Signal Processing 145 (2018) 202–213

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Quaternion representation based visual saliency for stereoscopic image quality assessment

Xu Wang[a], Lin Ma[b,*], Sam Kwong[c,d], Yu Zhou[a]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[b] Tencent AI Lab, Shenzhen 518060, China
[c] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong
[d] Shenzhen Research Institute, City University of Hong Kong, Shenzhen 5180057, China

ARTICLE INFO

Article history:
Received 15 July 2017
Revised 3 November 2017
Accepted 1 December 2017
Available online 8 December 2017

Keywords:
Stereoscopic image quality assessment (SIQA)
Visual saliency
Quaternion representation (QR)
Human visual system (HVS)

ABSTRACT

In this paper, a novel visual saliency detection method for stereoscopic images is proposed for the stereoscopic image quality assessment (SIQA) by considering the disparity map and difference image between the stereo image pairs. Firstly, a new quaternion representation (QR) of each stereo image (left/right view image) is constructed, which comprises the image content, the inter-view disparity, and the difference map. The quaternion Fourier transform (QFT) is performed on the constructed QR to generate the visual saliency maps for left and right views of stereoscopic image pairs, respectively. The generated visual saliency maps are further incorporated into the quality metrics for SIQA. Experimental results demonstrate that the visual saliency maps generated by the proposed method can help significantly boost the performance of SIQA, comparing with other visual saliency models proposed for stereoscopic images. It further confirms that the proposed visual saliency model can accurate depict the acuity property of human visual system (HVS) in judging the perceptual quality of stereoscopic images.

© 2017 Published by Elsevier B.V.

1. Introduction

With the rapid development of content generation and display technology, three-dimensional (3D) applications and services are becoming more and more popular for visual quality of experiences (QoE) of human viewers. The 3D contents displaying on the 3D devices, such as the 3D films and video games, have now brought vivid experiences to the consumers, which have attracted attentions from not only researchers but also the industries. For these applications, the quality of 3D content [1–4] is the most critical part to guarantee the visual QoE. However, in the 3D processing chain including capturing, processing, coding, transmitting, reconstruction, retrieving, etc., artifacts are inevitably introduced due to the resource shortage in processing [5,6]. Therefore, how to evaluate the perceptual quality of 3D content becomes a challenging issue in 3D visual signal processing, which can automatically evaluate, control, and optimize the perceptual quality of 3D contents during each processing stage. Then the best visual QoE can thus be provided to the consumers.

As human eyes are the ultimate receivers of the 2D/3D images, the properties of HVS are considered to develop an effective perceptual IQA metric [7]. For example, the widely known just-noticeable difference (JND) models [8,9], which employ the contrast sensitivity function (CSF), luminance masking, and contrast masking properties of the HVS, have demonstrated good performances for perceptual image/video quality assessment. Moreover, the horizontal effect property of HVS has been modeled in [10], which demonstrate that the HVS orientation preference can help improve the performance of IQA metrics. Among the HVS properties, visual saliency [11–18] is the most straightforward HVS characteristic for visual information processing. Visual saliency would selectively process the important part and ignore the unimportant part of the visual information. For quality assessment, the distortions presented in the salient regions would draw more attentions from the human viewers. In other words, the perceptual quality of the salient region tends to represent the perceptual quality of the whole image. Thus, it should be helpful to incorporate the saliency map into quality metrics. Over the past decades, many computational models [11,14–18] for visual saliency detection have been proposed. Itti et al. proposed a bottom-up model based on the neuronal architecture of the primates's early visual system [14]. The saliency map is derived from the color, intensity, and orientation features. Harel et al. [15] employed the graph-

* Corresponding author.
E-mail addresses: wangxu@szu.edu.cn (X. Wang), lma@ee.cuhk.edu.hk (L. Ma), cssamk@cityu.edu.hk (S. Kwong), yu.zhou@szu.edu.cn (Y. Zhou).

https://doi.org/10.1016/j.sigpro.2017.12.002
0165-1684/© 2017 Published by Elsevier B.V.

based theory to measure the saliency from the feature contrast. In [16], a saliency detection algorithm based on information maximization is proposed. Hou et al. proposed a spectral residual (SR) approach [17], where the saliency map is computed by log spectra representation of image in Fourier transform domain. Guo et al. [18] detected the video saliency map by considering the phrase residual (PR) features. For the application of image retargeting, Fang et al. [11] developed a saliency model in compression domain.

