Perceptual Full-Reference Quality Assessment of Stereoscopic Images by Considering Binocular Visual Characteristics

Feng Shao, Weisi Lin, *Senior Member, IEEE*, Shanbo Gu, Gangyi Jiang, *Member, IEEE*, and Thambipillai Srikanthan, *Senior Member, IEEE*

***Abstract*— Perceptual quality assessment is a challenging issue in 3D signal processing research. It is important to study 3D signal directly instead of studying simple extension of the 2D metrics directly to the 3D case as in some previous studies. In this paper, we propose a new perceptual full-reference quality assessment metric of stereoscopic images by considering the binocular visual characteristics. The major technical contribution of this paper is that the binocular perception and combination properties are considered in quality assessment. To be more specific, we first perform left–right consistency checks and compare matching error between the corresponding pixels in binocular disparity calculation, and classify the stereoscopic images into non-corresponding, binocular fusion, and binocular suppression regions. Also, local phase and local amplitude maps are extracted from the original and distorted stereoscopic images as features in quality assessment. Then, each region is evaluated independently by considering its binocular perception property, and all evaluation results are integrated into an overall score. Besides, a binocular just noticeable difference model is used to reflect the visual sensitivity for the binocular fusion and suppression regions. Experimental results show that compared with the relevant existing metrics, the proposed metric can achieve higher consistency with subjective assessment of stereoscopic images.**

***Index Terms*— Binocular fusion, binocular just-noticeable difference, binocular suppression, local amplitude, local phase, stereoscopic image quality assessment.**

I. Introduction

With the advancement of three-dimensional (3D) related technologies, including 3D content creation, 3D video compression, network transmission and stereoscopic display, 3D video applications have drawn increasing attention and are expected to be commercialized in the near future [1]. Stereoscopic imaging consists of two images (left and right views) captured by two cameras [2]. Depth perception can be realized through analysis of left and right images in the brain [3]–[5]. Compared with the traditional monoscopic image, stereoscopic images deal with two channels of information simultaneously. Image distortion and unwanted side effects existing in 3D scene affect the perceptual quality of stereoscopic images, such as the compression-induced artifacts (e.g., blurring, blockiness), and visual artifacts by stereoscopic display systems. Perceptual stereoscopic image quality assessment (SIQA) is an important issue for evaluating the performance of 3D technologies [6].

Similar to two-dimensional (2D) image quality assessment (2D-IQA), SIQA can be categorized into subjective and objective approaches. The former approach is based on subjective quality assessed by human observers [7], and the latter approach provides an objective index (i.e., generated by a machine) to measure the same. During the recent years, various factors that affect stereoscopic perception have been investigated by subjective experiments. IJsselsteijin [8], Tam [9], Wang [10], Kuijsters [11], and Pourazad [12] analyzed the factors of camera parameters, display duration, asymmetric quality, chroma variations and brightness on the visual quality of 3D videos. However, subjective assessment is time-consuming, laborious, expensive and not applicable to on-line operations. Therefore, objective SIQA in this research is to design an efficient objective criterion that agrees with human visual perception.

2D-IQA has been widely researched [13], and many quality assessment metrics, such as peak signal-to-noise ratio (PSNR), Structural SIMilarity (SSIM) [14], visual information fidelity (VIF) [15], etc., were proposed. During the recent years, the concept of quality of experience (QoE) has been used to reflect the overall experience of the consumer accessing and using the provided service [16]. 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. 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 [17]. Many works have been conducted in evaluating the visual comfort and visual fatigue of stereoscopic images [18], [19]. Chen *et al*. explored

Manuscript received May 1, 2012; revised October 19, 2012; accepted January 1, 2013. Date of publication January 14, 2013; date of current version March 14, 2013. This work was supported in part by the Natural Science Foundation of China under Grant 60902096, Grant 61271021, and Grant 61071120, and the K. C. Wong Magna Fund, Ningbo University. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stefan Winkler.

F. Shao, S. Gu, and G. Jiang are with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: shaofeng@nbu.edu.cn; gushanbo_1988@126.com; jianggangyi@nbu.edu.cn).

W. Lin is with the Centre for Multimedia and Network Technology, School of Computer Engineering, Nanyang Technological University, 639798 Singapore (e-mail: wslin@ntu.edu.sg).

T. Srikanthan is with the Centre for High Performance Embedded Systems, School of Computer Engineering, Nanyang Technological University, 639798 Singapore (e-mail: astsrikan@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2240003

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 [20]. However, the above methods [17]–[20] remained on a subjective level to explore the combination of various perceptual scales. For 3D/stereoscopic objective quality assessment, a more difficult problem than the 2D case is how to handle the added information (e.g., depth sensation) in a 3D/stereoscopic case, in order to measure the perceptual quality of stereoscopic image pair in line with the human vision system (HVS) [21]. To date, a number of objective SIQA methods have been proposed in the literature.

A straightforward way of applying the state-of-the-art 2D-IQA methods to SIQA is to evaluate the two views of the stereoscopic images separately and combine them into an overall score. Yasakethu *et al*. investigated the relationship between subjective quality measures and some 2D objective quality measures for 3D video [22]. Boev *et al*. combined the monoscopic quality component and the stereoscopic quality component for developing a stereo-video quality metric [23]. Since the HVS uses binocular disparity to perceive depth information in stereoscopic images, several SIQA methods were proposed by integrating the relative disparity (depth) map into the metric. You *et al*. investigated the capabilities of some common 2D quality metrics in SIQA, and integrated the disparity information into quality assessment [24]. Benoit *et al*. presented a linear combination for disparity distortion and the measurement of 2D image quality on both views [25]. Ha *et al*. designed a quality assessment method by considering the factors of temporal variation and disparity distribution [26]. Hewage *et al*. performed the evaluation for color plus depth 3D video by using the extracted edge information of depth maps and the extracted information from the corresponding color images [27]. Bosc *et al*. questioned the reliability of 2D objective metrics as well as 2D subjective protocols for the assessment of 3D contents [28]. 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. Thus, the related existing methods [22]–[27] are not appropriate to assess the perceived depth, because they are not consistent with human visual perception.

It has been known that various binocular artifacts, visual masking effects and monocular visual properties can all affect the 3D visual perception. Maalouf *et al*. computed the cyclopean image from left and right images to simulate the brain perception, and used contrast sensitivity coefficients of the cyclopean image as the basis of evaluation [30]. Jin *et al*. grouped the similar blocks from left and right views of stereoscopic video into a 3D stack, and evaluated the quality by 3D-DCT and considering contrast sensitive function and luminance masking [31]. Gorley *et al*. proposed a Stereo Band Limited Contrast (SBLC) algorithm to evaluate the stereoscopic image quality, which accounts for the HVS sensitivity to contrast and luminance changes at regions of high spatial frequencies [32]. Sazzad *et al*. proposed a no-reference perceptual quality assessment for JPEG coded stereoscopic images to evaluate the artifacts and disparity using local feature information, such as edge, flat and texture region classification and blockiness [33]. Hwang *et al*. devised a metric by considering the impact of visual attention, depth variation and stereo distortion prediction, to detect visually significant distortions based on the human visual properties [34]. Qi *et al*. used the correlation between the left and right views of a stereoscopic image pair to reflect the QoE, and represented the structure of low-level features using the phase congruency and saliency map [35]. 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.

Psychophysical researches have indicated that human visual perception is a complex visual process of the binocular fusion, binocular suppression and other factors, and forms the binocular vision from two rental images [36]. The binocular suppression theory has been used as a basis for validating the asymmetric stereoscopic video coding [37]: one view in a stereoscopic image pair can be encoded with lower quality, while the perceptual quality degradation is negligible to human eyes (this theory has been also adopted in 3D video coding in our previous work [38]). Some SIQA methods have been proposed by using the binocular properties. Mittal *et el*. used a machine learning technique for 3D quality evaluation to simulate the process of perceptual characteristics and mechanisms of binocular vision [39]. Wang *et al*. proposed a SIQA metric by considering the binocular spatial sensitivity to reflect the binocular fusion and suppression properties [40], but the process of the binocular fusion and suppression was not considered since only a weighted average of left and right views was used. 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 [41]. However, the binocular visual properties will be not uniformly distributed across the stereoscopic images because distortion effects on visual perception for different regions are not uniform.

In this paper, based upon the human visual characteristics, we propose a perceptual full-reference quality assessment metric for stereoscopic images. We tackle the binocular vision phenomenon that images from two eyes are fused and perceived as a single perception. The important effort of this work is devoted to modeling binocular vision mechanisms and developing a metric taking into account binocular perceptual characteristics. The main contributions of this work are as follows: 1) Analysis of the related binocular visual characteristics has been presented for stereoscopic image quality assessment. 2) By considering the binocular perception property, we classify the stereoscopic images into non-corresponding, binocular fusion and binocular suppression regions, evaluate them distinctively, and integrate them to get a total quality score. 3) By considering the binocular combination properties, we use the binocular just noticeable difference (BJND) to modulate the quality score, in which higher weights are assigned to more perceptually important retinal points. The rest of the paper is organized as follows. Section II analyzes the relevant binocular visual characteristics. Section III

(a) (b) (c) (d)

Fig. 1. Interchanging phase and amplitude of stereoscopic images. (a) Original left image. (b) Original right image. (c) Image constructed using the phase of (a) and the amplitude of (b). (d) Image constructed using the amplitude of (a) and the phase of (b).

presents the proposed perceptual quality assessment metric. The experimental results are given and discussed in Section IV, and finally conclusions are drawn in Section V.

