Digital Signal Processing ••• (••••) •••–•••

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Binocular energy response based quality assessment of stereoscopic images

Feng Shao *, Gang-yi Jiang, Mei Yu, Fucui Li, Zongju Peng, Randi Fu

Faculty of Information Science and Engineering, Ningbo University, Ningbo, 315211, China

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Stereoscopic image quality assessment
Binocular energy response
Binocular masking
Binocular just-noticeable difference
Quality of experience

ABSTRACT

Perceptual quality assessment in three-dimensional (3D) is challenging. In this research, we propose a binocular energy response based quality assessment metric of stereoscopic images. To be more specific, we first construct binocular energy responses of the original and distorted images, and measure the similarity between them as Image Quality Metric (IQM). Then, the binocular response and the binocular masking components are used to modulate the IQM, respectively. Finally, two evaluation results are nonlinearly integrated into an overall score by considering the importance of each component. Experimental results show that compared with the relevant existing metrics, the proposed metric can achieve higher consistency with the subjective assessment of stereoscopic image.

1. Introduction

Three-dimensional (3D) imaging technologies have been researched widely recently [1], the application of which ranges from content creation, video coding, network transmission and stereoscopic display. Therefore, designing for objective perceptual 3D image quality assessment (3D-IQA) approach is increasingly important [2], since such perceptual issue is hardly considered in the traditional 2D image quality assessment (2D-IQA). Following the research of 2D-IQA, 3D-IQA approaches fall into two categories: subjective assessment and objective assessment. Specifically, the development of objective quality assessment models has been a fruitful area of work.

In the aspect of objective/subjective assessment, the term 'quality of experience (QoE)' should be considered to capture the various factors that contribute to the overall visual experience of the 3D visual signal [3]. In contrast to the 2D case, QoE of 3D involves not only evaluating 2D image quality, but also additional aspects of quality, e.g., depth perception, visual comfort, and other visual experience. Many 2D-IQA metrics were proposed during the last decade, such as Structural SIMilarity (SSIM) [4], visual signal-to-noise ratio (VSNR) [5], visual information fidelity (VIF) [6], etc. However, the direct use of 2D-IQA in measuring 3D-IQA may not be straightforward, since the above 3D perceptual attributes needed to be considered. Lambooij et al. constructed a 3D quality model as a weighted sum of 2D image quality and perceived depth, and the model was validated by subjective experiments [7]. Chen et al. explored 3D QoE by constructing the visual experience as a weight sum of image quality, depth quantity and visual comfort, and subjective experiments were conducted to test the model [8]. However, these methods remained on a subjective level to explore the combination of various perceptual scales.

Currently, some publicly available 3D databases were provided, such as LIVE 3D IQA database [9], EPFL 3D image database [10], IRCCyN/IVC 3D image database [11], etc., by adding different types of stimuli (e.g., image distortion or camera distance) on both left and right images. Some objective 3D-IQA metrics were proposed by verifying on the databases. Research on objective 3D-IQA can be divided into two categories based on the involved information for evaluation. The most direct use of state-of-the-art 2D-IQA approaches in 3D-IQA is to evaluate the two views of the stereoscopic images, disparity/depth images separately by 2D metrics, and then combined into an overall score. Benoit et al. presented a linear combination solution for disparity distortion and 2D image quality on both views [11]. You et al. integrated the disparity information into quality assessment, and investigated the capabilities of some combination schemes [12]. Ha et al. designed a quality assessment method by considering the factors of temporal variation and disparity distribution [13]. Hewage et al. performed the evaluation for 3D video by using the extracted information from the depth maps and color images [14]. Obviously, it is not effective to assess the quality of perceived depth using image quality assessment methods (e.g., SSIM), because stimuli toward perceived depth are different with those for 2D image quality.

From another point of view, visual perceptual properties (e.g., monocular and binocular properties) were other important cues in 3D-IQA. Maalouf et al. computed the 'Cyclopean' image from left

* Corresponding author.

http://dx.doi.org/10.1016/j.dsp.2014.03.003
1051-2004/

and right images to simulate the brain perception, and used contrast sensitivity coefficients of cyclopean image as the basis of evaluation [15]. Lin et al. utilized binocular integration (i.e., binocular combination and the binocular frequency integration) behaviors as the bases for measuring the quality of stereoscopic 3D images [16]. Ryu et al. formulated a model for stereoscopic images based on binocular perception model considering asymmetric properties of stereoscopic images [17]. Ko et al. proposed a structural distortion parameter based binocular perception model for 3D image quality assessment [18]. Wang et al. proposed a metric by considering the binocular spatial sensitivity to reflect the binocular fusion and suppression properties [19]. Bensalma et al. proposed a Binocular Energy Quality Metric (BEQM) by modeling the simple cells responsible for the local spatial frequency analysis and the complex cells responsible for the generation of the binocular energy [20]. However, these methods are simple extensions of the monocular visual properties into the binocular vision, and how these monocular visual properties affect the binocular vision is still not accounted.

From the observation of the existing 3D-IQA metrics, both image quality and depth perception are expected to measure the QoE of 3D. However, the combination of these two parts is somewhat ill-defined in the existing metrics, since disparity map is estimated from the stereoscopic image, while the perception of depth from disparity is generally not well understood. In order to tackle the problem, we propose a binocular energy response based stereoscopic image quality assessment metric in this paper. The main contributions of this work are as follows: (1) We construct binocular energy responses of the original and distorted images based on Gabor filters and disparity information, and measure the similarity between them as Image Quality Metric (IQM); (2) By considering the binocular response and binocular masking characteristics, we use the binocular energy and the binocular just noticeable difference (BJND) components to modulate the IQM, respectively; (3) Two evaluation results are nonlinearly integrated into an overall score by considering the importance of each component. The rest of the paper is organized as follows. Section 2 discusses the background and motivation of the proposed metric. Section 3 presents the proposed quality assessment metric. The experimental results are analyzed in Section 4, and finally conclusions are drawn in Section 5.

2. Background and motivation

In order to explain the ideas of the proposed 3D-IQA metric, we first review some relevant works and point out the problems, and then present our innovation in the design of 3D-IQA.

2.1. Binocular energy response construction

It has been known that binocular vision is a complex visual process that requires the brain and both eyes working together to produce depth perception and clear vision [21]. The process of binocular visual perception can be regarded as responses of a pair of simple cells received from left and right eyes. As an example of one-dimensional signals, given the simple response functions, $C_l(x) = \rho_l(x)e^{i\varphi_l(x)}$ and $C_r(x) = \rho_r(x)e^{i\varphi_r(x)}$ for two input images, the binocular energy response, $BE(x)$, can be expressed as [22]

$$\begin{aligned} BE(x) &= \left\| C_l(x) + C_r(x) \right\|^2 \\ &= \rho_l^2(x) + \rho_r^2(x) + \rho_l(x) \cdot \rho_r(x) \cdot \cos\left(\Delta\phi(x)\right) \end{aligned} \tag{1}$$

where $\Delta\phi(x) = \phi_l(x) - \phi_r(x)$, being the phase difference between the left and right images.

