cells, as it provides a closely matching behaviour to that of a simple cell [11]. Sub bands are obtained (as shown in Figure 1) in sampling model with respective wavelet coefficients. BT splits up those sub bands in a quad tree of variable sizes called Dyadic Squares following the image geometry. An orientation is computed and assigned to each dyadic square depending on the wavelet coefficients. This dyadic square is characterised by its size, amplitude, and orientation as a simple cell [14].

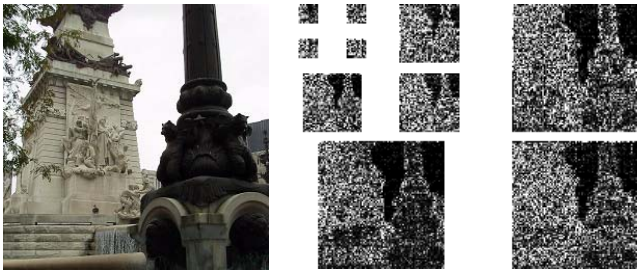


Figure 1. Sub bands of sampling model for the left eye view

A pair of real and imaginary dyadic squares, which are output of the BT, represents the excitatory and inhibitory responses of a group of simple cells with a defined orientation and spatial frequency selectivity. The coefficients of the wavelet transformation are sensitive to spatial impairments such as quantisation noise or blurring. Thus, the spatial impairments are reflected in the phase $\varphi(x)$ and amplitude $\rho(x)$ of the complex wavelet representation.

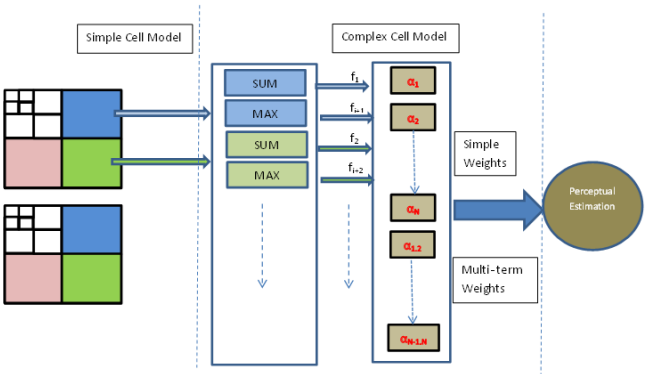


Figure 2. Illustration of simple and complex cell models

3.1.3 Complex cell model

The complex cell model is responsible for calculation of the binocular energy. The complex cells inherit most of the properties of simple cells, such as being orientation and spatial frequency selective, but are invariant of the spatial phase [5]. A complex cell receives inputs from several simple cells that are of same orientation and spatial frequency selectivity for generating the binocular energy. However, due to spatial phase invariance, a particular complex cell could receive inputs from both excitatory and inhibitory responses of the simple cells.

As a group, complex cells tend to be more heterogeneous than simple cells. Most common type of complex cells performs a summation-like operation on the responses of simple cells with similar orientation preference [15]. Recently, researchers have found evidence to suggest that complex cells perform a MAX-like operation on their inputs [16]. Furthermore, there are evidences of interactions between complex cells as proposed in the so-called recurrent excitation model. In such interactions, the output of one complex cell is modulated by the output of another complex cell

[17]. The proposed model of the primary visual cortex is designed in a way to capture the features discussed above.

A pair of dyadic squares (real and imaginary responses of the BT) from the left image and a pair from the right image are used in the binocular energy calculation. The real and imaginary parts of the luminance and chrominance are used to calculate the monocular amplitude ρ_i (x) and monocular phase φ_i (x) of the xth dyadic square of the left and right images, as in (1) and (2) [4].

$$\rho_i(x) = |C_i(x)| = [\mathrm{Re}(C_i(x))^2 + \mathrm{Im}(C_i(x))^2]^{1/2} \quad (1)$$

$$\varphi_i(x) = \arg(C_i(x)) = \tan^{-1}\left[\frac{\mathrm{Im}(C_i(x))}{\mathrm{Re}(C_i(x))}\right] \quad (2)$$

The output of a complex cell that performs a summation like operation is the resultant of the excitatory and inhibitory responses of the simple cells as given in (3).

$$E_s(x) = \rho_l{}^2(x) + \rho_r{}^2(x) + \rho_l(x)\rho_r(x)\cos(\varphi_r(x) - \varphi_l(x)) \quad (3)$$

In (3), l and r represent left and right images, respectively, and $E_S(x)$ is the binocular energy generated by a complex cell that performs a summation-like operation. Here $(\varphi_r(x,y) - \varphi_l(x,y))$ is the inter-ocular phase shift, which in turn relates to the horizontal disparity between the left and right views.