II. Binocular Visual Characteristics Analyses

It has been known that the binocular vision is a complex visual process that requires the brain and both eyes working together to produce depth perception and clear vision [42]. As an example of one-dimensional signals, considering a simple binocular cell with the left and right receptive fields, the binocular energy response to a stereoscopic image pair $I_l(x)$ and $I_r(x)$ at position x can be described as [43]:

$$\begin{aligned} r_q &= \left|\int_{-\infty}^{\infty}[f_l(x)I_l(x) + f_r(x)I_r(x)]dx\right|^2 \\ &= \left|\int_{-\infty}^{\infty} g(x)e^{j\omega x}\left[I_l(x) + e^{i\phi_-}I_r(x)\right]dx\right|^2 \end{aligned} \quad (1)$$

where $f_l(x) = g(x)e^{j(\omega x+\phi_l)}$ and $f_r(x) = g(x)e^{j(\omega x+\phi_r)}$, being Gabor functions for the left and right images, respectively; $g(x)$ is Gaussian kernel function; and $\phi_- = \phi_r - \phi_l$, being the phase difference between the left and right images. The stimulus disparity can be estimated by $D = \hat{\phi}_-/\omega$, where $\hat{\phi}_-$is the phase difference that maximizes the binocular energy response r_q, and ω is the radial frequency of the cell.

From another perspective, if we have known the position shift d between the left and right receptive field centers, the binocular energy response in Eq. (1) can be written as

$$r_q = \left|\int_{-\infty}^{\infty} e^{j\omega x}\left[g(x)I_l(x) + e^{j\omega d}g(x+d)I_r(x)\right]dx\right|^2 \quad (2)$$

and the stimulus disparity is given by $D = \hat{d}$, where $\hat{d}$ is the position shift that maximizes the binocular energy response r_q.

From the above equations, since the stimulus disparity can be estimated by $D = \hat{\phi}_-/\omega$ and $D = \hat{d}$, we can find that both phase difference and position shift can describe the same disparity information in the binocular vision (i.e., $\hat{\phi}_-/\omega = \hat{d}$). In other words, phase difference between left and right images provides the main cue for binocular disparity identification and depth perception in the binocular vision, and the distortion in phase may affect precisely identification of binocular disparity and further affect the perceived depth. We present two examples to illustrate the above phenomena. In the first example, the original left and right images of 'Lovebird1' test sequence are shown in Fig. 1(a) and (b). By interchanging the phase and amplitude, we construct the image in Fig. 1(c) using the phase of Fig. 1(a) and the amplitude of Fig. 1(b), and the image in Fig. 1(d) using the amplitude of Fig. 1(a) and the phase of Fig. 1(b). It can be noticed that the same object in Fig. 1(c) and Fig. 1(d) has different position shifts because the two images have different phase information, while the position shifts are the same in Fig. 1(a) and Fig. 1(c), or in Fig. 1(b) and Fig. 1(d) if two images have the same phase information. This illustrates that the phase conveys the disparity information in the binocular vision.

In the second example, without considering the position shift between images, phase and amplitude have different contributions in determining image quality. The first row of Fig. 2 shows the (a) original, (b) Gaussian blurred and (c) JPEG compressed left images, respectively. The second row of Fig. 2 shows the constructed images using their respective phases of Figs. 2(a)-(c) but constant amplitude (luminance-inverted for better display), while the third row in Fig. 2 shows the constructed images using their respective amplitudes Figs. 2(a)–(c) but constant phase. We can find that most important features such as edges and contours are preserved in Figs. 2(d)-(f), and the structure is degraded due to blurring and JPEG compression. In other words, the constructed images in Figs. 2(g)-(i) convey less useful information although they reflect the blurring and JPEG compression distortions. Existing studies have shown that phase information is very important in feature description [44]–[46]. Therefore, phase similarity (difference) between the original and distorted images is expected to give a reasonable estimation of quality degradation (i.e., phase has a larger impact on the quality score than amplitude, as demonstrated in the next subsection IV.B).

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 [47]. 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 [48]. In the following, we summarize the derivation of the BJND model. By incorporating the luminance and contrast

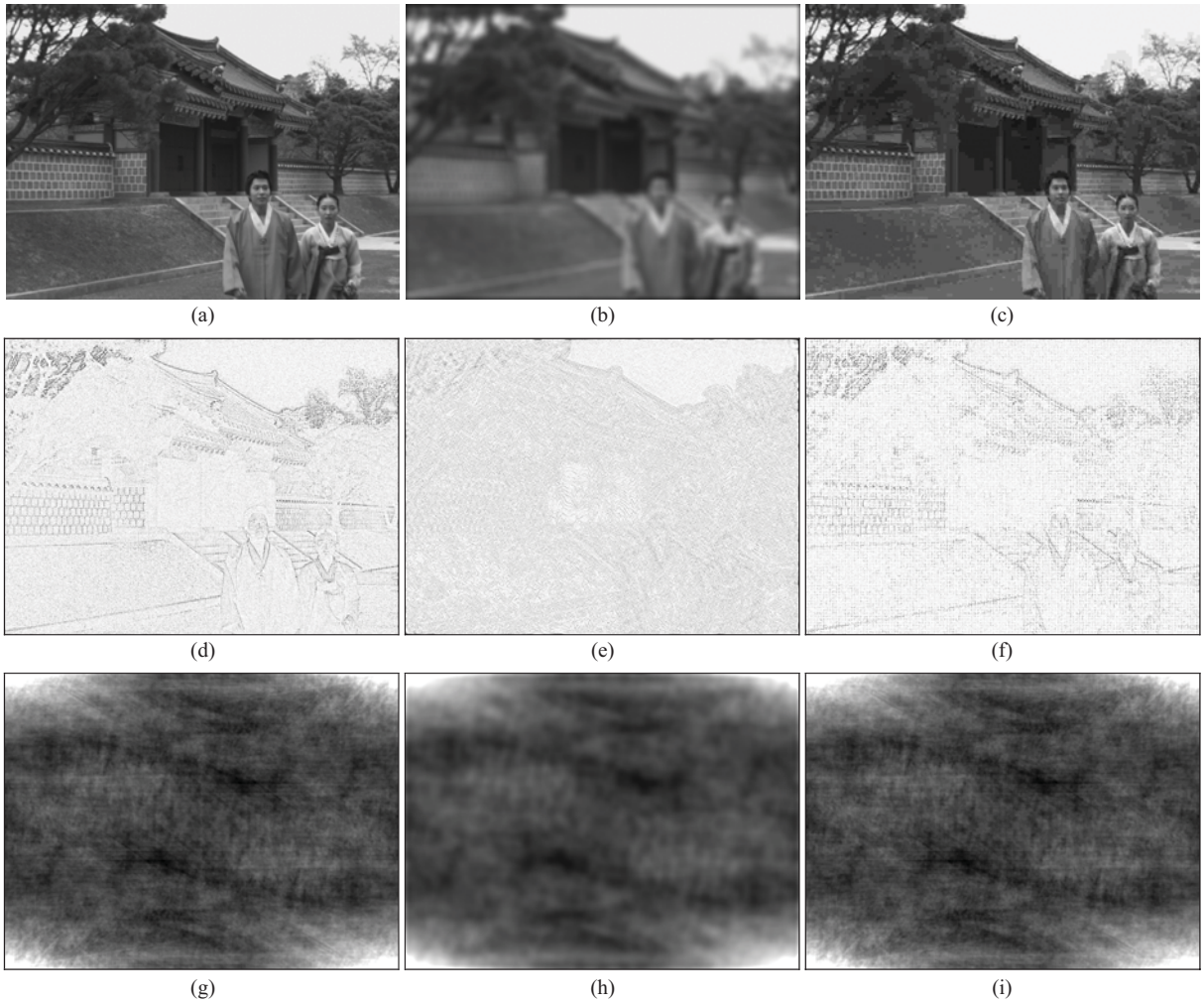


Fig. 2. Example of image construction using phase and amplitude. (a) Original left image. (b) Gaussian blurred image. (c) JPEG compressed image. (d) Image constructed using the phase of (a) and constant amplitude (luminance-inverted for better display). (e) Image constructed using the phase of (b) and constant amplitude (luminance-inverted for better display). (f) Image constructed using phase of (c) and constant amplitude (luminance-inverted for better display). (g) Image constructed using the amplitude of (a) and constant phase. (h) Image constructed using the amplitude of (b) and constant phase. (i) Image constructed using the amplitude of (c) and constant phase.

masking effects, as well as considering the correspondence matching between two views, the BJND at the left view is defined as

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

where d is the horizontal disparity value at pixel (x, y), $bg_r(x+d, y)$ denotes the background luminance level, $eh_r(x+d, y)$ denotes the edge height, $A_r(x+d, y)$ denotes the noise amplitude, and $\lambda = 1.25$ as in [48].

Note that $BJND_l$ is dependent on the background luminance, edge height and noise amplitude. 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; the visibility thresholds are determined as

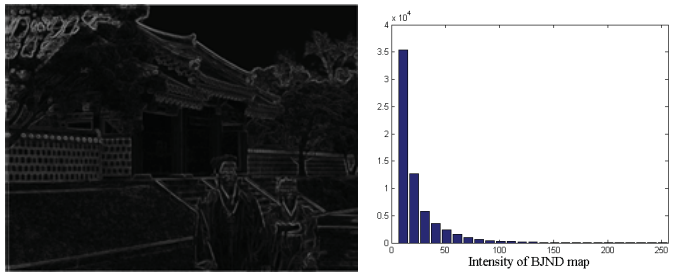
$$A_{C,limit}(bg, eh) = A_{limit}(bg) + K(bg) \cdot eh \tag{4}$$

$$K(bg) = -10^{-6} \cdot (0.7 \cdot bg^2 + 32 \cdot bg) + 0.07 \tag{5}$$

$$A_{limit}(bg) = \begin{cases} 0.0027 \cdot (bg^2 - 96 \cdot bg) + 8, & if\ 0 \le bg < 48 \\ 0.0007 \cdot (bg^2 - 32 \cdot bg) + 1.7, & if\ 48 \le bg \le 255 \end{cases} \tag{6}$$

Eqs. (5)-(6) are the approximation of findings from psychophysical experiments, and the detailed representation of eh can be referred to [48]. Similarly, the BJND of the right view can be obtained. Fig. 3 shows the BJND map of the left image in Fig. 2(a), where, to facilitate display, the BJND values are mapped to [0, 255]. The corresponding histogram for BJND is shown in Fig. 3(b). We also find that with the increasing distortion strength, the BJND values become smaller, and this agrees with the sensitivity property of the binocular vision.