For 3D images, it is claimed that the artifacts of 3D content affect more on the perceptual quality [19,20], compared with the conventional 2D contents. Therefore, the modeling of visual saliency properties on 3D content is expected to more accurately evaluate perceptual quality of the 3D contents. Nowadays, the visual saliency models for 3D images have been researched by incorporating the depth cues from the 3D images. However, most of existing saliency models mainly focus on simulating the behavior of human eye fixation of the image, which may not be suitable for depicting the HVS property for quality perception. Thus, the saliency map derived by these visual saliency models may not be helpful for stereoscopic image quality assessment (SIQA), as demonstrated in Section 5. In order to handle the limitations of these saliency maps, we propose a novel saliency map model targeting at SIQA. By incorporating the saliency map, the performances of the quality metrics can be significantly improved. Compared with the state of the art works, our contributions are listed as follows:

- A visual saliency model for stereoscopic image targeting at SIQA is developed. Compared with existing saliency models, which explicitly require the depth image, the proposed saliency model takes the stereoscopic image as the input to generate a better saliency map for SIQA.
- The stereoscopic image is represented as a new quaternion representation (QR), which considers the spatial image content and the inter-view relationships, specifically the disparity map and the difference image between left and right views. With such considerations, the depth cues are implicitly considered for saliency map generation. And the experimental results demonstrated that the new QR is very effective for saliency map generation, especially for SIQA.

The rest of this paper is organized as follows. Section 2 overviews the related works. In Section 3, the proposed visual saliency model for stereoscopic image is introduced. Section 4 illustrates the incorporation of visual saliency map into the quality metrics for stereoscopic images. Experimental results are provided in Section 5. Finally, conclusions are given in Section 6.

2. Related works

As introduced in the previous section, many computational models for visual saliency have been proposed for 2D images or video sequences. With the popularity of 3D images, several studies have been researched for the 3D images. In [21], a stereo attention framework is proposed by extending the existing attention model from 2D to the binocular domain. Multiple perceptual stimuli are employed for a stereoscopic visual attention model in [22]. Region-of-interest (ROI) extraction method is proposed by Chamaret et al. for adaptive rendering [23]. The depth information is employed to weight the 2D saliency map for generating the final saliency map of 3D image in [22,23]. Ouerhani et al. [24] took the depth cues into consideration to develop the 3D saliency map. And Potapova et al. [25] proposed a 3D saliency detection model for robotics tasks by incorporating the top-down depth cues into the bottom-up saliency detection method. Eye tracking experimental results are carried out on 2D and 3D images for depth saliency analysis in [26], where 3D saliency map is calculated by extending previous 2D saliency detection models. Moreover, the work in [27] also extended the saliency detection method for 2D images to 3D images. The features of color and depth are employed in [28] to generate the saliency map for image segmentation. Wang et al. [29], proposed a computational attention model for 3D images by extending the traditional 2D saliency models. More recently, Fang et al. [12] proposed to incorporate the color, luminance, texture, and depth cues to generate the saliency map for 3D images.

It can be observed that the key of visual saliency map for 3D images are the depth cue. The 2D saliency models consider the low-level features, such as color, intensity, orientation, and so on. For 3D image, the depth information is critical for human perception. Therefore, there are many research works on 3D visual saliency by incorporating the depth information, such as [12,26,27]. However, the depth map of the 3D image, specifically the stereoscopic image pairs are not always available, as the accurate depth map is hard to be sensed and captured. Also most of the real 3D applications only provide the images of two different views without the depth map. Therefore, the visual saliency models that explicitly incorporate the depth map are not practical for most applications. In this paper, in order to handle such drawbacks, we do not explicitly use the depth map for modeling visual saliency. Instead, the disparity map and difference image between different views are employed to derive the 3D visual saliency model. As such, the depth cue is implicitly considered.

As human eyes are the ultimate receivers of the stereoscopic image, HVS properties such as binocular vision and depth perception have been considered in developing the SIQA metrics. For example, the depth (or disparity) information and 2D quality metrics are fused together to analyze 3D visual quality in [30,31]. The concept of cyclopean image was investigated to fuse the left and right views, where the monoscopic and stereoscopic quality component are combined together in stereo-video quality assessment metric designing [32]. To further improve the performance of SIQA metric, the binocular fusion and rivalry properties are widely investigated. For example, Wang et al. [33] proposed a binocular spatial sensitivity (BSS) weighted metric based on the binocular JND model [34]. Chen et al. [35] proposed a SIQA metric to improve the prediction performance on asymmetric distortion types. In [36], the linear rivalry model was developed to exploit the binocular rivalry property of HVS. Wang et al. [37] proposed an information content and divisive normalization-based pooling scheme to improve the performance of structural similarity metric for estimating the quality of single-view images. The binocular rivalry inspired multi-scale model is designed to predict the final quality of stereoscopic images. The HVS modeling can help to improve the performances of SIQAs. Therefore, as the most straightforward and important property of HVS, the visual saliency needs to be investigated further for quality assessment of the stereoscopic image pairs. In this paper, we aim to develop an effective visual saliency model for stereoscopic images. Unlike the visual saliency models in prior arts, the proposed saliency model targets at the performance improvement of the SIQA.