Binocular suppression is applicable when left and right images have undergone asymmetric impairments, which causes the HVS to ignore the view of one eye and perceive through the other eye. This behaviour of binocular suppression can be modelled using the complex cells that perform MAX-like operation. The binocular energy output $E_m(x,y)$ of a complex cell that performs a MAX-like operation is given as:

$$E_m(x) = \max(\rho_l{}^2(x), \rho_r{}^2(x)) \quad (4)$$

The proposed method also considers interactions between the complex cells that perform summation-like and MAX-like operations.

The total binocular energy for the luminance component E_L is calculated as a weighted sum of all the complex cell outputs as in (5).

$$E_L = \sum_{i=1}^{n} \alpha_i E_{s_i} + \sum_{i=1}^{n} \beta_i E_{m_i} + \sum_{i=1}^{n-1} \gamma_i E_{s_i} E_{s_{i+1}} + \sum_{i=1}^{n-1} \delta_i E_{m_i} E_{m_{i+1}} \quad (5)$$

In (5) $\alpha_i, \beta_i, \gamma_i$ *and* δ_i are corresponding weights for energy outputs of complex cells E_{s_i}, E_{m_i} and also for interactions of similarly operational energy outputs of two complex cells as shown in Figure 2. Similar to (5), the total binocular energy for the chrominance component E_C is calculated. As given in (6), the total binocular energy E is the addition of both the luminance energy and chrominance energy.