A major limitation of the above BJND model is that it only focuses on binocular combination property of injected noises, luminance masking and contrast masking, and it is assumed that the disparity has minor impact on the binocular combination. However, this assumption is not always held in the binocular vision. In fact, in the process of binoc-



(a) (b)

Fig. 3. BJND results of the left image in Fig. 2(a). (a) BJND map. (b) Histogram of (a).

ular combination, some binocular visual phenomenon, e.g., binocular fusion and binocular rivalry, may occur. According to visual psychophysical research, if the similar contents fall on the corresponding retinal points, binocular fusion will integrate the two retinal points into one single binocular perception [49]. However, when attempting to fuse the two corresponding retinal points with distinctly dissimilar content, they will be in competition with each other. The competition is known as rivalry in visual psychology, i.e., a sense of failed fusion between the left and right images. During rivalry, the alternative fluctuation between left and right eyes will continue a period of time. However, the HVS normally cannot tolerate rivalry for long. It usually reconciles the conflicting data by suppression. The entire image from one retina may be suppressed, but in most cases, parts of the right eye's visual field will be suppressed while other parts of the left eye's visual field are suppressed [50]. This phenomenon is known as binocular suppression. Therefore, binocular vision is generally regarded as a combination of binocular fusion and suppression [50]. In order to facilitate the comparison and analysis in SIQA, we give the definitions of binocular fusion and binocular suppression as follows.

Binocular fusion is a process of superimposing and combining similar content from the two views into one perception.

Binocular suppression is caused when noticeably mismatched or conflict contents are presented to two eyes.

In summary, binocular visual characteristics analyzed in this section have major impact in SIQA: 1) the phase information allows better representation of image features, and provides theoretical basis for identification of binocular disparity, 2) the BJND model, which takes the binocular combination property into account, can be used to reflect visual sensitivity in the binocular vision, 3) the relevant binocular fusion and binocular suppression properties can be applied to simulate the binocular visual response process, for more effective evaluation.

III. Proposed Perceptual Full-Reference Stereoscopic Image Quality Assessment Metric

In the existing full-reference SIQA work, it is usually assumed that original stereoscopic images are well calibrated so that the binocular suppression phenomenon does not occur. If a pixel has no matched correspondence in the other view, monocular vision occurs in non-corresponding retinal points; otherwise, binocular fusion occurs in corresponding retinal points. However, the introduction of distortion in a stereoscopic image affects the identification of binocular disparity and then the binocular perception. Thus, in this work, we categorize the distorted stereoscopic image into three kinds of regions: non-corresponding region, binocular fusion region and binocular suppression region. Each region is evaluated independently based on its binocular perception property, and afterward combination is made to get a total quality score. The procedure of the proposed quality assessment metric is illustrated in Fig. 4.

A. Region Classification

As analyzed in Section II, binocular disparity provides information to classify stereoscopic image regions. In this work, we do not intend to study the estimation methods of disparity maps and their impact on the quality assessment. We use the stereo matching algorithm [51] for disparity calculation since its performance is prominent for high-quality stereoscopic images. The changes between the disparity values from the original and the distorted stereoscopic images are usually located at the positions where the distortions are clearly visible, e.g., noise added regions, because phase degradation in such regions will affect the identification of binocular disparity. In this work, rather than directly comparing the magnitude of the disparity values between the original and the distorted stereoscopic images, we perform left-right consistency check and compare matching error between the corresponding pixels in binocular disparity calculation from the distorted stereoscopic images for region classification. The detailed process of region classification is described as follows.

a) Noncorresponding region detection: In stereoscopic vision, occlusion/disocclusion will inevitably occur due to viewing angle difference. For example, the background behind the foreground in the left image will be disoccluded in the right image. That is, the right image does not provide any information about those background areas in the left one. As a result, occlusion or disocclusion will appear in the estimated disparity maps. The occluded/disoccluded regions in the estimated disparity maps are regarded as non-corresponding region. The non-corresponding region of the distorted stereoscopic images consists of the left non-corresponding region R_{nc}^{l} in the left image and the right non-corresponding region R_{nc}^{r} in the right image. Since R_{nc}^{l} and R_{nc}^{r} do not overlap each other, the non-corresponding region R_{nc} of the distorted stereoscopic images is described as

$$R_{nc} = R_{nc}^{l} \cup R_{nc}^{r} \tag{7}$$

b) Binocular suppression region detection: According to the definition in Section II, binocular suppression occurs when (1) the contents in left and right images are mismatched, (2) the matched contents are not consistent, i.e., with large matching error, even though they are correctly matched. Based on the above two considerations, we define the blending condition under which the binocular suppression region is detected. Let $d_L(x, y)$ denote the left disparity map and $d_R(x, y)$ denote the right disparity map, left-right consistency check is used to identify inconsistent pixels in the disparity maps. The pixels are mismatched if the sum of the matched disparity values is

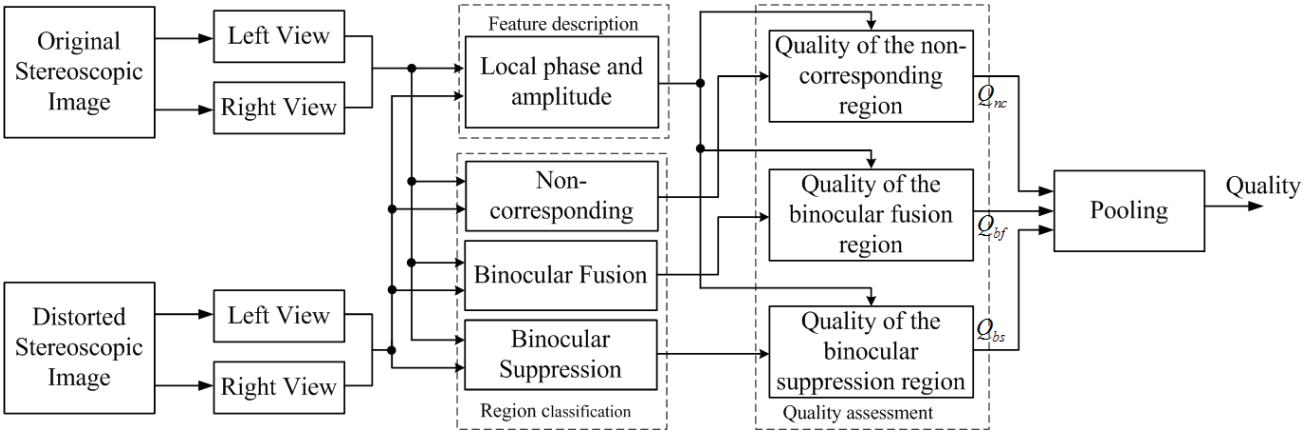


Fig. 4. Procedure of the proposed full-reference stereoscopic image quality assessment metric.

larger than a threshold T_{LR} (we use $T_{LR} = 1$)

$$|d_L(x, y) + d_R(x - d_L(x, y), y)| \geq T_{LR} \tag{8}$$

Furthermore, if the above condition is not satisfied for the matched pixels, we further determine if the matching error is large than a threshold T_{ME}, i.e.,

$$|I_L(x, y) - I_R(x - d_L(x, y), y)| \geq T_{ME} \tag{9}$$

where T_{ME} is adaptively selected according to the BJND threshold [48]. That is, if the inter-difference between two values is lower than the threshold T_{ME}, binocular fusion is still maintained. Finally, the binocular suppression region of the distorted stereoscopic image can be described as

$$R_{bs} = \sup\left(R_{bs}^l, R_{bs}^r\right) \tag{10}$$

where sup( ) denotes the binocular suppression operation, and R_{bs}^l and R_{bs}^r denote the left binocular suppression region and right binocular suppression region, respectively. Note that R_{bs} is not a simple superposition of R_{bs}^l and R_{bs}^r, it involves the complex non-linear neurophysiological process, e.g., formation of the cyclopean image [52]. Therefore, in order to avoid directly evaluating the binocular suppression region, we evaluate the left suppression region and right suppression region independently, and combine them together to generate the quality score for the region. Moreover, the detected binocular suppression regions increase steadily when the stereoscopic image has increasing distortion strength. Consequently, the perceived quality is reduced since binocular suppression usually accompanies with visual discomfort [53].

c) Binocular fusion region detection: After excluding the above detected non-corresponding and binocular suppression regions from the distorted stereoscopic image, binocular fusion region of the distorted stereoscopic image can be described as

$$R_{bf} = \text{fus}\left(R_{bf}^l, R_{bf}^r\right) \tag{11}$$

where fus( ) denotes the binocular fusion operation, and R_{bf}^l and R_{bf}^r denote the left binocular fusion region and right binocular fusion region, respectively. Of course, the proposed region classification method cannot completely coincide with the actual binocular visual response process. This is because, on one hand, the accurate binocular disparity cannot be available from the distorted stereoscopic images, so that certain errors exist in region classification. On the other hand, besides binocular disparity, many other factors, such as edge direction, monocular depth cues, etc., also affects the perceived depth. Existing technologies cannot fully model the binocular visual response process [39]–[41]. In this work, we try to simulate the process of binocular visual response and make the evaluation results more effective.