However, the above binocular energy response may not be able to characterize depth perception (e.g., position shift between views is not considered in the above expression). In order to measure depth perception, the existing technologies [12–14] directly evaluate the quality of estimated disparity by 2D-IQA methods. However, the limitations of these evaluations are that: (1) binocular disparity is obtained by stereo matching method. The quality of the estimated disparity has a great relationship with the specialized stereo matching method since ground truth disparity is generally not available; (2) the HVS perceives a single mental image (e.g., cyclopean image) [23] of a scene by combining two images received from the two eyes. The disparity information is only assisted to the construction and formation of the cyclopean image. Fig. 1 shows the estimated disparity maps from different distorted stereoscopic images. It is clearly demonstrated that the quality of the estimated disparity has weak correlation with the subjective perceived quality.

Considering that two simple cells, belonging to the left and right images, have different spatial positions, the right response function $C_r(x)$ can become a shifted version of the left response function, i.e. $C_r(x+d) = C_l(x)$, thus, the binocular energy response $BE(x)$ can be re-expressed as

$$\begin{aligned} BE(x) &= \left\| C_l(x) + C_r(x+d) \right\|^2 \\ &= \rho_l^2(x) + \rho_r^2(x+d) + \rho_r(x) \cdot \rho_r(x+d) \cdot \cos\left(\Delta\phi(x)\right) \end{aligned} \tag{2}$$

where d is the estimated disparity value at pixel (x, y), $\rho_l^2(x)$ and $\rho_r^2(x+d)$ are the energy magnitudes of $C_l(x)$ and $C_r(x+d)$, respectively, and $\Delta\phi$ is the corresponding phase difference. It is noticed that the estimated disparity will affect the energy magnitude $\rho_r^2(x+d)$ and phase difference $\Delta\phi$. It should be emphasized that the disparity d is estimated from reference stereoscopic images (i.e., undistorted original stereoscopic images) in the above binocular energy response, because we believe that the identification of matched regions should mainly depend on camera geometry, while image distortions will do not seriously affect the camera geometry.

2.2. Perceptual properties analysis

The above binocular energy reflects the strength of response for different retinal points. However, in the process of binocular summation, the masking effect should not be ignored. It is well known that visual masking effect (e.g., formulated as just-noticeable difference (JND)) has played an important role in pro-HVS signal processing [24]. For example, the HVS can tolerate more error in higher frequency components while the distortion in lower frequency components has a larger impact on the visual quality. Recently, Zhao et al. proposed a BJND model to measure the minimum distortion in the two views of stereoscopic images with psychophysical experiments [25]. It is assumed that disparity between the patterns is zero, and disparity has minor impact on the binocular combination. In the following, we summarize the derivation of the BJND model. By incorporating the luminance and contrast masking effects, as well as considering the correspondence matching between two views, the BJND at the left view is defined as

$$\begin{aligned} BJND_l &= BJND_l\left(bg_r(x+d, y), eh_r(x+d, y), A_r(x+d, y)\right) \\ &= A_{C,limit}\left(bg_r(x+d, y), eh_l(x+d, y)\right) \\ &\quad \times \left(1 - \left(\frac{A_r(x+d, y)}{A_{C,limit}(bg_r(x+d, y), eh_r(x+d, y))}\right)^{\lambda}\right)^{1/\lambda} \end{aligned} \tag{3}$$

where d is the estimated disparity value computed similarly with Eq. (2), $bg_r(x+d, y)$ denotes the background luminance level, $eh_r(x+d, y)$ denotes the edge height, $A_r(x+d, y)$ denotes the noise amplitude, and $A_{C,limit}$ is the visibility threshold determined

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003

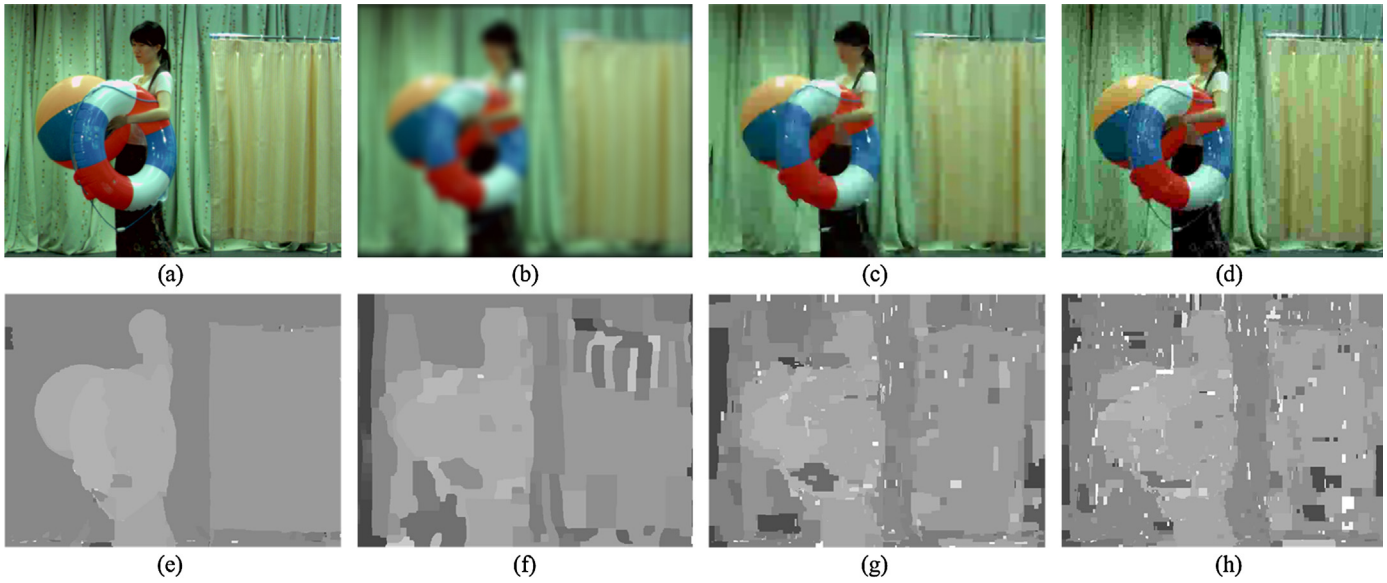

Fig. 1. Examples of quality degraded left images and the corresponding disparity maps: (a) Original left image; (b) Gaussian blurred left image; (c) H.264 compressed left image; (d) JPEG compressed left image; (e) The corresponding disparity map of (a); (f) The corresponding disparity map of (b); (g) The corresponding disparity map of (c); (h) The corresponding disparity map of (d).

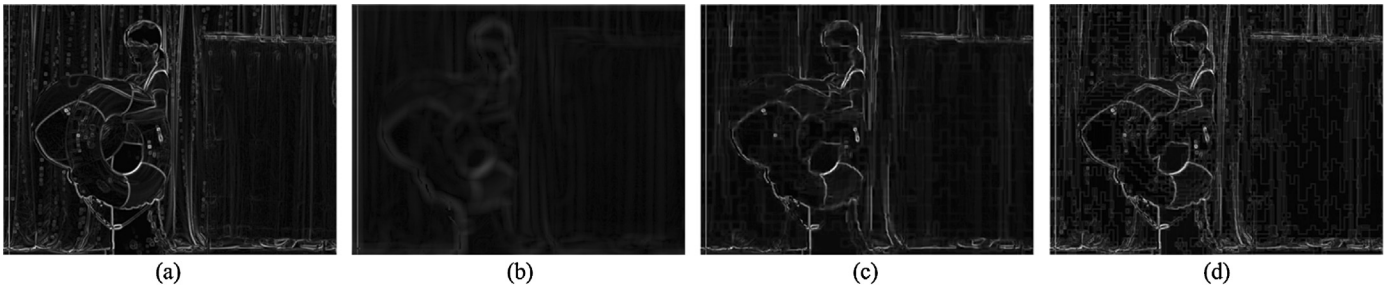

Fig. 2. Examples of BJND results of left images for different types of distortion: (a) Original left image; (b) Gaussian blurred left image; (c) H.264 compressed left image; (d) JPEG compressed left image.