3. Stereoscopic image visual saliency

The framework of our proposed visual saliency model for stereoscopic image is illustrated in Fig. 1. As we target at a saliency model for SIQA, two different saliency maps are generated by our proposed saliency models for the left and right view images, respectively. Firstly, each view image is represented as a QR by referring to the other view image. The QR of each view image roughly considers two different types of cues, specifically the image content and disparity cues. The disparity cue considers the inter-view correlation between left and right view image. The

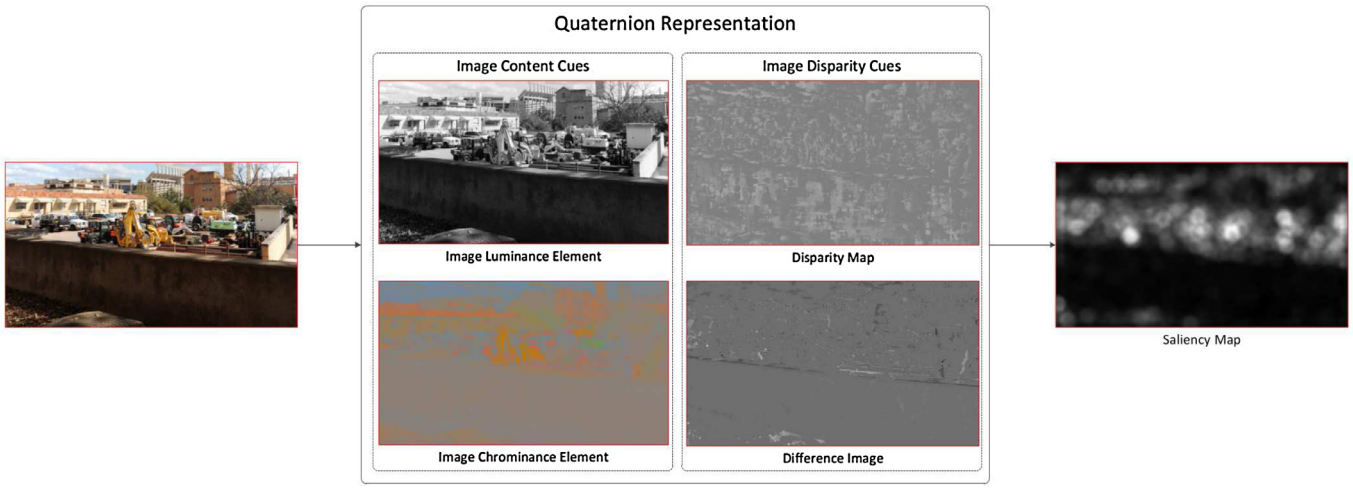


Fig. 1. The framework of our proposed stereoscopic image visual saliency model. For better visualization, the difference image is scaled within the range of [1,255]. And each pixel value of the disparity map with the addition of 128 is illustrated.

QR is further employed to derive the saliency map for each view image. Afterwards, the obtained saliency map will be incorporated into quality metrics to improve their performances thereafter.

3.1. Stereoscopic image quaternion representation

As discussed in Section 2, saliency models for 2D images mainly consider the low-level features, such as color and intensity features, while saliency models for 3D images incorporate the depth cues which are critical for 3D perception. Our new QR of the stereoscopic image considers both the image low-level features and depth information.

3.1.1. Image content cues

As discussed in [11,12], the color and luminance information is helpful for saliency detection of 2D images. Following their approaches, we extract the color and luminance information for visual saliency detection. However, instead of extracting low-level features from the 2D image for characterizing the color and luminance properties, we simply use the luminance and color components of the image to construct the stereoscopic image QR from the image content perspective.

Firstly, each view of stereoscopic image pairs is converted from RGB color space to YUV color space. Then the luminance component Y denoted the image intensity is extracted as one element of the stereoscopic image QR. The chrominance components U and V of each view are merged together as another element of the stereoscopic image QR. In our preliminary exploratory experiments, different merging strategies are tested, such as the averaging, the root of sum squared value, and so on. It is demonstrated that different merging strategies slightly affect the final results of SIQA. Therefore, the simple averaging process is used to merge the U and V components together. The luminance and chrominance components of the reference stereoscopic image are illustrated in Fig. 2. It can be observed that the luminance component comprises most information of the left/right view image. However, the chrominance component indeed depict the salient color information, which will attract the viewers' attention and be helpful for visual saliency detection.

As mentioned before, prior saliency models focus on extracting the low-level features to depict the luminance and chrominance components, which are believed to be useful for saliency detection. In contrary,we use the raw image luminance and chrominance component in this paper. We leave our saliency model to compose and make interactions between the luminance and chrominance components to predict the saliency properties of the stereoscopic images.