B. Local Phase and Amplitude Feature Description

As analyzing in Section II, phase is an important feature in quality assessment because much of the position and structure information of an image is captured in phase. However, global phase from the direct Fourier transform is not an effective means in feature description, because different regions are evaluated independently. Obviously, the constant frequency response is not a good choice for feature description. According to the phase congruency theory, the local phase feature information can be extracted by maximizing the Fourier components in phase [54]. Previous researches have shown that simple cells in the primary visual cortex can be well-modeled using log-Gabors [55]. In this paper, we use the method in [56] for phase congruency calculation with the log-Gabor filter. A set of responses on different scales and along different orientations, denoted as $[\eta_{s,o}, \zeta_{s,o}]$, can be obtained by applying the log-Gabor filter $G_{s,o}$(denoted by spatial scale index s and orientation index o) in the Fourier frequency domain:

$$G_{s,o}(r, \theta) = \exp\left[-\frac{(\log(\omega/\omega_s))^2}{2\sigma_s^2}\right] \cdot \exp\left[-\frac{(\theta - \theta_o)^2}{2\sigma_o^2}\right] \tag{12}$$

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. The parameters σ_s and σ_o determine the strength of the filter. The design of the filter is based on the work in [57].

Then, by modulating the scale and orientation, local amplitude at location $\mathbf{x}$ on scale α and along orientation o is given

by

$$A_{s,o}(\mathbf{x}) = \sqrt{\eta_{s,o}(\mathbf{x})^2 + \xi_{s,o}(\mathbf{x})^2} \tag{13}$$

At the same time, the local energy along orientation o is given by

$$E_o(\mathbf{x}) = \sqrt{F_o(\mathbf{x})^2 + H_o(\mathbf{x})^2} \tag{14}$$

where $F_o(\mathbf{x}) = \sum_s \eta_{s,o}(\mathbf{x})$ and $H_o(\mathbf{x}) = \sum_s \zeta_{s,o}(\mathbf{x})$. Then, the phase congruency along orientation o is computed by

$$PC_o(\mathbf{x}) = \frac{E_o(\mathbf{x})}{\varepsilon + \sum_s A_{s,o}(\mathbf{x})} \tag{15}$$

where ε is a small positive constant. In this work, unlike the existing methods [35], [57] that directly use phase congruency for feature description, both local phase and amplitude are adopted. The local phase is defined as the angle of $F_o(\mathbf{x})$ and $H_o(\mathbf{x})$ along the orientation with the maximum phase congruency value:

$$LP(\mathbf{x}) = \arctan\left(H_{o_m}(\mathbf{x}), F_{o_m}(\mathbf{x})\right) \tag{16}$$

where o_m denotes the orientation which corresponds to the maximum phase congruency value. Then, the local amplitude is defined as the sum of the local amplitudes of all the scales along the orientation o_m

$$LA(\mathbf{x}) = \sum\nolimits_s A_{s,o_m}(\mathbf{x}) \tag{17}$$

In the experiment, regarding the parameter selection of log-Gabor filter, $\omega_s = 1/6$, $\theta_o = 0$, $\sigma_s = 0.3$, and $\sigma_o = 0.4$, and the numbers of scale and orientation of the filter are all set to 4.

C. Quality Assessment Metric

After obtaining the above local phase and amplitude features, the next step is to evaluate the image quality by using these features. Let $LP_{l1}(\mathbf{x})$ and $LP_{l2}(\mathbf{x})$ denote the local phases extracted from the original and distorted left images, $LP_{r1}(\mathbf{x})$ and $LP_{r2}(\mathbf{x})$ denote the local phases extracted from the original and distorted right images, $LA_{l1}(\mathbf{x})$ and $LA_{l2}(\mathbf{x})$ denote the local amplitudes extracted from the original and distorted left images, and $LA_{r1}(\mathbf{x})$ and $LA_{r2}(\mathbf{x})$ denote the local amplitudes extracted from the original and distorted right images. According to the definition of SSIM (Structural SIMilarity) [14], the phase and magnitude similarities for each pixel in the left image are defined as

$$S^l_{LP}(\mathbf{x}) = \frac{2LP_{l1}(\mathbf{x}) \cdot LP_{l2}(\mathbf{x}) + C_1}{LP_{l1}(\mathbf{x})^2 + LP_{l2}(\mathbf{x})^2 + C_1} \tag{18}$$

$$S^l_{LA}(\mathbf{x}) = \frac{2LA_{l1}(\mathbf{x}) \cdot LA_{l2}(\mathbf{x}) + C_2}{LA_{l1}(\mathbf{x})^2 + LA_{l2}(\mathbf{x})^2 + C_2} \tag{19}$$

where $C_1$and C_2 are constants to avoid the denominator being zero [14]. Eqs. (18) and (19) are commonly used measures to define the similarity of two positive real numbers and the results range in [0, 1], in which 0 indicates no similarity between two numbers and 1 implies perfect similarity between two numbers. Similarly, $S^r_{LP}(\mathbf{x})$ and $S^r_{LA}(\mathbf{x})$ for each pixel in the right image can be measured.

Then, in order to integrate the effects of the changes in phase and magnitude, we employ linear regression to combine $S^l_{LP}(\mathbf{x})$ and $S^l_{LA}(\mathbf{x})$ into an overall quality score. Although more sophisticated regression models may be employed, we chose linear regression as its lower computational complexity. Assuming that $S_l(\mathbf{x})$ denotes the quality scores for each pixel in the left image, we can express the solution as

$$S_l(\mathbf{x}) = w_{LP} S^l_{LP}(\mathbf{x}) + w_{LA} S^l_{LA}(\mathbf{x}) + b \tag{20}$$

where w_{LP} and w_{LA} are the weights assigned to the phase and amplitude, while b is the intercept (a constant). It is expected that phase has a larger impact on the quality score and so $|w_{LP}| > |w_{LA}|$. The parameters can be determined by the training method in [58] that minimizes the sum of squared errors over the training data. Similarly, the quality score $S_r(\mathbf{x})$ for the right image can be measured by the same manner.

After having obtained the quality scores $S_l(\mathbf{x})$ and $S_r(\mathbf{x})$ at each location respectively, the quality score for each region can be calculated. For the non-corresponding region, the quality score is determined by an average of individual quality scores of each pixel in the left and right non-corresponding regions:

$$Q_{nc} = \frac{\sum_{\mathbf{x}\in R^l_{nc}} S_l(\mathbf{x}) + \sum_{\mathbf{x}\in R^r_{nc}} S_r(\mathbf{x})}{N^l_{nc} + N^r_{nc}} \tag{21}$$

where N^l_{nc} and N^r_{nc} are the numbers of pixels of the left and right non-corresponding regions, respectively.

Based on the analysis in the previous section, 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 the importance of the pixel in the binocular perception is low. A similar idea has been applied to sharpness enhancement of stereoscopic images [59]. Therefore, it would be more effective to impose a higher weight on the perceptually important components, and from this perspective, the quality score for the left binocular suppression region is calculated by

$$Q^l_{bs} = \frac{\sum_{\mathbf{x}\in R^l_{bs}} w_l(\mathbf{x}) \cdot S_l(\mathbf{x})}{\sum_{\mathbf{x}\in R^l_{bs}} w_l(\mathbf{x})} \tag{22}$$

where $w_l(\mathbf{x})$ denotes the weight at location $\mathbf{x}$, and $w_l(\mathbf{x}) = 1/BJND_l(\mathbf{x})$; $BJND_l(\mathbf{x})$ denotes the BJND value of the distorted left image at location $\mathbf{x}$. Similarly, the quality scores Q^l_{bs}, Q^l_{bf} and Q^r_{bf} can be calculated. Since the BJND model in [48] does not consider the accuracy of the disparity information, the original disparity information (estimated from the original stereoscopic images) is used to calculate the BJND in this work.

According to the binocular suppression theory, the visual quality of the binocular suppression region is dominated by the higher quality one. In other words, binocular suppression is equivalent to find the maximum score from the left and right quality scores. Therefore, the quality score for the binocular suppression region is calculated as

$$Q_{bs} = \max(Q^l_{bs}, Q^r_{bs}) \tag{23}$$

Research has shown that binocular summation is more complex than just the sum of two left and right inputs, and the

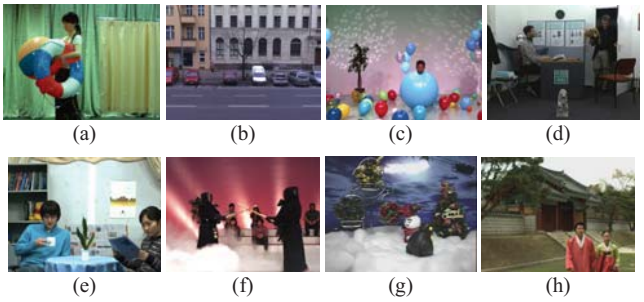


Fig. 5. Eight selected reference left images in the 3D database [61]. (a) *Akko&Kayo*. (b) *Altmoabit*. (c) *Balloons*. (d) *Doorflower*. (e) *Newspaper*. (f) *Kendo*. (g) *Xmas*. (h) *Lovebird1*.

binocular sensitivity of binocular summation is approximately 1.4 times better than monocular sensitivity [60]. In other words, binocular summation is better than monocular viewing by a factor of 1.4 in the process of binocular fusion. Thus, the quality score for the binocular fusion region is calculated by

$$Q_{bf} = 1.4 \cdot \frac{Q_{bf}^{l} + Q_{bf}^{r}}{2} \tag{24}$$

Finally, Q_{nc},Q_{bf} and Q_{bs} are integrated into an overall quality score Q. We use linear weighting to combine the three quality scores because the evaluation for different regions is independent in this work. The overall quality score can be expressed as

$$Q = w_{nc} \cdot Q_{nc} + w_{bf} \cdot Q_{bf} + w_{bs} \cdot Q_{bs} \tag{25}$$

where w_{nc}, w_{bf} and w_{bs} are the weights assigned to different regions, respectively, with the constraint $w_{nc} + w_{bf} + w_{bs} = 1$ to quantify the importance of each region in the binocular vision.