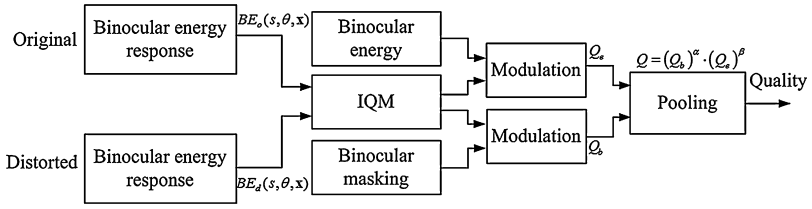

Fig. 3. The framework of the proposed quality assessment metric.

by $bg_r(x+d, y)$ and $eh_r(x+d, y)$. A more detailed representation of the visibility threshold $A_{C,limit}$ is given in [25]. If the noise is omitted, that is, $A_r(x+d, y) = 0$, the visibility thresholds are elevated linearly with the edge height, but the elevating effect decreases as the background luminance increases.

From Eq. (3), since $BJND_l$ for one pixel in the left image is dependent on the corresponding background luminance, edge height and noise amplitude of the matched pixel in the right image, only those matched regions in the left and right images are considered in this paper. Such consideration is reasonable because the BJND model focuses on binocular combination, while the unmatched regions in the left and right images (e.g., the occluded/disoccluded regions in the estimated disparity map) cannot provide clear binocular vision. The matched regions between two views are obtained based on the estimated disparity map in Eqs. (2), (3), assuming that the HVS matches an object in the two views in a way similar to stereo matching. Fig. 2 shows the BJND maps of the left images in Figs. 1(a)–(d), where, to facilitate display, the BJND values are mapped to [0, 255]. We can find that with different distortion strength, the BJND values become different, and this agrees with the sensitivity property of the binocular vision.

From the above analyses, important innovations of the 3D-IQA framework include: (1) the binocular energy response is constructed to measure the strength of response for different retinal points; (2) the BJND model is used to reflect binocular masking effect in measuring image quality. To be more specific, these factors are taken into account in quality assessment by weighing the measured quality score, and combined into an overall score.

3. Proposed quality assessment metric

The framework of the proposed quality assessment metric is illustrated in Fig. 3. Given the original and distorted stereoscopic im-

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003

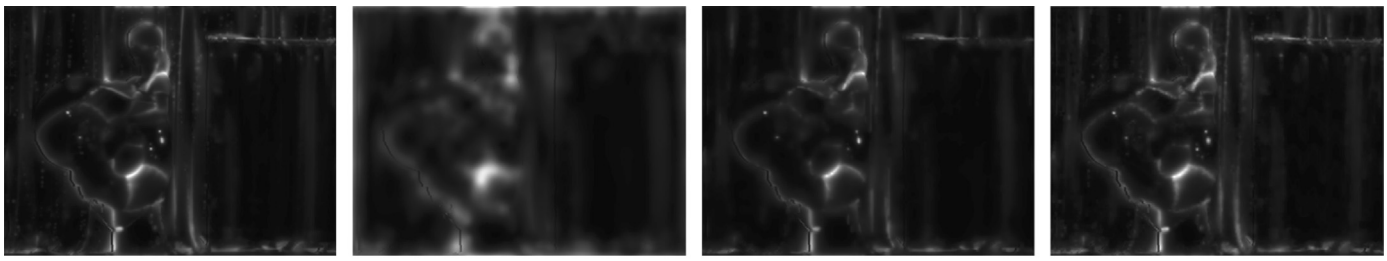

Fig. 4. Examples of binocular energy maps across all scales and orientations: (a) The constructed binocular energy map from the stereoscopic image in Fig. 1(a); (b) The constructed binocular energy map from the stereoscopic image in Fig. 1(b); (c) The constructed binocular energy map from the stereoscopic image in Fig. 1(c); (d) The constructed binocular energy map from the stereoscopic image in Fig. 1(d).

ages, the corresponding binocular energy responses are generated based on Gabor filter responses and the estimated disparity maps. Then, the similarity between the binocular energy responses of the original and distorted stereoscopic images is measured as Image Quality Metric (IQM). Finally, the binocular energy and binocular masking (e.g., BJND) features are used to modulate the IQM, and afterward combination is made to get a total quality score. Therefore, by considering the binocular energy and binocular masking characteristics in the proposed IQM, we hope to produce accurate estimates of the perceptual quality.

3.1. Local phase and amplitude features

As discussed in Section 2, the response functions, $C_l(x) = \rho_l(x)e^{i\varphi_l(x)}$ and $C_r(x) = \rho_r(x)e^{i\varphi_r(x)}$ are used to represent the energy responses. In this work, different with [20] that use complex wavelet transform (CWT) to calculate the energy, we apply log-Gabor filter on the left and right images to model binocular response. Previous researches have shown that simple cells in the primary visual cortex can be well-modeled using log-Gabors [26]. A set of responses at location $\mathbf{x}$ on different scales and along different orientations, denoted as $[\eta_{s,\theta}(\mathbf{x}), \zeta_{s,\theta}(\mathbf{x})]$, can be obtained by applying the log-Gabor filter $G_{s,\theta}$ (denoted by spatial scale index s and orientation index θ) in the Fourier frequency domain

$$G_{s,\theta}(\omega,\theta) = \exp\left(-\frac{(\log(\omega/\omega_0))^2}{2\sigma_r^2}\right)\cdot\exp\left(-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\right) \tag{4}$$

where the parameters ω and θ are the normalized radial frequency and the orientation angle of the filter, and ω_s and θ_s are the corresponding center frequency and orientation of the filter, respectively.

Then, the local amplitude (LA) at location $\mathbf{x}$ on different scale s along different orientation θ is given by

$$LA(s,\theta,\mathbf{x}) = \sqrt{\eta_{s,\theta}(\mathbf{x})^2 + \zeta_{s,\theta}(\mathbf{x})^2} \tag{5}$$

Similarly, the local phase (LP) at location $\mathbf{x}$ on different scale s along different orientation θ is given by

$$LP(s,\theta,\mathbf{x}) = \arg\tan\big(\eta_{s,\theta}(\mathbf{x}), \zeta_{s,\theta}(\mathbf{x})\big) \tag{6}$$

Regarding the parameter selection of log-Gabor filter, $\omega_0 = 1/6$, $\theta_0 = 0$, $\sigma_r = 0.3$, and $\sigma_\theta = 0.4$. In the experiment, the numbers of scale and orientation of the filter are all set to 4. The design of the filter is based on the work in [27].