3.1.2. Image disparity cues

The depth cue is demonstrated to be critical for visual saliency for 3D image, as the depth map depicts the correlations between the two view images as well as the relative positions of the objects in the image. Therefore, many research papers [12,24] employed the depth map to derive the visual saliency map for the 3D image. However, the accurate depth map is always unavailable, as it is hard to be sensed and captured. In this paper, instead of the truly captured depth map, the disparity map estimated from the stereoscopic image is employed to depict their correlations. Moreover, the difference image is further computed by referring to the disparity map. These two images, regarded as the disparity cues of the stereoscopic image, are employed as the disparity elements of the stereoscopic image QR. As such, although without truly captured depth map, the relationship between the left and right view images as well as the object relative locations are implicitly considered.

We employed the work in [38] to obtain the disparity map for each view image by referring to the other view image. The contrast-invariant correspondence between the two different view images are obtained by performing local matching using phase information from a bank of Gabor filters. As the phase differences for local matching is only used and not for explicitly computing the correspondence, the filters of large spatial extent do not need to be computed for large shifts, which prevents degradation of boundaries. And the algorithm is able to handle significant changes in contrast between the two images even if the changes vary spatially over the image, and performs well in the presence of noise. As the matching between the two view images is not our contributions, we do not provide the detailed approach here. Detailed information about the method can be found in [38].

We assume that the disparity map of the left view image by referring to the right view image is obtained by the method [38], which is denoted as M_d. As the correspondence between the image pixels are bidirectional, the disparity map of the right view image referring to the left view image will be $-M_d$. Then the difference image between the two view images is obtained by:

$$I_d(i,j) = I_l(i,j) - I_r(i, clip(1, I_{width}, j + M_d(i,j))) \quad (1)$$

where (i, j) is the pixel position. I_l and I_r are the left and right view images, respectively. I_{width} denotes the width of the image. I_d