IV. Experimental Results and Analyses

A. Stereoscopic Image Quality Database

We have used the database presented in [61]. The details of the database are described as follows.

a) Source images: Twelve stereoscopic images from MPEG were chosen for our study. The content varies widely and includes indoor and outdoor scenes with a large variety of color, texture, and depth structure. All the images are captured with 50-75mm camera spacing. Eight selected reference left images used in the database are shown in Fig. 5.

b) Participants: Twenty-six non-expert adult viewers with ages range from 20 to 25 were participated in the subjective evaluation of the database. All the participants in the experiment met the minimum visual acuity of 20/30, stereo acuity up to 40 seconds of arc (sec-arc), and passed a color vision test. The participants were asked to rank the stereoscopic images based on their own judgment.

c) Environment: The subjective tests were conducted in the laboratory designed for subjective quality tests according to the recommendation BT.500-11 [62]. The experiment was conducted using a linear polarization stereoscopic display. Stereoscopic images were played back through a duality stereoscopic projection system where two BenQ P8265 DLP projectors were used as front projectors. The duality projection system was used in conjunction with DELL real-time 3D graphics workstations to form a complete stereoscopic production and replay system. The size of stereoscopic projection screens was 150 inches. The viewing distance was 3 times the height of the screen.

TABLE I
Parameters in the Database

Distortion	Control Parameter	Range	Level
JPEG	Quality parameter	[15, 90]	5
JP2K	Bitrate	[0.02, 2]	5
GB	Standard deviation	[0.2, 50]	5
WN	Standard deviation	[0.012, 2.0]	5
H.264	Quantization parameter	[22, 47]	6

d) Stimuli: The symmetric distortions were added on left and right images. The distortions that were simulated inclusive of JPEG, JPEG2000 and H.264 compression, additive Gaussian blur, and White Noise. The degradation of stimuli was varied by control parameters within pre-defined ranges, as reported in Table I. More specifically, there are different distortion levels for each distortion type, and 60, 60, 60, 60 and 72 distorted stereoscopic images in the database with JPEG, JPEG2000, Gaussian Blur, White Noise and H.264 distortions, respectively.

e) Test methodology: A Double Stimulus Continuous Quality Scale (DSCQS) test methodology described in ITU-R recommendation BT.500-11 was used in the experiment. The subjective ratings for the distorted stereoscopic images were obtained on a scale of 0-10. In the data processing stage, after outlier detection and subject rejection using Kurtosis method, the difference mean opinion scores (DMOS) were evaluated using the recommendations adapted from ITU-R BT.500 on a scale of 0-100.

B. Performance Measures and Parameters Determination

Four commonly used performance indicators are employed to evaluate the SIQA metrics: Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC), and root mean squared error (RMSE), between the objective scores after nonlinear regression and the subject scores. Among these four criteria, SROCC and KROCC 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 = SROCC = KROCC = 1 and RMSE = 0. For the nonlinear regression, we use the following five-parameter logistic function [63]:

$$DMOS_p = \beta_1 \cdot \left[\frac{1}{2} - \frac{1}{1 + \exp(\beta_2 \cdot (x - \beta_3))}\right] + \beta_4 \cdot x + \beta_5 \tag{26}$$

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 w_{LP}, w_{LA} and b in Eq. (20), and w_{nc}, w_{bf} and w_{bs} in Eq. (25) by

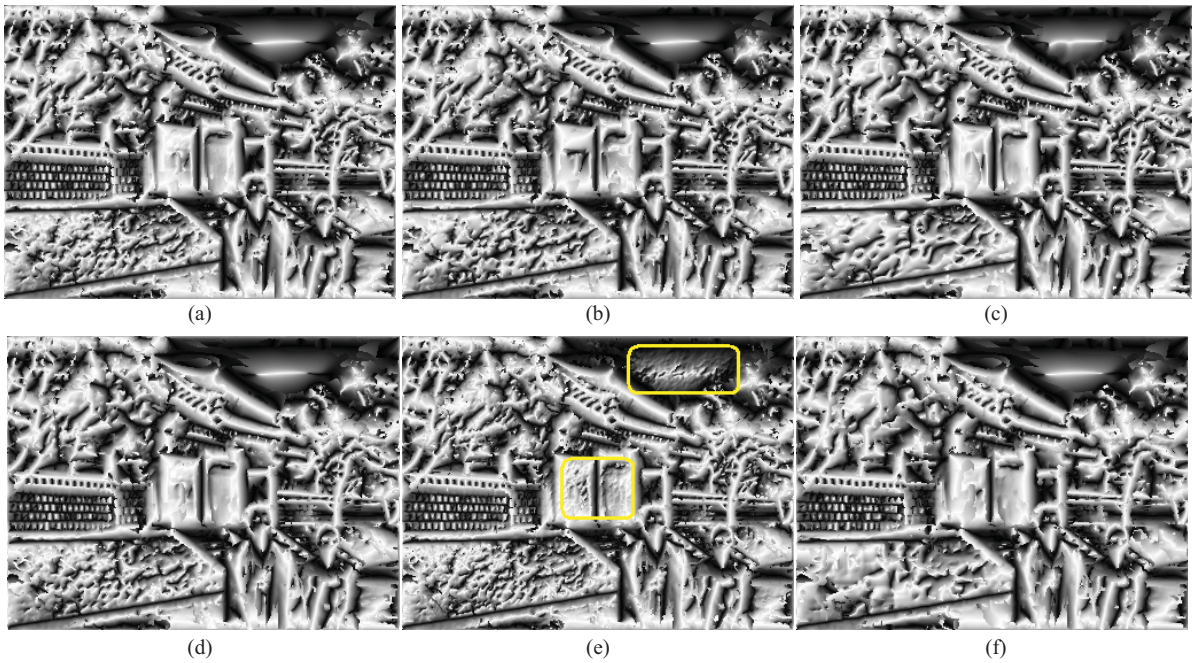


Fig. 6. Local phase maps extracted from the original and distorted images. (a) Local phase maps of the original image in Fig. 1(a). (b)–(f) Local phase maps of the distorted images of JPEG, JPEG 2000, Gaussian Blur, White Noise, and H.264, respectively.

training to optimize the PLCC values between the objective and subjective scores. In the experiments, we select a subset of the database (House, Puppy, and Soccer2 test sequences) to train the parameters. For simplicity, the parameters are chosen by linear regression optimization. Here, two-stage parameter determination is adopted. In the first stage, to determine the parameters w_{LP}, w_{LA} and b, the quality score is evaluated for the whole stereoscopic image (the left and right images are evaluated separately, and weight-averaged to obtain the quality score), so that the influence of w_{oc}, w_{bf} and w_{bs} can be omitted. The parameter determination results are $w_{LP} = 0.9834$, $w_{LA} = 0.2915$ and $b = 0$. As expected, the phase component is more important than the magnitude component (i.e., $w_{LP} > w_{LA}$). In the second stage, with the determined parameters w_{LP}, $w_{LA}$and b, the further parameter determination results are $w_{nc} = 0$, $w_{bf} = 0.5445$ and $w_{bs} = 0.4555$. As expected, binocular fusion and binocular suppression dominate the binocular vision. In the following experiments, the proposed metric is tested on the remaining test sequences in the database (in this way, we avoid same sequences for training and testing). Thus, totally 234 distorted stereoscopic images are adopted in the evaluation.

In order to evaluate the performance of the proposed scheme, we compare with the existing schemes, including three 2D-IQA schemes, i.e., signal-to-noise ratio (PSNR), multi-scale SSIM (MS-SSIM) [64], VIF [15], and four SIQA schemes, i.e., You's scheme [24], Benoit's scheme [25], Qi's scheme [35] and Wang's scheme [40]. The former three schemes directly estimate the quality of each view separately and generate a weighted average score. For You's scheme, we adopt the best combination approach, in which the image quality metric (IQM) is the average result of the left and right images using SSIM, and depth quality metric (DQM) is the quality result of the disparity map using universal quality index (UQI) [65]. 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. Since both algorithms require estimated disparity maps from both the original and distorted stereoscopic images, we use the same stereo matching algorithm [51] to create disparity maps.