3.2. Binocular energy response

As discussed earlier, the quality of the stereoscopically viewed 3D image generally cannot be predicted based on the average quality of the two individual images [28]. According to the definition in Eq. (2), we use the LA and LP of the filter responses to predict the binocular energy response. The binocular energy response for the distorted stereoscopic image at location $\mathbf{x}$ on different scale s along different orientation θ is given by

$$\begin{aligned} BE_d(s,\theta,\mathbf{x}) = {} & \big(LA_d^l(s,\theta,\mathbf{x})\big)^2 + \big(LA_d^r(s,\theta,\mathbf{x}')\big)^2 \\ & + LA_d^l(s,\theta,\mathbf{x})\cdot LA_d^r(s,\theta,\mathbf{x}')\cdot\cos\big(\Delta\varphi_d(s,\theta,\mathbf{x})\big) \end{aligned} \tag{7}$$

where the subscripts "o" and "d" denote the original and the distorted images, respectively, the superscripts "l" and "r" denote the left and right images, respectively, $\mathbf{x}$ and $\mathbf{x}'$ are the matched locations in the left and right images, and $\Delta\varphi_d(s,\theta,\mathbf{x}) = LP_d^l(s,\theta,\mathbf{x}) - LP_d^r(s,\theta,\mathbf{x}')$. In this work, we use state-of-the-art stereo matching algorithm in [29] to estimate the disparity from reference stereoscopic images. Similarly, the binocular energy response $BE_o(s,\theta,\mathbf{x})$ for the original stereoscopic image can be calculated by the same manner. Fig. 4 shows the constructed binocular energy maps across all scales and orientations for the stereoscopic images in Figs. 1(a)–(d). It is obvious that the important features (e.g., edges and contours) are preserved in Figs. 4(a)–(d). Thus, the similarity between the binocular energy responses of the original and the distorted images is expected to give a reasonable estimation of quality degradation.

3.3. Image quality measure

Since image distortions will lead to unequal distribution of the binocular energy response with different levels of change, to balance the inconsistence, their phase shift is defined as

$$ES(s,\theta,\mathbf{x}) = \arccos\left(\frac{2BE_d(s,\theta,\mathbf{x})\cdot BE_o(s,\theta,\mathbf{x}) + T_1}{BE_d(s,\theta,\mathbf{x})^2 + BE_o(s,\theta,\mathbf{x})^2 + T_1}\right) \tag{8}$$

where T_1 is a positive constant to increase the stability of the energy shift. In this paper, $T_1 = 16$.

Then, the phase shift across all scales and orientations are calculated as the IQM

$$IQM(\mathbf{x}) = \sum_s\sum_\theta ES(s,\theta,\mathbf{x}) \tag{9}$$

Based on the analysis in Fig. 4, binocular energy can reflect the perceptual importance in binocular vision. That is, if a pixel has a significant binocular energy value, it implies that this position $\mathbf{x}$ will have a high impact on binocular response. Therefore, we use the binocular energy as a modulation component for IQM pooling. The binocular energy modulated quality score is calculated by

$$Q_e = \frac{\sum_{\mathbf{x}\in R_{mtr}} w_e(\mathbf{x})\cdot IQM(\mathbf{x})}{\sum_{\mathbf{x}\in R_{mtr}} w_e(\mathbf{x})} \tag{10}$$

where $w_e(\mathbf{x}) = BE_n(\mathbf{x})$ denotes the weight at location $\mathbf{x}$, and R_{mtr} indicates the matched regions in the estimated disparity map from reference stereoscopic images, because those unmatched regions

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003

(e.g., the occluded/disoccluded regions in the estimated disparity map) cannot produce clear binocular vision. Since the matched regions are obtained based on the estimated disparity map from reference stereoscopic images, the proposed IQM can be applied on symmetrically or asymmetrically distorted stereoscopic images. Considering that it is the distorted stereoscopic image that are delivered to the viewers, the possible overlapping effects between the binocular energy responses from both the original and the distorted stereoscopic should be considered [30]. Therefore, the $BE_n(\mathbf{x})$ is calculated by

$$BE_n(\mathbf{x}) = \frac{BE_o(\mathbf{x}) + BE_d(\mathbf{x})}{2} - min\{BE_o(\mathbf{x}), BE_d(\mathbf{x})\} \tag{11}$$

where $BE_o(\mathbf{x}) = \sum_s \sum_\theta BE_o(s,\theta,\mathbf{x})$ and $BE_d(\mathbf{x}) = \sum_s \sum_\theta BE_d(s,\theta,\mathbf{x})$ denote the binocular energy values across all scales and orientations of the original and distorted stereoscopic images at location $\mathbf{x}$, respectively.

From another perspective, the retinal points with higher binocular energy response will have lower masking value, but the property is not absolute (e.g., luminance and contrast masking effects). The BJND model can be used to reflect the visual sensitivity of different retinal points. Intuitively, if a pixel has a significant BJND value, it implies that this pixel can tolerate large distortion, and thus the importance of the pixel in binocular perception will be low. Therefore, it would be more effective to impose a higher weight on the perceptually important pixels, and from this perspective, the BJND modulated quality score is calculated by

$$Q_b = \frac{\sum_{\mathbf{x}\in R_{mtr}} w_b(\mathbf{x}) \cdot IQM(\mathbf{x})}{\sum_{\mathbf{x}\in R_{mtr}} w_b(\mathbf{x})} \tag{12}$$

where $w_b(\mathbf{x})$ denotes the weight at location $\mathbf{x}$, and $w_b(\mathbf{x}) = 1/BJND_l(\mathbf{x})$; $BJND_l(\mathbf{x})$ denotes the BJND value of the distorted left image at location $\mathbf{x}$.

Finally, considering that the above binocular energy and BJND features are necessary in modulating quality score, and the relationship between them is not a simple superimposed, we investigate nonlinear combination of the two quality scores Q_b and Q_e to obtain the final quality score

$$Q = (Q_e)^\alpha \cdot (Q_b)^\beta \tag{13}$$

where α and β are parameters adjusting the importance of binocular energy and BJND features. In this paper, we train these parameters by optimizing the evaluation results between the objective and subjective scores.

4. Experimental results and analyses

4.1. Stereoscopic image quality databases

In the experiment, we have used two databases, NBU 3D IQA database [31,32] and LIVE 3D IQA database [9], to verify the performance of the proposed metric. The NBU 3D IQA database has been released and can be downloaded at http://cise.nbu.edu.cn/MPC-lab/resourse.htm. The NBU 3D IQA database consists of 312 distorted stereoscopic pairs generated from 12 reference images. Five types of distortions are symmetrically applied to the reference images at various levels: Gaussian Blur (60), White Noise (60), JPEG (60), JPEG2000 (60) and H.264 (72). The LIVE 3D IQA database consists of 365 distorted stereoscopic pairs generated from 20 reference stereoscopic images. Five types of distortions are also symmetrically applied to the reference images at various levels: Gaussian Blur (45), White Noise (80), JPEG (80), JPEG2000 (80) and Fast Fading (80). Each image in these databases has been evaluated by human subjects, and then assigned a quantitative subjective quality score: Difference Mean Opinion Score (DMOS).