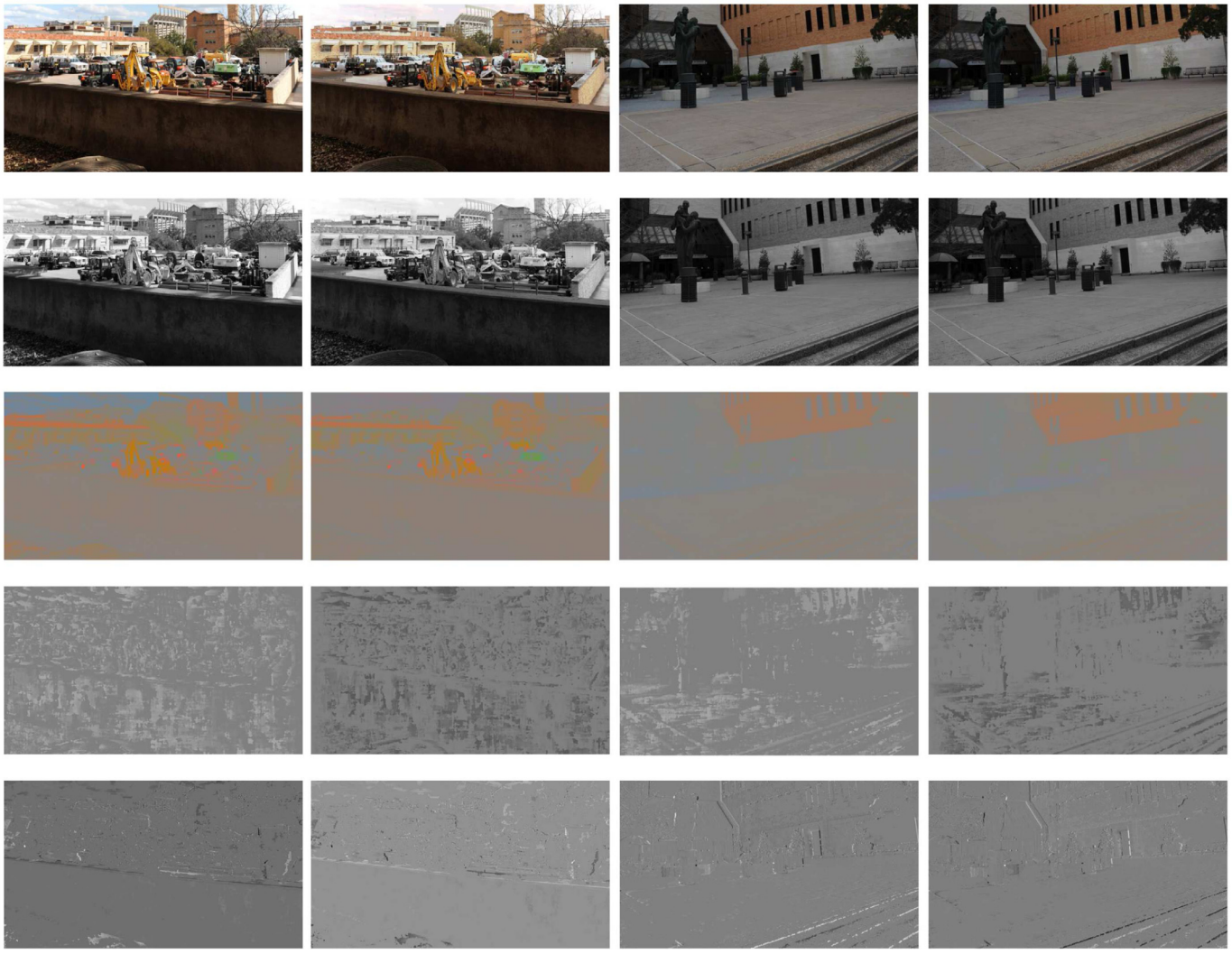


Fig. 2. Elements for composing the stereoscopic image QR. From top to bottom: the left/right view image, the luminance component, the chrominance component, the disparity component, and the difference image component.

is the obtained difference image between the left and right view image. The *clip*(·) function ensures that the mapped pixel locates within the image. The disparity map and the difference image are illustrated in Fig. 2. It can be observed that the disparity map depicts the object locations within the image, while the difference image depicts the image differences introduced by inter-view dissimilarities.

3.1.3. Quaternion representation

With the above processes, we obtain four elements for each view image from both the image content and disparity perspectives. Afterwards, each view image I (I can be left or right view image) is represented as a quaternion image (I_i, I_c, I_d, M_d), where I_i and I_c denote the image luminance and chrominance components respectively. In order to generate the visual saliency map, a new quaternion representation (QR) I_q [39] of each view image is represented by considering the four different quaternion elements as:

$$\begin{aligned} &I_q = I_i + I_c\mu_1 + I_d\mu_2 + M_d\mu_3 \\ &\text{where,}\quad \mu_i^2 = -1, i = 1,2,3 \\ &\qquad \mu_1 \perp \mu_1, \mu_2 \perp \mu_3, \mu_3 \perp \mu_1 \\ &\qquad \mu_3 = \mu_1\mu_2 \end{aligned} \tag{2}$$

A symplectic form of I_q can be further expressed by:

$$\begin{aligned} &I_q = f_1 + f_2\mu_2, \\ &\text{where,}\quad f_1 = I_i + I_c\mu_1 \\ &\qquad\quad f_2 = I_d + M_d\mu_1 \end{aligned} \tag{3}$$

In [18,40], a quaternion image is composed to depict each frame of the video sequence. For the quaternion image in [18], one intensity element, two color elements, and one motion element are employed to compose the quaternion image. For [40], one intensity element and three motion elements (considering the motion vectors in two dimension and the prediction error) are used to compose the quaternion image. Their quaternion representations are not practical for stereoscopic image. In this paper, we consider both the image content and disparity cues to compose the quaternion image, which not only includes the low-level features, such as intensity and color from 2D image, but also considers the depth information for 3D perception, such as the disparity map and the difference image. These four elements are expected to compose and interact with each other to generate the visual saliency of the stereoscopic images.

3.2. Quaternion representation based stereoscopic image visual saliency (QRSIVS)

As demonstrated in [18,40], the phase spectrum is employed to generate the saliency information for each video frame. Providing an image $I(i, j)$, the saliency map is generated by:

$$SM(i,j) = g(i,j) * \| F^{-1}(e^{i\cdot p(x,y)}) \|^2, \quad \text{where,} \quad f(x,y) = F(I(i,j)) \quad p(x,y) = P(f(i,j)) \tag{4}$$

where F and F^{-1} denote the Fourier transform and inverse Fourier transform, respectively. $f(x, y)$ is the Fourier representation of the given image, $p(x, y)$ denotes the phase information of the $f(x, y)$. $g(i, j)$ is a smoothing filter. The saliency map SM is generated by only considering the phase spectrum of the given image.

As a quaternion image is constructed for each view image, the quaternion Fourier transform (QFT) [39] is thus employed instead of Fourier transform to generate the corresponding visual saliency map. For the quaternion image illustrated in Eq. (3), the QFT is performed according to:

$$I_Q(u,v) = F_1(u,v) + F_2(u,v)\mu_2, \tag{5}$$

where

$$F_i(u,v) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} e^{-\nu_1 2\pi(\frac{mv}{M}+\frac{nu}{N})} f_i(n,m), \tag{6}$$

where (n, m) and (u, v) denote the locations in the spatial and frequency domain. N and M indicate the image height and width, respectively. f_i, $i \in \{1, 2\}$ is obtained from Eq. (3). F_i is the obtained Fourier representation of f_i. The QFT I_Q of the quaternion image I_q can be further expressed in the polar form as:

$$I_Q = \| I_Q \| e^{\mu \cdot p} \tag{7}$$

where p is the phase spectrum of the Fourier representation I_Q, and μ is a unit pure quaternion.