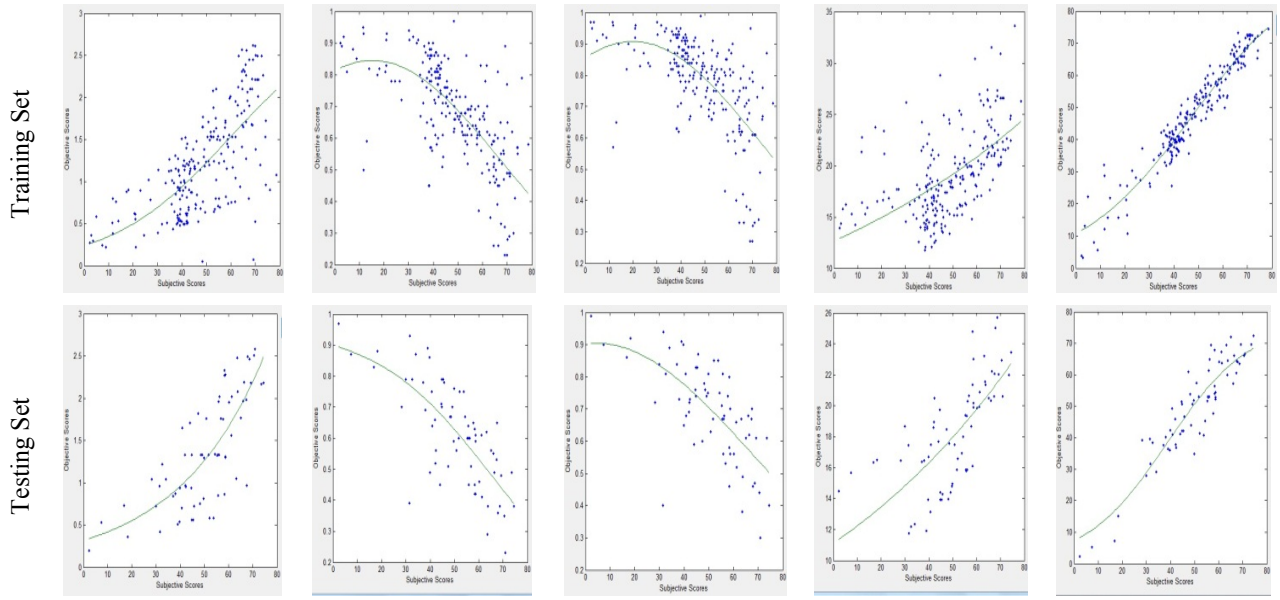
$$E = E_L + E_C \quad (6)$$

3.1.4 Statistical training of the analytical model

To estimate the coefficients of the analytical model presented in the previous section, the model is trained utilising two publicly available stereoscopic image databases.Stereoscopic image pairs used from the first database [6] consist of 8 different scenes with each scene having 27 stereoscopic images consisting of 3 distortion types (JPEG 2000, JPEG, and Gaussian Blur) and 9 profiles on each distortion. These profiles have 3 symmetrically impaired image pairs and 6 asymmetrically impaired image pairs. In the second database [1], there were 5 scenes with each scene having 15 stereoscopic images consisting of the 3 distortion types and 5 symmetric profiles on each distortion type.

The total of 291 stereoscopic stimuli from the two databases is divided into two sets, one for training and the other for testing the analytical model. The training set was defined with 222 stimuli, whereas the remaining 69 stimuli were used as the testing set.

Finally, the coefficients described in (5) were calculated using the stepwise linear regression modeling function in Mat Lab® R2013. The final solution yielded in 25 terms including 13 interaction terms. However, due to the space limitations, we have not provided the final coefficients and related terms in the paper.



BEQM [4] SSIM_Ddl [1] SSIM_Avg [2] NRIM [9] Proposed

Figure 3. The correlation plots between the objective and subjective scores, for the training and testing sets

4. RESULTS AND DISCUSSION

Performance of the proposed method is compared with a number of state-of-the-art metrics, based on several performance criteria as outlined in [18]. Specifically, the following parameters are measured after logistic transformation of the objective scores on to the scale of subjective scores, as shown in Figure 3: the prediction consistency by the Pearson's linear Correlation (PCC), prediction monotonicity by Spearman's Rank Order Correlation (SROCC) and prediction accuracy by Average Absolute Error (AAE) and Root Mean Squared Error (RMSE). The proposed metric is compared against four different stereoscopic image quality metrics that are proposed recently, which are SSIM_Ddl metric [1], the average SSIM of the left and right views (SSIM_Avg) [2], the No Reference stereoscopic Image Metric (NRIM) [10], and the Binocular Energy Quality Metric (BEQM) [4]. The performance comparison results are provided in Tables 1 and 2.

Table 1 Performance measures for the training set

	PCC	SROCC	AAE	RMSE
BEQM [4]	0.695	0.689	46.551	0.427
SSIM_Ddl [1]	0.681	0.681	47.084	0.116
SSIM_Avg [2]	0.644	0.669	46.986	0.113
NRIM [10]	0.561	0.587	29.727	3.419
Proposed	**0.961**	**0.956**	**3.822**	**4.194**

Table 2 Performance measures for the testing set

	PCC	SROCC	AAE	RMSE
BEQM [4]	0.771	0.764	48.77	0.4179
SSIM_Ddl [1]	0.733	0.718	49.53	0.1116
SSIM_Avg [2]	0.683	0.691	49.45	0.1107
NRIM [10]	0.703	0.763	32.56	2.4300
Proposed	**0.920**	**0.894**	**4.94**	**6.0984**

The results presented in Tables 1 and 2, as well as the correlation plots in Figure 3, clearly illustrate that the proposed metric significantly outperforms all the considered other metrics in every aspect of performance comparison. Furthermore, the results illustrate that the proposed analytical model of the primary visual cortex is quite suitable for the purpose of stereoscopic quality assessment. The closest contender to the proposed method is the BEQM, which inspired the current work, and which also consider a binocular energy model. The improvements of the presented work over the BEQM can be attributed to three major novelties introduced in the proposed analytical model. Firstly, in BEQM, all the simple cells are treated in a similar way without considering different spatial orientations and frequencies, whereas in the proposed model, different weights are assigned based on individual contributions. Secondly, the model in BEQM does not consider complex cells with a MAX operation, and thus fails to predict the quality of asymmetrically processed stereoscopic stimuli. Finally, the proposed method incorporates interaction between complex cells, which has been ignored in BEQM.