Note that the two databases differ in the following respects: (1) the reference images in the LIVE 3D IQA database are acquired using an advanced terrestrial range scanner, while the reference images in the NBU 3D IQA database are chosen from the MPEG test set; (2) for distortion stimulus, H.264 compression and Fast Fading distortions are not all included in the two databases; (3) for subjective tests, images in the NBU 3D IQA database are displayed on screens through a duality stereoscopic projection system, while images in the LIVE 3D IQA database are displayed on passive 3D monitor; (4) the subjective test methodology of the NBU 3D IQA database is Double Stimulus Continuous Quality Scale (DSCQS), while Single Stimulus Continuous Quality Evaluation (SSCQE) is used in the LIVE 3D IQA database. Thus, experiments on the two databases can comprehensively reflect the performance of the proposed metric.

4.2. Parameters determination and performance measurement

Four commonly used performance indicators are employed to evaluate the metrics: Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SRCC), Kendall rank-order correlation coefficient (KRCC), and root mean squared error (RMSE), between the objective scores after nonlinear regression and the subject scores. Among these four criteria, SRCC and KRCC are employed to assess prediction monotonicity, and PLCC and RMSE are used to evaluate prediction accuracy. For a perfect match between the objective and subjective scores, PLCC = SRCC = KRCC = 1 and RMSE = 0. For the nonlinear regression, we use the following five-parameter logistic function [33]

$$DMOS_p = \beta_1 \cdot \left(\frac{1}{2} - \frac{1}{1+\exp(\beta_2 \cdot (x-\beta_3))}\right) + \beta_4 \cdot x + \beta_5 \tag{14}$$

where β_1, β_2, β_3, β_4 and β_5 are determined by using the subjective scores and the objective scores.

In the proposed scheme, we determine the parameters α and β in Eq. (13) by training to optimize the PLCC values between the objective and subjective scores. In the experiments, we select a small subset of the NBU 3D IQA database (House, Puppy, and Soccer2 test sequences) to train the parameters. For simplicity, the parameters are chosen by linear regression optimization. The parameter determination results are $\alpha = 0.0505$ and $\beta = 0.7755$. As expected, the binocular masking component is more important than the binocular energy component (i.e., $\alpha < \beta$). In the following experiments, the proposed scheme is tested on the remaining test sequences in the NUB 3D IQA database (in this way, we avoid same sequences for training and testing). Thus, totally 234 distorted stereoscopic images are adopted in the NUB 3D IQA database. The same parameters are adopted in the LIVE 3D IQA database to guarantee training and testing sets be not the same.

In order to evaluate the performance of the proposed scheme, we compare with the existing state-of-the-art schemes, including four 2D-IQA schemes, i.e., signal-to-noise ratio (PSNR), VIF [6], multi-scale SSIM (MS-SSIM) [34] and mean singular value decomposition (MSVD) [35], and three 3D-IQA schemes, i.e., Benoit's scheme [11], You's scheme [12] and Wang's scheme [19]. The former four schemes directly estimate the quality of each view separately and generate a weighted average score. For Benoit's scheme, we adopt the d_1 metric in the paper, in which the 2D image quality metric is the average result of the left and right images using SSIM, and disparity distortion is the global correlation coefficient between the original and distorted disparity maps. For You's scheme, we adopt the best combination approach in the paper, in which the image quality metric (IQM) is the average result of the left and right images using universal quality index (UQI) [36], and depth quality metric (DQM) is the quality result of the disparity

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003

Table 1
Performance comparison of the eight schemes (the cases in bold: the best performance).

Database	Metric	PSNR	MS-SSIM	VIF	MSVD	Benoit [11]	You [12]	Wang [19]	Proposed
NBU 3D	PLCC	0.8132	0.8067	0.8464	0.8605	0.7936	0.7435	0.7578	**0.9220**
	SRCC	0.8914	0.8728	0.8982	0.9033	0.8529	0.7762	0.8740	**0.9371**
	KRCC	0.6950	0.6719	0.7103	0.7248	0.6565	0.5668	0.6711	**0.7750**
	RMSE	9.4784	9.6244	8.6735	8.2966	9.9091	10.8905	10.6256	**6.6533**
LIVE 3D	PLCC	0.7159	0.8290	0.9247	0.9127	0.8867	0.9162	0.8676	**0.9295**
	SRCC	0.7071	0.9258	0.9195	0.9029	0.8854	0.9248	0.8922	**0.9297**
	KRCC	0.5154	0.7524	0.7388	0.7169	0.6876	0.7503	0.6972	**0.7574**
	RMSE	11.4489	9.1701	6.2422	6.7009	7.5810	6.3206	8.1544	**6.0472**

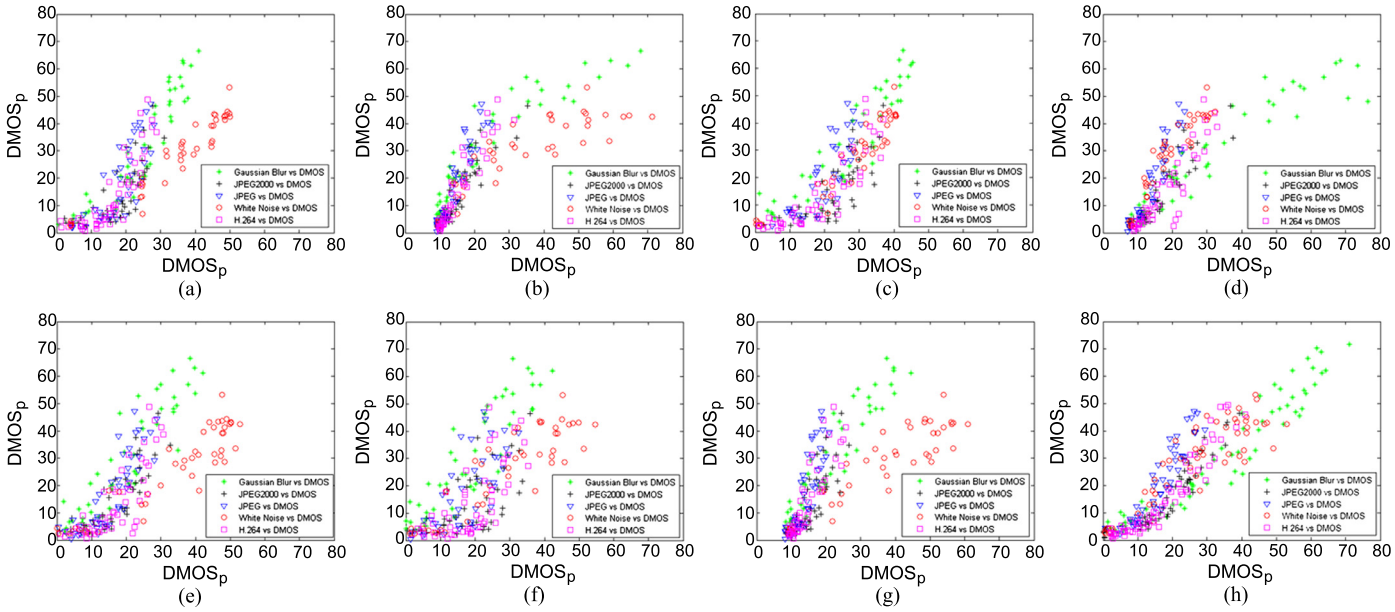

Fig. 5. Scatter plots of objective scores vs. subjective scores of the eight schemes on the NUB 3D IQA database: (a) PSNR; (b) MS-SSIM; (c) VIF; (d) MSVD; (e) Benoit's scheme [11]; (f) You's scheme [12]; (g) Wang's scheme [19]; (h) Proposed scheme.

map using SSIM. To benchmarking these schemes, the same stereo matching method [29] have been used for disparity generation.