C. Feature Extraction Performance

To facilitate display for the figures that follow (Figs. 6 and 7), the local phase and amplitude values are mapped to [0, 255]. In order to show different characteristics of distortion types, the distorted images with the similar DMOS values are selected (cases with exactly the same DMOS values do not exist in the database). The DMOS values used in this regard are 25.652, 20.957, 20.217, 23.435 and 21.783 for the distortions of JPEG, JPEG2000, Gaussian Blur, White Noise and H.264, respectively. Fig. 6(a)–(f) show the local phase maps of the original image in Fig. 1(a) and its distorted images, respectively. Fig. 7(a)–(f) show the local amplitude maps of the corresponding local phase maps in Fig. 6(a)–(f), respectively. We can see that the local phase maps in Fig. 6(b), (c), (d) and (f) are more similar to the map in Fig. 6(a) than the maps in Fig. 6(e). In order to facilitate visual examination, regions with obvious differences in Fig. 6(e) are marked by rectangle boxes. Besides, from the visual examination, the local phase map is a good description of local features because important structure information is preserved. We find that the local amplitude maps in Fig. 7(b), (c), (d) and (f) are also more similar to the map in Fig. 7(a) than the maps in Fig. 7(e), as denoted by the error images marked by red rectangles. This indicates that the feature

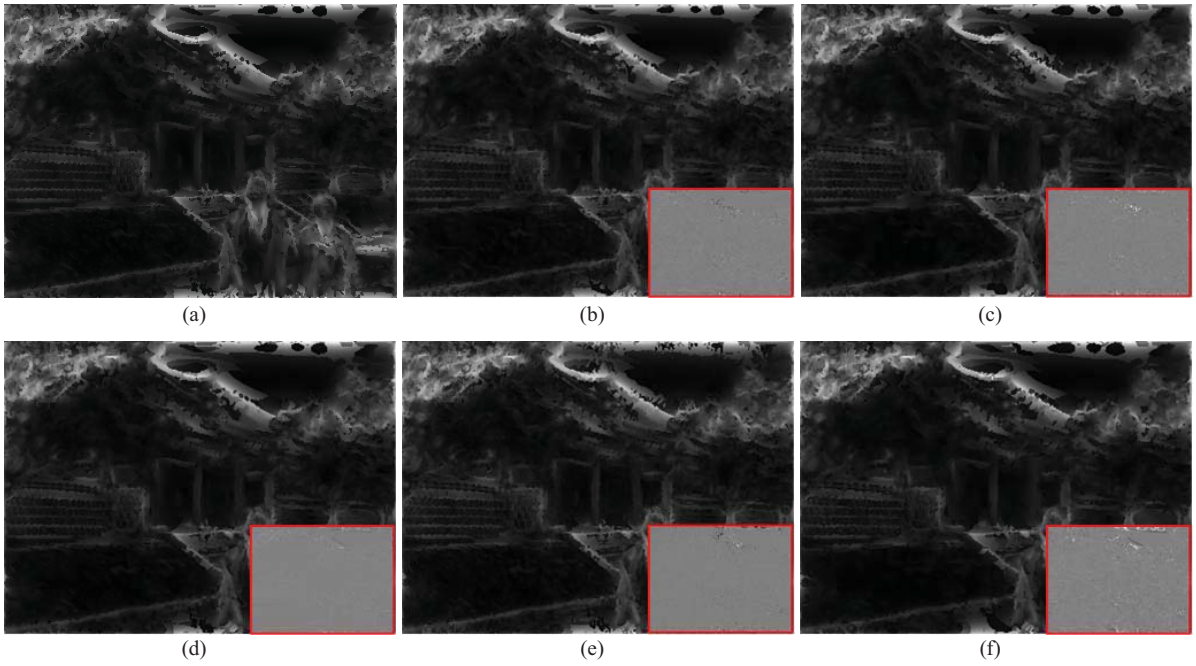


Fig. 7. Local amplitude maps extracted from the original and distorted images. (a) Local amplitude maps of the original image in Fig. 1(a). (b)–(f) Local amplitude maps of the distorted images of JPEG, JPEG 2000, Gaussian Blur, White Noise, and H.264, respectively.

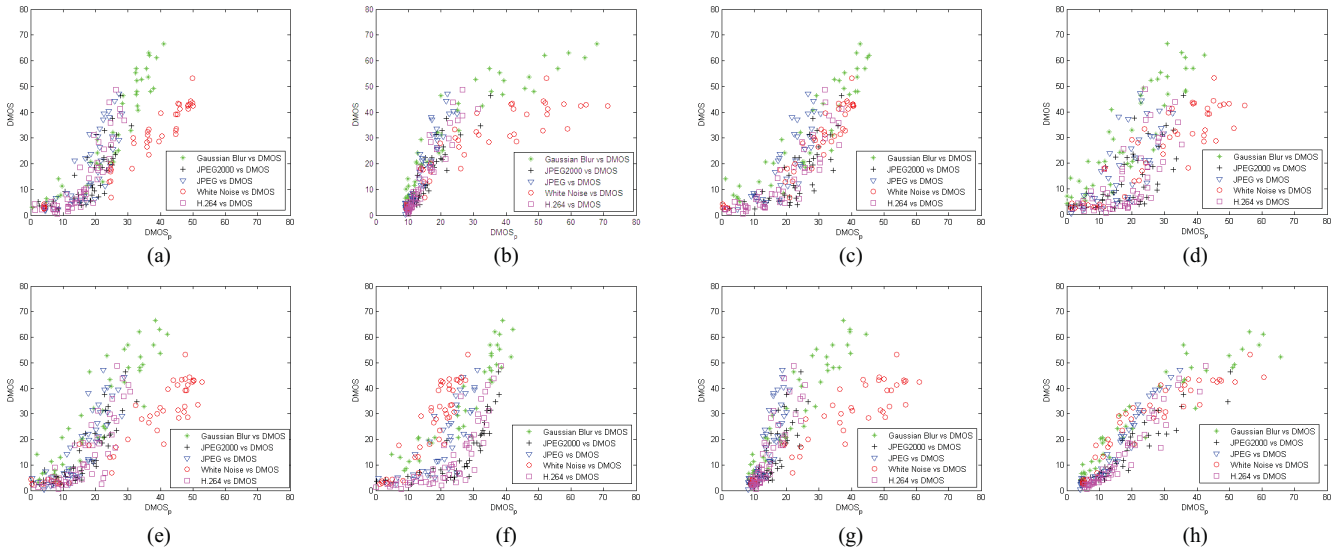


Fig. 8. Scatter plots of objective scores versus subjective scores for the eight schemes. (a) PSNR. (b) MS-SSIM. (c) VIF. (d) Scheme in [24]. (e) Scheme in [25]. (f) Scheme in [35]. (g) Scheme in [40]. (h) Proposed scheme.

extraction method may be not effective for White Noise. Overall, by properly making use of local phase and amplitude maps, we can estimate the quality degradation consistently with human visual perception. Most results will be presented in the next subsections.

D. Overall Assessment Performance

Fig. 8 shows the scatter plots for the eight schemes under comparison. The vertical axis denotes the subjective ratings of the perceived distortions and the horizontal axis denotes the predicted DMOS values. The values of PLCC, SROCC, KROCC and RMSE of each distortion type with the database are listed in Table II. From the table, we can see that the proposed scheme outperforms the 2D-IQA and SIQA schemes. For PSNR, MS-SSIM and VIF schemes, since they are directly extended from the 2D case and do not take the binocular visual characteristics into account, the overall performance is far worse than the proposed scheme in general, even though they may be effective for some special distortion types. For example, similar to the 2D case, PSNR is a useful measure for additive white noise [66]. The proposed scheme is not very capable in dealing with the distortions of "noise". For You's and Benoit's schemes, they are the combination of 2D

TABLE II
PERFORMANCE COMPARISON OF THE EIGHT SCHEMES (CASES IN BOLD DENOTE BEST PERFORMANCE)

	Criteria	PSNR	MS-SSIM	VIF	You [24]	Benoit [25]	Qi [35]	Wang [40]	Proposed
JPEG	PLCC	0.7353	0.9413	0.7951	0.7522	0.8370	0.8010	0.8566	**0.9737**
	SROCC	0.8666	0.9410	0.8874	0.8087	0.8824	0.9042	0.8799	**0.9819**
	KROCC	0.6781	0.7610	0.6781	0.5953	0.6882	0.7489	0.6781	**0.9005**
	RMSE	9.4061	4.6864	8.4172	9.1457	7.5954	8.3082	7.1617	**3.1624**
JPEG2000	PLCC	0.7054	0.9111	0.7379	0.6623	0.8243	0.7561	0.8158	**0.9419**
	SROCC	0.8760	0.9227	0.8518	0.6876	0.8779	0.9260	0.8619	**0.9557**
	KROCC	0.6970	0.7576	0.6525	0.5030	0.7374	0.7879	0.6667	**0.8222**
	RMSE	7.9167	4.6045	7.5383	8.3686	6.3244	7.3097	6.4591	**3.7503**
Gaussian Blur	PLCC	0.8879	0.9177	0.9124	0.9054	0.9149	0.9383	**0.9475**	0.9275
	SROCC	0.9445	**0.9683**	0.9444	0.9112	0.9153	0.9598	0.9485	0.9579
	KROCC	0.8014	**0.8540**	0.7873	0.7408	0.7549	0.8338	0.7913	0.8277
	RMSE	9.1153	7.8720	8.1095	8.4110	7.9981	6.8500	**6.3356**	7.4056
White Noise	PLCC	**0.9570**	0.8846	0.9362	0.8830	0.9197	0.9040	0.9234	0.9211
	SROCC	**0.9588**	0.9003	0.9378	0.8561	0.8791	0.8813	0.8700	0.9568
	KROCC	**0.8202**	0.7152	0.7778	0.6768	0.7030	0.7192	0.6848	**0.8242**
	RMSE	**4.3459**	6.9857	5.2643	7.0337	5.8827	6.4067	5.7499	5.8347
H.264	PLCC	0.8394	0.9454	0.8575	0.7062	0.8038	0.7868	0.8908	**0.9593**
	SROCC	0.8871	0.9357	0.8957	0.7279	0.8227	0.8955	0.8862	**0.9672**
	KROCC	0.7149	0.7849	0.7051	0.5373	0.6814	0.7317	0.7079	**0.8520**
	RMSE	6.9398	4.1622	6.5685	9.0409	7.5966	7.8802	5.8022	**3.6043**
All	PLCC	0.8132	0.8320	0.8464	0.7435	0.7936	0.7192	0.7578	**0.9168**
	SROCC	0.8914	0.9245	0.8982	0.7762	0.8529	0.7572	0.8740	**0.9324**
	KROCC	0.6950	0.7491	0.7103	0.5668	0.6565	0.5612	0.6711	**0.7684**
	RMSE	9.4784	9.0349	8.6735	10.8905	9.9091	11.3155	10.6256	**6.5031**

image quality metrics for stereoscopic images and disparity maps. The performance of the two schemes is lower than the 2D-IQA schemes. The reason is that the quality of the estimated disparity is highly dependent on the stereo matching algorithms, and the 2D image quality metric for disparity maps does not coincide with human perception of disparity. This is another demonstration of the importance of binocular region classification in the proposed scheme. For Wang's schemes, even though it may be effective for some individual distortion types, e.g., Gaussian Blur, White Noise, the overall assessment performance is not very good. The reason is that uniform assessment is adopted for left and right images of different distortion types, and thus, poor convergence across different distortion types is occurred, as observed from the scatter plots in Fig. 8. For Qi's scheme, since it is designed to assess the perceived depth (camera distances for stimuli), the generated local quality map cannot reflect the changes of image quality for some compression-induced distortions, e.g., JPEG, JPEG2000 and H.264, but is particularly effective for Gaussian blur because blur affects the perceived depth [67]. Overall, the proposed scheme can achieve the best assessment performance against existing 2D-IQA and SIQA schemes involving all the distortion types.