On the limitations of the presented work, it should be noted that the proposed method did not perform well on stimuli that were impaired with white noise. Thus, such stimuli have not been considered in this study. To overcome this limitation, an algorithm is required to identify the type of noise, and treat the stimuli accordingly.

5. CONCLUSION AND FUTURE WORK

This paper proposed a cellular model of the primary visual cortex of the HVS to predict the subjective quality of impaired stereoscopic images. Specifically, the complex cells of the primary visual cortex have been modelled as consistent with the physiological literature, where they have been modelled as orientation and spatial frequency selective cells that respond to perform summation and MAX operation on simple cell responses. Furthermore, interactions between complex cells have also been considered. The binocular energy has been calculated as a weighted sum of different complex cell outputs, thus treating different complex cell types asymmetrically to produce an accurate prediction of the stereoscopic image quality. The proposed model has been trained using two publicly available stereoscopic image databases. The experimental results of the proposed metric have illustrated a correlation of 0.92 with subjective results. Further improvements are sought to improve the prediction performance of the proposed metric by conforming to further studies on physiology of the HVS, while also making it suitable to measure the quality of impaired stereoscopic videos.

6. ACKNOWLEDGEMENT

This work was supported by the ROMEO project (grant number: 287896), which was funded by the EC FP7 ICT collaborative research programme.

7. REFERENCES

[1] Benoit A., Callet P. L., Campisi P., Cousseau R., "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–13, Jan. 2009.

[2] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] De Silva, V., Arachchi, H. K., Ekmekcioglu, E. Kondoz, A., "Toward an Impairment Metric for Stereoscopic Video: A Full-Reference Video Quality Metric to Assess Compressed Stereoscopic Video", IEEE Transactions on Image Processing, vol. 22, no 9, 2013.

[4] Bensalma R., Larabi M. C., "A perceptual metric for stereoscopic image quality assessment based on the binocular energy", Multidimensional Syst. Signal Process., vol. 24, no. 2, pp. 281–316, 2012.

[5] Hubel D. H., Wiesel T. N., "Stereoscopic vision in macaque monkey cells sensitive to binocular depth in area 18 of the macaque monkey cortex", Journal of Nature, 225, 41–42, 1970.

[6] Sheikh H. R., Wang Z., Cornack L., Bovik A. C., Live Image Quality Assessment Database, Release2 2005 [online], Available: http://live.ece.utexas.edu/research/quality

[7] Hewage C., Martini M.G, "Reduced-reference quality metric for 3D depth map transmission", IEEE 3DTV conference (pp. 1–4), Tampere, Finland, 2010.

[8] Xing L., You J., Ebrahimi T, Perkis A. "A perceptual quality metric for stereoscopic crosstalk perception", IEEE international conference on image processing (pp. 4033–4036), Hong Kong, 2010.

[9] Cheng I., Boulanger P., "A 3D perceptual metric using just-noticeable-difference", Eurographics Short Presentations (pp. 97–100), 2005.

[10] Akhter R., Sazzad Z. M. P., Horita Y., Baltes J, “No reference stereoscopic image quality assessment”, Image quality and system performance (vol. 7524, pp. 17–21), San Jose, California, USA, 2010.
[11] Mallat S., Peyré G., “Orthogonal bandelet bases for geometric image approximation”, Communications on Pure and Applied Mathematics, 61(9), 1173–1212, 2006.
[12] Kingsbury N., “Image processing with complex wavelets”, Philosophical Transactions on Royal Society London A. 357, 2543–2560, 1997.
[13] Schanda J., “Colorimetry: Understanding the CIE System”, Hoboken, NJ, USA, 2007.
[14] Peyre G., “Geometrie multi-échelles pour les images et les textures”, Ph.D. thesis, Ecole Polytechnique, 2005.
[15] Finn and Ferster, Computational Diversity in Complex Cells of Cat Primary Visual Cortex, 2007
[16] Movshan et al. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex, 1978
[17] Tao et al. An egalitarian network model for the emergence of simple and complex cells in visual cortex, 2004
[18] Final report from the video quality experts group on the validation of objective models of multimedia quality assessment. Technical Report, PHASE I 2008.