4.3. Overall assessment performance

Table 1 lists the performance evaluation results of the eight schemes on the two databases. The best results across the eight schemes for each database are highlighted in boldface. From the table we can see that the proposed scheme outperforms the other schemes. For PSNR, MS-SSIM, VIF and MSVD schemes, since they are directly extended from the 2D case and do not take the binocular visual characteristics into account, the overall performance is worse than the proposed scheme. For Benoit's and You's schemes, they are the combination of 2D-IQA metrics for stereoscopic images and disparity maps. The performance of the two schemes is lower than the proposed scheme on the NBU 3D IQA database, but is comparatively good on the LIVE 3D IQA database. The reason is that the quality of the estimated disparity is highly dependent on the stereo matching algorithms, but UQI has the best performance on the LIVE 3D IQA database (in agreement with the result in [9]). Thus, the overall combined performance will be weakened. This is another demonstration of the disadvantage in the 2D quality metric for disparity maps. For Wang's scheme, the overall performance is not very high because uniform assessment is adopted for the corresponding and non-corresponding regions of the left and right images, while in the proposed scheme, only the matched regions are evaluated, and the binocular response and the binocular masking features are used to modulate the quality score. Figs. 4 and 5 show the scatter plots of predicted quality scores against subjective quality scores (in terms of DMOS) of the eight schemes on the two databases. Overall, the proposed scheme has an impressive consistency with human perception.

4.4. Performance on individual distortion types

In this testing, to more comprehensively evaluate the prediction performance of the proposed metric, we compare the performance of competing methods on each type of distortion. To save space, only the results of PLCC and SRCC are presented in Table 2 and Table 3. For each type of distortion, the best (highest) value across the eight schemes is highlighted in boldface. Even though some schemes may be effective for some special types of distortion, e.g., PSNR scheme is more effective for WN distortion and MSVD scheme is more effective for JPEG distortion on the NBU 3D IQA database, the proposes scheme is more stable across different distortion types on the NBU 3D IQA database. Since the parameters α and β are trained from the NBU 3D IQA database, the performance of the proposed scheme on the LIVE 3D IQA database may be affected by different distortion stimulus, subjective tests and test methodologies of the two databases, and thus the proposed scheme has variable performance across different distortion types but the overall performance is prominent on the LIVE 3D IQA database. Referring to the scatter plots in Figs. 5 and 6, the scatter

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003

Table 2
PLCC performance comparison of the eight schemes (the cases in bold: the best performance).

Distortion	PSNR	MS-SSIM	VIF	MSVD	Benoit [11]	You [12]	Wang [19]	Proposed
NBU 3D IQA database								
GB	0.8879	0.7689	0.9124	0.9430	0.9149	0.9054	0.9475	**0.9573**
WN	**0.9570**	0.8858	0.9362	0.9440	0.9197	0.8830	0.9234	0.9083
JPEG	0.7353	0.9019	0.7951	**0.9368**	0.8370	0.7522	0.8566	0.9171
JP2K	0.7054	0.7130	0.7379	**0.9233**	0.8243	0.6623	0.8158	0.9200
H.264	0.8394	0.8483	0.8575	0.9267	0.8038	0.7062	0.8908	**0.9429**
LIVE 3D IQA database								
GB	0.8395	0.9443	0.9574	0.9190	0.9154	0.9543	0.9280	**0.9553**
WN	0.9262	0.9052	0.9012	**0.9460**	0.9160	0.9369	0.9390	0.9400
JPEG	0.2238	**0.6242**	0.6196	0.4162	0.4774	0.6147	0.4244	0.6033
JP2K	0.6978	0.9143	0.8972	0.8853	0.8772	**0.9311**	0.8641	0.9166
FF	0.6673	0.7171	0.8393	0.7307	0.7132	**0.8490**	0.7296	0.8284

Table 3
SRCC performance comparison of the eight schemes (the cases in bold: the best performance).

Distortion	PSNR	MS-SSIM	VIF	MSVD	Benoit [11]	You [12]	Wang [19]	Proposed
NBU 3D IQA database								
GB	0.9445	0.9043	0.9444	0.9494	0.9153	0.9112	0.9485	**0.9654**
WN	**0.9588**	0.8014	0.9378	0.9490	0.8791	0.8561	0.8700	0.8710
JPEG	0.8666	0.9334	0.8874	0.9338	0.8824	0.8087	0.8799	**0.9592**
JP2K	0.8760	0.8535	0.8518	0.9353	0.8779	0.6876	0.8619	**0.9693**
H.264	0.8871	0.8928	0.8957	0.9047	0.8227	0.7279	0.8862	**0.9540**
LIVE 3D IQA database								
GB	0.7361	0.9231	0.9341	0.8900	0.8726	**0.9323**	0.9126	0.9258
WN	0.9317	0.9433	0.9316	**0.9459**	0.9388	0.9397	0.9395	0.9392
JPEG	0.2297	**0.6252**	0.5808	0.3879	0.4756	0.6034	0.3875	0.5887
JP2K	0.7310	0.8932	**0.9007**	0.8902	0.8667	0.8983	0.8626	0.8941
FF	0.6011	0.7424	0.8053	0.7162	0.6106	**0.8172**	0.6298	0.7973