E. Impact of Each Components in the Proposed Scheme

To demonstrate the impact of each component in the proposed scheme, we design four different schemes for comparison, denoted by Scheme-A, Scheme-B, Scheme-C and Scheme-D, respectively. For Scheme-A, the BJND modulation in Eq. (22) is not included, and other operations are same with the proposed scheme. For Scheme-B, only the local phase feature is used, and other operations are same with the proposed scheme. For Scheme-C, only the local amplitude feature is used, and other operations are same with the proposed scheme. For Scheme-D, the quality scores for the left and right images are directly weight-averaged to obtain the final quality score. The results of PLCC and SROCC are presented in Table III and Table IV. From the tables, we can see that proposed BJND modulation component has an important impact on the performance improvement for most of the distortion types. Besides, by comparing the evaluation results of Scheme-B (local phase only) and Scheme-C (local amplitude only), the phase component has more important contribution than the amplitude component, and

TABLE III
PLCC COMPARISONS FOR EACH COMPONENT OF THE PROPOSED SCHEME

	Scheme-A	Scheme-B	Scheme-C	Scheme-D	Proposed
JPEG	0.9653	0.9685	0.9261	0.9644	**0.9737**
JP2K	0.9473	**0.9506**	0.8701	0.9452	0.9419
GB	0.9123	0.9189	0.9160	0.9121	**0.9275**
WN	**0.9237**	0.9227	0.8678	0.9225	0.9211
H.264	0.9553	**0.9628**	0.9029	0.9519	0.9593
All	0.9004	0.9126	0.8843	0.8933	**0.9168**

TABLE IV
SROCC COMPARISONS FOR EACH COMPONENT OF THE PROPOSED SCHEME

	Scheme-A	Scheme-B	Scheme-C	Scheme-D	Proposed
JPEG	0.9790	**0.9847**	0.9321	0.9792	0.9819
JP2K	0.9544	**0.9599**	0.9043	0.9485	0.9557
GB	**0.9601**	0.9553	0.9560	0.9611	0.9579
WN	0.9498	**0.9680**	0.9141	0.9540	0.9568
H.264	0.9531	0.9626	0.9358	0.9572	**0.9672**
All	0.9063	0.9272	0.9142	0.8955	**0.9324**

TABLE V
CROSS-DISTORTION PERFORMANCE EVALUATION OF THE PROPOSED SCHEME

	JPEG	JP2K	GB	WN	H.264
JPEG	**0.9741**	0.9422	0.9240	0.9214	0.9577
JP2K	0.9729	**0.9425**	0.9190	0.9211	0.9569
GB	0.9736	0.9420	**0.9275**	0.9210	0.9594
WN	0.9740	0.9423	0.9223	**0.9214**	0.9570
H.264	0.9716	0.9418	0.9264	0.9204	**0.9598**

the overall evaluation performance can be further improved by properly combining the phase and amplitude components, even though Scheme-B has better performance for some distortion types. For Scheme-D, the evaluation performance is obviously lower than the proposed full-fledged scheme, and this further indicates the improvement of the feature description and binocular visual processes modeled in the proposed scheme.

F. Cross-Distortion Performance Evaluation

The effectiveness of the scheme has been investigated when applying the same training parameters for different distortion types. Table V shows the experimental results of cross-distortion performance evaluation of the proposed scheme. In the table, only the PLCC values are listed (to save space), because the other measures give similar results. Each column indicates the results for a same test set, while each row indicates the results for a same training set. As illustrated in subsection 4.2, the parameters w_{LP}, w_{LA} and b in Eq. (20), and the parameters w_{nc}, w_{bf} and w_{bs} in Eq. (25) need to be determined by training. It can be seen from the table that the diagonal elements yield the best performance because the same training and test sets are used. What is more important is that, no matter which the training set is, the test performance for different distortions is always good. By observing the experimental results in Table II and Table V, the parameter determination method of the proposed scheme tends to be effective across different distortion types. This demonstrates that the proposed scheme is a good metric for predicting the perceived quality of stereoscopic images.

V. CONCLUSION

This paper has presented a perceptual full-reference quality assessment method of stereoscopic images by considering the related binocular visual characteristics. Compared with the existing two-dimensional (2D) metrics, the theoretical significance of the proposed method is that we try to quantify the stereoscopic image quality in a manner that conforms to human binocular vision. To be more specific, a stereoscopic image is separated into different binocular regions, each region is evaluated independently by considering their visual properties, and all effects are finally integrated into an overall quality score. The advantages of the proposed method are as follows: 1) the local phase and amplitude allow better representation of image features, 2) by considering the binocular combination property, the binocular just noticeable difference (BJND) model is used to modulate the quality score, 3) by considering binocular perception property, different binocular regions are evaluated independently. It can be observed from the experimental results that the proposed method can achieve much higher consistency with the subjective assessments.

The important contribution of this work is the demonstration that accounting for binocular visual characteristics can greatly improve the performance of three-dimensional (3D) metrics. The proposed metric can be used to evaluate the quality of stereoscopic images that are distorted in various processes. In the future work, to further advance the performance of the proposed metric, the following issues should be addressed: 1) the effect of disparity/depth masking needs to be further addressed in evaluating the perceived depth, and besides binocular disparity, many other factors, such as visual direction, monocular depth cues, etc., should be addressed in modeling the sensation of depth; 2) we should investigate the performance of the proposed metric by considering various perceptual scales for image quality, depth perception and visual comfort; 3) a more comprehensive database should be constructed to include various types of 3D distortion, e.g., asymmetric distortion, spatial resolution reduction, vertical disparities, color changes, and camera distances.

REFERENCES

[1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, "Three-dimensional video postproduction and processing," *Proc. IEEE*, vol. 99, no. 4, pp. 607–625, Apr. 2011.

[2] A. A. Alatan, Y. Yemez, U. Güdükbay, X. Zabulis, K. Müller, C. E. Erdem, C. Weigel, and A. Smolic, "Scene representation technologies for 3DTV—a survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1587–1605, Nov. 2007.

[3] C. Wheatstone, "Contributions to the physiology of vision—part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision," *Phil. Trans. Royal Soc. London*, vol. 128, no. 1838, pp. 371–394, 1983.

[4] S. E. Palmer, *Vision Science: Photons to Phenomenology*. Cambridge, MA, USA: MIT Press, 1999.

[5] I. P. Howard and B. J. Rogers, *Seeing in Depth: Depth Perception*, vol. 2. Toronto, ON, Canada: I Porteous, 2002.

[6] S. Winkler and D. Min, "Stereoscopic image quality compendium," in *Proc. Int. Conf. Inf., Commun. Signal Process.*, Dec. 2011, pp. 1–5.

[7] J. S. Lee, L. Goldmann, and T. Ebrahimi, "Paired comparison-based subjective quality assessment of stereoscopic images," in *Multimedia Tools Applications*. New York, USA: Springer-Verlag, Feb. 2012, pp. 1–18.

[8] W. A. Ijsselsteijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 225–233, Mar. 2000.

[9] W. J. Tam, L. B. Stelmach, and S. Subramaniam, "Stereoscopic video: Asymmetrical coding with temporal interleaving," *Proc. SPIE*, vol. 4297, pp. 299–306, Jun. 2001.

[10] X. Wang, M. Yu, Y. Yang, and G. Y. Jiang, "Research on subjective stereoscopic image quality assessment," *Proc. SPIE*, vol. 7255, p. 725509, Jan. 2009.

[11] A. Kuijsters, W. A. Ijsselsteijn, M. T. M. Lambooij, and I. Heynderickx, "Influence of chroma variations on naturalness and image quality of stereoscopic images," *Proc. SPIE*, vol. 7240, p. 72401E, Jan. 2009.

[12] M. T. Pourazad, Z. Mai, P. Nasiopoulos, K. Plataniotis, and R. K. Ward, "Effect of brightness on the quality of visual 3D perception," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 989–992.

[13] W. Lin and C. C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, May 2011.

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

[15] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[16] M. G. Martini, B. Villarini, and F. Fiorucci, "A reduced-reference perceptual image and video quality metric based on edge preservation," *Eur. Associat. Signal Image Process. J. Adv. Signal Process.*, vol. 66, no. 2012, pp. 1–13, Mar. 2012.

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

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

[19] M. Lambooij, M. Fortuin, W. A. Ijsselsteijn, B. J. Evans, and I. Heynderickx, "Susceptibility to visual discomfort of 3-D displays by visual performance measures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 12, pp. 1913–1923, Dec. 2011.

[20] W. Chen, J. Fournier, M. Barkowsky, P. Le Callet, "Quality of experience model for 3DTV," *Proc. SPIE*, vol. 8288, Feb. 2012.

[21] H. T. Quan, P. Le Callet, and M. Barkowsky, "Video quality assessment: From 2D to 3D—challenges and future trends," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4025–4028.

[22] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969–1976, Nov. 2008.

[23] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G. B. Akar, "Toward compound stereo-video quality metric: A specific encoder-based framework," in *Proc. IEEE Southwest Symp. Image Anal. Interpret.*, Jun. 2006, pp. 218–222.