As mentioned before, only the phase spectrum is enough to construct the visual saliency map. Therefore, only the phase spectrum of I_Q is preserved to generate the saliency map. The magnitude value $\|I_Q\|$ is set as 1 to eliminate the affection of the magnitude spectrum. The QFT representation is further modified as:

$$I_Q^m = e^{\mu \cdot p} \tag{8}$$

Afterwards, the inverse QFT is performed on I_Q^m, which is defined as:

$$f_i^m(n,m) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} e^{-\nu_1 2\pi(\frac{mv}{M}+\frac{nu}{N})} F_i^m(u,v) \tag{9}$$

where F_i^m is the modified Fourier representation by setting the magnitude value as 1 according to Eq. (8). By performing the inverse QFT and composing all the f_i^m images:

$$I_S = f_1^m + f_2^m \mu_2 \tag{10}$$

The quaternion image I_S is constructed. We further employ a filter to smooth I_S by:

$$SM = g * \| I_S \|^2 \tag{11}$$

where g is the smoothing filter. I_S is the quaternion image constructed by inverse QFT. $\|I_S\|^2$ is the constructed image in the image domain from the saliency model. In this paper, the Gaussian filter is employed to smooth the image for simplicity.

4. Quaternion representation based stereoscopic image visual saliency (QRSIVS) for stereo image quality assessment

For traditional 2D IQA metrics, the saliency maps have been widely applied for guiding the spatial pooling stage to improve the performance. Based on the previous discussion, stereoscopic image visual saliency map can indicate the relative importance of pixels in the spatial domain for left/right views. Therefore, it is natural to incorporate the constructed QRSIVS into the designing of SIQA metrics. Fig. 3 presents the framework of proposed QRSIVS weighted SIQA model. For each stereoscopic image pair, the saliency maps SM_l and SM_r for left and right views are extracted by Eq. (11), respectively. The traditional spatial domain based 2D IQA metrics can be employed to generate the error maps EM_l and EM_r for the left and right view image, respectively. Finally, the saliency maps are employed to pool the error maps as the image quality score Q_s of distorted stereoscopic image pairs. The general mathematic form of proposed QRSIVS weighted SIQA index is given by:

$$Q_s = \sum_{i\in\{l,r\}} c_i \frac{\sum_{x\in\Omega_i} SM_i(x)\cdot EM_i(x)}{\sum_{x\in\Omega_i} SM_i(x)}, \tag{12}$$

where c_l and c_r are the weighting factors of the left and right views, respectively. Ω_l and Ω_r are the spatial domains of the left and right views, respectively.

5. Experimental results

In this section, we implement the proposed SIQA metric and make performance comparisons with the state-of-the-art methods. To validate the robustness of proposed metric, it is necessary to evaluate the SIQA metrics on different 3D image quality databases (IQDs). Currently, there are two categories for existing 3D IQDs. One is symmetric IQD where the left/right views of the stereoscopic image are symmetric distorted. The other category is the asymmetric IQD where the left/right views of the stereoscopic image are degraded with different distortion types and levels. In this paper, we evaluate the effectiveness of the SIQA metrics on two typical symmetric IQDs as well as its generality on one asymmetric IQD. The detailed information of the selected IQDs is described as follows:

- **LIVE 3D IQD Phase I (LIVE-Phase-I)** [41] consists of 20 outdoor stereoscopic scenes. Each scene contains one stereoscopic pairs (left/right view) and the corresponding range maps of the views. All the reference stereoscopic images are with resolution 640×360. For each reference stereoscopic image, its left/right views are symmetrically degraded by five different distortion types with different degradation levels. The distortion types include JPEG compression (denoted as JPEG), JPEG2000 compression (denoted as JP2K), white noise contamination (denoted as WN), Gaussian blur (denoted as GBLUR), and fast fading channel distortion of JPEG2000 compressed bitstream (denoted as FF). The database contains 365 subject-rated stereoscopic image pairs (80 each for JP2K, JPEG, WN and FF; 45 for GBLUR).
- **Ningbo University IQD Phase II (Ningbo-Phase-II)** [42] aims to build a diverse database that consists of a wide variety of scenes and distortions. The database contains 12 outdoor and indoor stereoscopic scenes. The resolutions are from 480×270 to 1280×960. The distortion types include JPEG, JP2K, WN, GBLUR and H.264 compressed bitstream (denoted as H264). The database consists of 312 subject-rated stereoscopic image pairs (60 each for JP2K, JPEG, WN and GBLUR; 72 for H264).
- **LIVE 3D IQD Phase II (LIVE-Phase-II)** [35] consists of both symmetrically and asymmetrically distorted stereoscopic pairs. Same as LIVE-Phase-I, the introduced distortion types include JPEG, JP2K, WN, GBLUR and FF. The database consists of 360 subject-rated stereoscopic images (72 each for JP2K, JPEG, WN, GBLUR and FF).