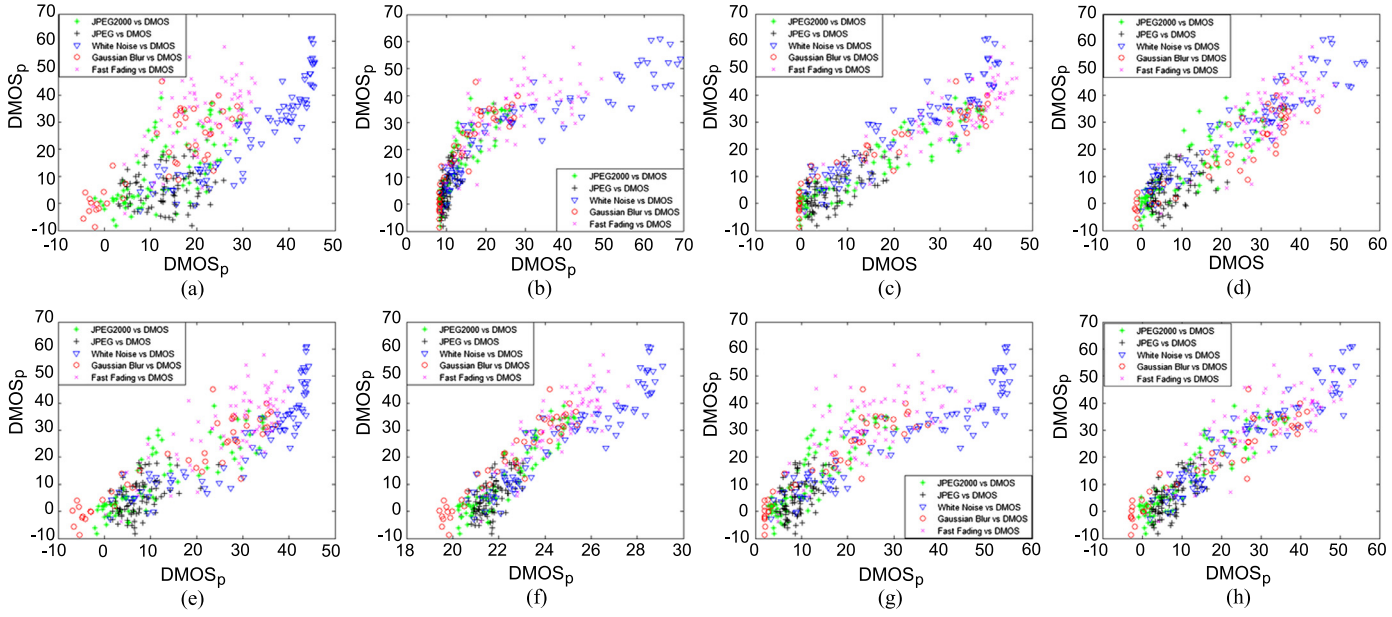

Fig. 6. Scatter plots of objective scores vs. subjective scores of the eight schemes on the LIVE 3D IQA database: (a) PSNR; (b) MS-SSIM; (c) VIF; (d) MSVD; (e) Benoit's scheme [11]; (f) You's scheme [12]; (g) Wang's scheme [19]; (h) Proposed scheme.

plot of the proposed scheme is more concentrated across different groups of distortion types.

4.5. Impact of each components in the proposed scheme

To demonstrate the impact of each component in the proposed scheme, we design three different schemes for comparison, denoted by Scheme-A, Scheme-B and Scheme-C, respectively. For Scheme-A, only Eq. (9) is used to evaluate the quality. For Scheme-B, only binocular energy modulation in Eq. (10) is considered, while for Scheme-C, only binocular masking modulation in Eq. (12) is considered, and other operations are similar with the proposed scheme. The results of PLCC and SRCC are presented in Table 4 and Table 5. From the table, we can see that the proposed binocular energy and binocular masking modulation components have important impacts on the performance improvement for some independent distortions, e.g., Scheme-B is more effective for the

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003

Table 4
PLCC comparisons for each component of the proposed scheme.

Database	Scheme	GB	WN	JPEG	JP2K	H.264	FF	All
NBU 3D	Scheme-A	0.9423	0.9109	0.9300	0.8986	0.9090	–	0.9113
	Scheme-B	**0.9645**	**0.9181**	0.9237	**0.9506**	0.9481	–	0.8982
	Scheme-C	0.9631	0.8944	**0.9318**	0.9365	**0.9514**	–	0.9192
	Proposed	0.9573	0.9083	0.9171	0.9200	0.9429	–	**0.9220**
LIVE 3D	Scheme-A	0.9384	**0.9421**	**0.6473**	0.9139	—	0.7652	0.9223
	Scheme-B	0.9531	0.9328	0.5826	0.9051	–	0.8310	0.9201
	Scheme-C	0.9534	0.9373	0.6054	0.9131	–	0.8202	0.9256
	Proposed	**0.9553**	0.9400	0.6033	**0.9166**	–	**0.8384**	**0.9295**

Table 5
SRCC comparisons for each component of the proposed scheme.

Database	Scheme	GB	WN	JPEG	JP2K	H.264	FF	All
NBU 3D	Scheme-A	0.9495	0.8705	0.9573	0.9377	0.9343	–	0.9284
	Scheme-B	0.9609	**0.9079**	0.9537	**0.9710**	0.9391	–	0.9277
	Scheme-C	0.9648	0.8698	**0.9605**	0.9691	**0.9541**	–	0.9370
	Proposed	**0.9654**	0.8710	0.9592	0.9693	0.9540	–	**0.9371**
LIVE 3D	Scheme-A	0.9192	**0.9421**	**0.6248**	**0.8992**	–	0.7330	0.9276
	Scheme-B	0.9232	0.9375	0.5614	0.8936	–	**0.8069**	0.9281
	Scheme-C	**0.9267**	0.9385	0.5875	0.8938	–	0.7949	0.9295
	Proposed	0.9258	0.9392	0.5887	0.8941	–	0.7943	**0.9297**

GB, WN and JP2K distortions on the NBU 3D IQA database, and Scheme-C is more effective for the JPEG and H.264 distortions on the NBU 3D IQA database. Even thought the performance improvement is limited by using the above modulations on the LIVE 3D IQA database, the overall performance can be gradually promoted by the proposed scheme.

5. Conclusions

This paper has presented a quality assessment method of stereoscopic image based on binocular energy response. Compared with the existing two-dimensional (2D) metrics, the technical contribution of the proposed method is that we try to use binocular energy and binocular masking features to quantify the binocular visual characteristics. The prominent advantage of the proposed method is as follows: 1) We construct binocular energy responses of the original and distorted images, respectively, and measure the similarity between them as Image Quality Metric (IQM); 2) By considering the binocular response and binocular masking characteristics, we use the binocular energy and the binocular just noticeable difference (BJND) components to modulate the IQM, respectively; 3) Two evaluation results are nonlinearly integrated into an overall score by considering the importance of each component. Experimental results show that the proposed method can achieve much higher consistency with the subjective assessments. In the future work, to further advance the performance of the proposed method, more comprehensive study of various binocular visual characteristics (e.g., visual attention and visual comfort) should be considered in quality assessment.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 61271021, 61271270, U1301257), the Natural Science Foundation of Zhejiang Province (grant Y1111061). It was also sponsored by K.C. Wong Magna Fund in Ningbo University.

References

[1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, M. Lang, Three-dimensional video postproduction and processing, Proc. IEEE 99 (4) (April 2011) 607–625.

[2] H.T. Quan, P. Le Callet, M. Barkowsky, Video quality assessment: from 2D to 3D – challenges and future trends, in: Proc. of IEEE International Conference on Image Processing, Hong Kong, Sep. 2010, pp. 4025–4028.

[3] M. Lambooij, W. IJsselsteijn, D.G. Bouwhuis, I. Heynderickx, Evaluation of stereoscopic images: beyond 2D quality, IEEE Trans. Broadcast. 57 (2) (June 2011) 432–444.

[4] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (April 2004) 600–612.

[5] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise-ratio for natural images, IEEE Trans. Image Process. 16 (Sept. 2007) 2284–2298.

[6] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (Feb. 2006) 430–444.

[7] M. Lambooij, M. Fortuin, W.A. IJsselsteijn, B.J. Evans, I. Heynderickx, Susceptibility to visual discomfort of 3-D displays by visual performance measures, IEEE Trans. Circuits Syst. Video Technol. 21 (12) (Dec. 2011) 1913–1923.