[24] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, Scottsdale, AZ, USA, Sep. 2010, pp. 4033–4036.

[25] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 389–392.

[26] K. Ha and M. Kim, "A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2525–2528.

[27] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, S. Villette, and A. M. Knodoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 304–318, Apr. 2009.

[28] E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Toward a new quality metric for 3-D synthesized view assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1332–1343, Nov. 2011.

[29] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H. P. Seidel, "A perceptual model for disparity," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–8, 2011.

[30] A. Maalouf and M. C. Larabi, "CYCLOP: A stereo color image quality assessment metric," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 1161–1164.

[31] L. Jin, A. Boev, A. Gotchev, and K. Egiazarian, "3D-DCT based perceptual quality assessment of stereo video," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2521–2524.

[32] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE*, vol. 6803, p. 680305, Apr. 2008.

[33] Z. M. P. Sazzad, S. Yamanaka, Y. Kawayoke, and Y. Horita, "Stereoscopic image quality prediction," in *Proc. Int. Conf. Quality Multimedia Exper.*, Jul. 2009, pp. 180–185.

[34] J. J. Hwang and H. R. Wu, "Stereo image quality assessment using visual attention and distortion predictors," *KSII Trans. Internet Inf. Syst.*, vol. 5, no. 9, pp. 1613–1631, Sep. 2011.

[35] F. Qi, D. Zhao, T. Jiang, and S. Ma, "Quality of experience assessment for stereoscopic images," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 1712–1715.

[36] H. Ono, R. Angus, and P. Gregor, "Binocular single vision achieved by fusion and suppression," *Percept. Psychophys.*, vol. 21, no. 6, pp. 513–521, 1977.

[37] G. Saygili, C. G. Guŕler, and A. M. Tekalp, "Quality assessment of asymmetric stereo video coding," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4009–4012.

[38] F. Shao, G. Jiang, M. Yu, K. Chen, Y. S. Ho, "Asymmetric coding of multi-view video plus depth based 3D video for view rendering," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 157–167, Feb. 2012.

[39] A. Mittal, A. K. Moorthy, J. Ghosh, and A. C. Bovik, "Algorithmic assessment of 3D quality of experience for images and videos," in *Proc. IEEE Digit. Signal Process. Workshop*, Jan. 2011, pp. 338–343.

[40] X. Wang, S. Kwong, and Y. Zhang, "Considering binocular spatial sensitivity in stereoscopic image quality assessment," in *Proc. IEEE Visual Commun. Image Process.*, Tainan, Taiwan, Nov. 2011, pp. 1–4.

[41] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimen. Syst. Signal Process.*, doi: 10.1007/s11045-012-0178-3.

[42] O. J. Braddick, "Binocular single vision and perceptual processing," *Proc. Royal Soc. London. Ser. B, Biol. Sci.*, vol. 204, no. 1157, pp. 503–512, Jun. 1979.

[43] D. J. Fleet, H. Wagner, and D. J. Heeger, "Neural encoding of binocular disparity: Energy models, position shifts and phase shifts," *Vis. Res.*, vol. 36, no. 12, pp. 1839–1857, Jun. 1996.

[44] M. C. Morrone and D. C. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proc. Royal Soc. London. Ser. B, Biol. Sci.*, vol. 235, no. 1280, Jun. 1988, pp. 221–245.

[45] M. G. A. Thomson, D. H. Foster, and R. J. Summers, "Human sensitivity to phase perturbations in natural images: A statistical framework," *Perception*, vol. 29, no. 9, pp. 1057–1070, 2000.

[46] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.

[47] X. Yang, W. Lin, Z. Lu, X. Lin, S. Rahardja, E. Ong, and S. Yao, "Rate control for videophone using local perceptual cues," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 4, pp. 496–507, Apr. 2005.

[48] Y. Zhao, Z. Chen, C. Zhu, Y. P. Tan, and L. Yu, "Binocular JND model for stereoscopic images," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 19–22, Jan. 2011.

[49] I. P. Howard and B. J. Rogers, *Binocular Fusion and Rivalry in Binocular Vision and Stereopsis*, New York, USA: Oxford Univ. Press, 1995.

[50] S. B. Steinman, B. A. Steinman, and R. P. Garzia, *Foundations of Binocular Vision: A Clinical Perspective*, New York, USA: McGraw-Hill, 2000.

[51] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. IEEE Int. Comput. Vis.*, vol. 2. Vancouver, BC, Canada, Jul. 2001, pp. 508–515.

[52] B. Julesz, *Foundations of Cyclopean Perception*, Chicago, IL, USA: Univ. Chicago Press, 1971.

[53] C. Richardt, L. Świrski, I. P. Davies, and N. A. Dodgson, "Predicting stereoscopic viewing comfort using a coherence-based computational model," in *Proc. Int. Symp. Comput. Aesthet. Graph., Visualizat., Imag.*, Vancouver, BC, Canada, Aug. 2011, pp. 97–104.

[54] N. Qian and S. Mikaelian, "Relationship between phase and energy methods for disparity computation," *Neural Comput.*, vol. 12, no. 2, pp. 279–292, 2000.

[55] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[56] P. Kovesi, "Image features from phase congruency," *Videre: J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 1–26, 1999.

[57] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[58] M. Narwaris, W. Lin, I. V. McLoughlin, S. Emmanuel, and L. T. Chia, "Fourier transform based scalable image quality measure," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3364–3377, May 2012.

[59] S. W. Jung, J. Y. Jeong, and S. J. Ko, "Sharpness enhancement of stereo images using binocular just-noticeable difference," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1191–1199, Mar. 2012.

[60] J. M. Nelson-Quigg, K. Cello, and C. A. Johnson, "Predicting binocular visual field sensitivity from monocular visual field results," *Investigat. Ophthalmol. Visual Sci.*, vol. 41, no. 8, pp. 2212–2221, 2000.
[61] J. Zhou, G. Jiang, X. Mao, M. Yu, F. Shao, Z. Peng, and Y. Zhang, "Subjective quality analyses of stereoscopic images in 3DTV system," in *Proc. IEEE Visual Commun. Image Process.*, Tainan, Taiwan, Nov. 2011, pp. 1–4.
[62] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Standard BT.500-11, 2002.
[63] P. G. Gottschalk and J. R. Dunn, "The five-parameter logistic: A characterization and comparison with the four-parameter logistic," *Anal. Biochem.*, vol. 343, no. 1, pp. 54–65, Aug. 2005.
[64] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.
[65] Z. Zhou and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[66] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[67] J. Wang, M. Barkowsky, V. Ricordel, and P. Le Callet, "Quantifying how the combination of blur and disparity affects the perceived depth," *Proc. SPIE*, vol. 7865, pp. 78650K-1–78650K-10, Feb. 2011.

Feng Shao received the 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 with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. He was a Visiting Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore, from February 2012 to August 2012. His current research interests include 3-D video coding, 3-D quality assessment, and image perception.

Weisi Lin (M'92–SM'98) received the B.Sc. and M.Sc. degrees from Zhongshan University, Guangzhou, China, and the Ph.D. degree from the King's College, London University, London, U.K.

He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He was the Head of the Visual Processing Laboratory and the Acting Manager of the Department of Media Processing, Institute for Infocomm Research, Singapore. He has authored or co-authored more than 200 refereed papers in international journals and conferences. His current research interests include expertise include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.

Dr. Lin is on the editorial boards of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*. He was the Lead Guest Editor of a special issue on Perceptual Signal Processing, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, in 2012. He is the Chair of the IEEE MMTC Special Interest Group on Quality of Experience. He was elected as a Distinguished Lecturer of APSIPA for 2012–2013. He is the Lead Technical Program Chair of Pacific-Rim Conference on Multimedia 2012, and the Technical Program Chair of the IEEE International Conference on Multimedia and Expo 2013. He is a Chartered Engineer in U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.

Shanbo Gu received the B.S. degree in electronic and information engineering from Hangzhou Dianzi University, Hangzhou, China, in 2010. She is currently pursuing the M.S. degree with Ningbo University, Ningbo, China.

Her current research interests include image and video processing, and quality assessment.

Gangyi Jiang received the M.S. degree from Hangzhou University, Hangzhou, China, and the Ph.D. degree from Ajou University, Gyeonggi-do, Korea, in 1992 and 2000, respectively.

He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. His current research interests include digital video compression and multiview video coding.

Thambipillai Srikanthan (SM'92) received the B.Sc. degree (Hons) in computer and control systems and the Ph.D. degree in system modeling and information systems engineering from CNAA, Milton Keynes,U.K.

He joined Nanyang Technological University (NTU),Singapore, in 1991, where, in February 2010, he was invited to be the Head of the School of Computer Engineering (SCE) which has about 85 faculty, 1800 students (UG and M.Sc.), 380 Ph.D. students, 74 non-faculty and 175 research staff. Beginning in 2005, he served as the founding Director of the Intelligent Devices and Systems Cluster (IDeAS), NTU for two years. He remains the Director of a 100-strong Centre for High Performance Embedded Systems (CHiPES), NTU, which he founded in 1998. Bearing his name are three invention disclosures and more than 340 technical papers including 83 journals in IEEE TRANSACTIONS, IEE proceedings and other reputed international journals. He has successfully supervised to completion more than 25 graduate students and he is the main supervisor to more than 10 Ph.D. students at present. His current research interests include design methodologies for high-productivity embedded systems, architectural translations of compute-intensive algorithms, computer arithmetic and high-speed techniques for vision- enabled systems.

Prof. Srikanthan was the recipient of the Public Service Public Administration Medal (Bronze) on National Day for outstanding contributions to education and research in Singapore in 2006. He is the invited fellow of Singapore Academy of Engineering.