For fair comparisons, both the 2D IQA extension models and binocular vision inspired metrics (denoted as 3D IQA model)

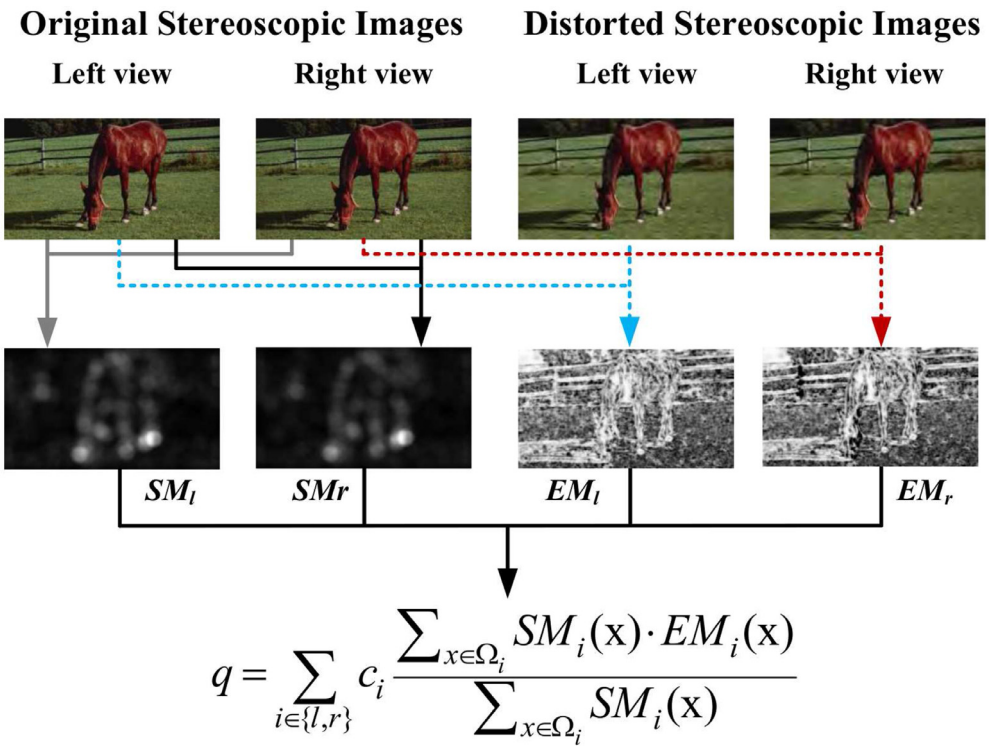


Fig. 3. The framework of proposed QRSIVS weighted SIQA model for the stereoscopic image pair.

are evaluated in the experiment. Two 3D IQA models, including FI-PSNR [43] and MJ3DQA [35] are compared in the experiment. To verify the effectiveness of the proposed QRSIVS model, three different IQA metrics, including SSIM, multi-scale SSIM (MS-SSIM) [44], and edge-strength-similarity (ESSIM) [45] are employed as the basic IQA metrics. For MS-SSIM, the extracted visual saliency map is processed with the same filters in the MS-SSIM. Besides, to demonstrate the effectiveness of proposed QRSIVS index, three state-of-the-art methods, including spectral residual (SR) approach [17], saliency detection (SD) approach [11], and 3D saliency detection (3DSD) [12] are also implemented and compared. As shown in Tables 1 and 2, 15 metrics in total (12 saliency map weighted SIQA metrics) are tested and compared.

To remove the nonlinearity introduced by the subjective rating process and further facilitate the empirical comparison of different IQA metrics, the nonlinear least-squares regression function *nlinfit* of Matlab is employed to map the objective quality score Q_s to the predicted subjective quality score $DMOS_P$. The mapping function is the five parameters logistic function defined as:

$$DMOS_p = \frac{p_1}{2} - \frac{p_1}{1 + exp(p_2 \cdot (Q_s - p_3))} + p_4 \cdot q + p_5, \tag{13}$$

where p_1, p_2, p_3, p_4 and p_5 are the parameters of the logistic function. Three criteria are employed to evaluate the corresponding performance: (1) correlation coefficient (CC): accuracy of objective metrics; (2) Spearman's rank order correlation coefficient (SROCC): monotonicity of objective metrics; and (3) root mean-squared-error (RMSE). Detailed experimental results are provided in Tables 1 and 2. For each group of saliency map weighted SIQA metrics, the metric with the best performance is highlighted in bold. Also we provided the scatter plots of subjective *DMOS* values against the predicted $DMOS_p$ values of the SIQA metrics on the 3D IQDs in Figs. 4–6.