[8] W. Chen, J. Fournier, M. Barkowsky, P. Le Callet, Quality of experience model for 3DTV, in: Stereoscopic Displays and Applications, vol. XXIII, in: Proc. of SPIE, vol. 8288, San Jose, CA, USA, Feb. 2012.

[9] A.K. Moorthy, C.C. Su, A. Mittal, A.C. Bovik, Subjective evaluation of stereoscopic image quality, Signal Process. Image Commun. 28 (8) (Dec. 2013) 870–883.

[10] L. Goldmann, F. De Simone, T. Ebrahimi, A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video, in: Three-Dimensional Image Processing and Applications, Proc. of SPIE, Feb. 2010, p. 75260S.

[11] A. Benoit, P. Le Callet, P. Campisi, P. Campisi, R. Cousseau, Quality assessment of stereoscopic images, EURASIP J. Image Video Process. 2008 (2008), Article ID 659024, 23 pp.

[12] J. You, L. Xing, A. Perkis, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, in: Proc. of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA, 2010.

[13] K. Ha, M. Kim, A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video, in: Proc. of IEEE International Conference on Image Processing, Brussels, Belgium, Sep. 2011, pp. 2525–2528.

[14] C.T.E.R. Hewage, S.T. Worrall, S. Dogan, S. Villette, A.M. Knodoz, Quality evaluation of color plus depth map-based stereoscopic video, IEEE J. Sel. Top. Signal Process. 3 (2) (April 2009) 304–318.

[15] A. Maalouf, M.C. Larabi, CYCLOP: a stereo color image quality assessment metric, in: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, May 2011, pp. 1161–1164.

[16] Y. Lin, J. Wu, Quality assessment of stereoscopic 3D image compression by binocular integration behaviors, IEEE Trans. Image Process. 23 (4) (April 2014) 1527–1542.

[17] S. Ryu, D.H. Kim, K. Sohn, Stereoscopic image quality metric based on binocular perception model, in: Proc. of IEEE Intentional Conference on Image Processing, 2012.

[18] H. Ko, C.-S. Kim, S. Choi, C.-C. Jay Kuo, 3D image quality index using SDP-based binocular perception model, in: Proc. of IEEE IVMSP Workshop, Seoul, Korea, June 2013.

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003

[19] X. Wang, S. Kwong, Y. Zhang, Considering binocular spatial sensitivity in stereoscopic image quality assessment, in: Proc. of IEEE Visual Communications and Image Processing, Tainan City, Taiwan, Nov. 2011.
[20] R. Bensalma, M.C. Larabi, A perceptual metric for stereoscopic image quality assessment based on the binocular energy, Multidimens. Syst. Signal Process. 23 (1–2) (2012) 1–36.
[21] O.J. Braddick, Binocular single vision and perceptual processing, Proc. R. Soc. Lond. B, Biol. Sci. 204 (1157) (June 1979) 503–512.
[22] D.J. Fleet, H. Wagner, D.J. Heeger, Neural encoding of binocular disparity: energy models, position shifts and phase shifts, Vis. Res. 36 (12) (June 1996) 1839–1857.
[23] R.D. Henkel, Fast stereovision by coherence detection, Lect. Notes Comput. Sci. 1296 (1997) 297–304.
[24] A.B. Watson, J.A. Solomon, Model of visual contrast gain control and pattern masking, J. Opt. Soc. Am. A 14 (9) (Sep. 2007) 2379–2391.
[25] Y. Zhao, Z. Chen, C. Zhu, Y.P. Tan, L. Yu, Binocular just-noticeable-difference model for stereoscopic images, IEEE Signal Process. Lett. 18 (1) (Jan. 2011) 19–22.
[26] D.J. Field, Relations between the statistics of natural images and the response properties of cortical cells, J. Opt. Soc. Am. A 4 (12) (1987) 2379–2394.
[27] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (Aug. 2011) 2378–2386.
[28] M.-J. Chen, C.C. Su, D.K. Kwon, L.K. Cormack, A.C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, in: Proc. of Forty-Sixth Annual Asilomar Conference on Signals, Systems, and Computers, Monterey, California, Nov. 2012.
[29] V. Kolmogorov, R. Zabih, Computing visual correspondence with occlusions using graph cuts, in: Proc. of IEEE International on Computer Vision, vol. 2, Vancouver, Canada, Jul. 2001, pp. 508–515.
[30] K. Gu, G.T. Zhai, X.K. Yang, L. Chen, W.J. Zhang, Nonlinear additive model based saliency map weighting strategy for image quality assessment, in: Proc. of IEEE 14th International Workshop on Multimedia Signal Processing, The Banff Park Lodge, Canada, Sep. 2012, pp. 313–318.
[31] J. Zhou, G. Jiang, X. Mao, M. Yu, F. Shao, Z. Peng, Y. Zhang, Subjective quality analyses of stereoscopic images in 3DTV system, in: Proc. of IEEE Visual Communications and Image Processing, Tainan City, Taiwan, Nov. 2011.
[32] F. Shao, W. Lin, S. Gu, G. Jiang, T. Srikanthan, Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics, IEEE Trans. Image Process. 22 (5) (May 2013) 1940–1953.
[33] P.G. Gottschalk, J.R. Dunn, The five-parameter logistic: a characterization and comparison with the four-parameter logistic, Anal. Biochem. 343 (1) (Aug. 2005) 54–65.
[34] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, IEEE Asilomar Conf. Signals, Systems, and Computers (2003) 1398–1402.
[35] A. Shnayderman, A. Gusev, A.M. Eskicioglu, An SVD-based grayscale image quality measure for local and global assessment, IEEE Trans. Image Process. 15 (2) (2006) 422–429.
[36] Z. Zhou, A.C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (March 2002) 81–84.

Feng Shao received his B.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in Electronic Science and Technology. He is currently an Associate Professor in Faculty of Information Science and Engineering, Ningbo University, China. He was a visiting Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore, from February 2012 to August 2012. His research interests include 3D video coding, 3D quality assessment, and image perception, etc.

Gangyi Jiang received his M.S. degree from Hangzhou University in 1992, and received his Ph.D. degree from Ajou University, Korea, in 2000. He is now a Professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include digital video compression, multi-view video coding, etc.

Mei Yu received her M.S. degree from Hangzhou Institute of Electronics Engineering, China, in 1993, and Ph.D. degree from Ajou University, Korea, in 2000. She is now a Professor in Faculty of Information Science and Engineering, Ningbo University, China. Her research interests include image/video coding and video perception.

Fucui Li received her M.S. degree from HeFei University of Technology, China, in 2004, and is currently pursuing the Ph.D. degree in Ningbo University, China. Her research interests include image/video coding and video perception.

Zongju Peng received his B.S. degree from Sichuan Normal College, China, in 1995, and M.S. degree from Sichuan University, China, in 1998, and received his Ph.D. degree from Institute of Computing Technology, Chinese Academy of Science, in 2010. He is now an Associate Professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include image/video compression, 3D video coding and video perception.

Randi Fu received his M.S. degree from the PLA Information Engineering University, China, in 2001. He is now an Associate Professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include remote sensing image processing and pattern recognition.

Please cite this article in press as: F. Shao et al., Binocular energy response based quality assessment of stereoscopic images, Digital Signal Process. (2014), http://dx.doi.org/10.1016/j.dsp.2014.03.003