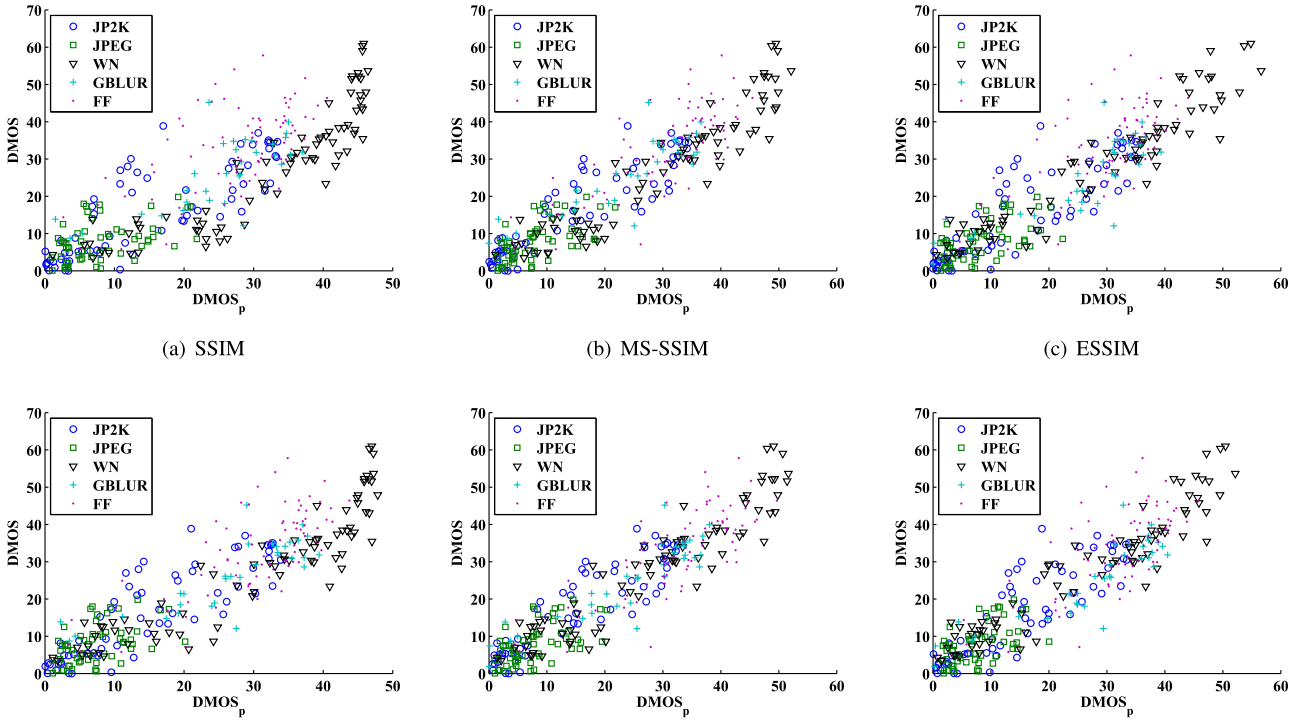
5.1. Comparison with the stereoscopic image quality metrics

The stereoscopic image present different visual experiences for the human viewers, where the depth perception is most important. Therefore, there are a thread of work on SIQA by considering the depth information, such as MJ3DQA [35] and FI-PSNR [43]. In MJ3DQA [35], the authors proposed to construct an intermediate image which when viewed stereoscopically is designed to have a perceived quality close to that of the cyclopean image. They hypothesized that performing stereoscopic QA on the intermediate image yields higher correlations with human subjective judgments. In FI-PSNR [43], besides the traditional 2D image metrics, the HVS behaviors on 3D content perception, specifically the binocular integration behaviors-the binocular combination and the binocular frequency integration, are utilized as the bases for measuring the quality of stereoscopic 3D images.

Compared with MJ3DQA and FI-PSNR, in most cases, the proposed saliency map based SIQA framework can achieve better performances on both LIVE-Phase-I and Ningbo-Phase-II. The reason can be attributed to that the mechanism of binocular summation is still an open issue. Thus the computation model of the rivalry property may not be accurate enough for assessing the perceptual quality of 3D images. That is also the main reason why the performances of existing binocular vision inspired metrics are limited. In contrary, the saliency map as the most straightforward and effective HVS property has been extensively researched and studied. Thus the saliency map weighted approach is demonstrated to be an effective and simple way to improve the performance of quality prediction.

5.2. Comparison with other visual saliency models

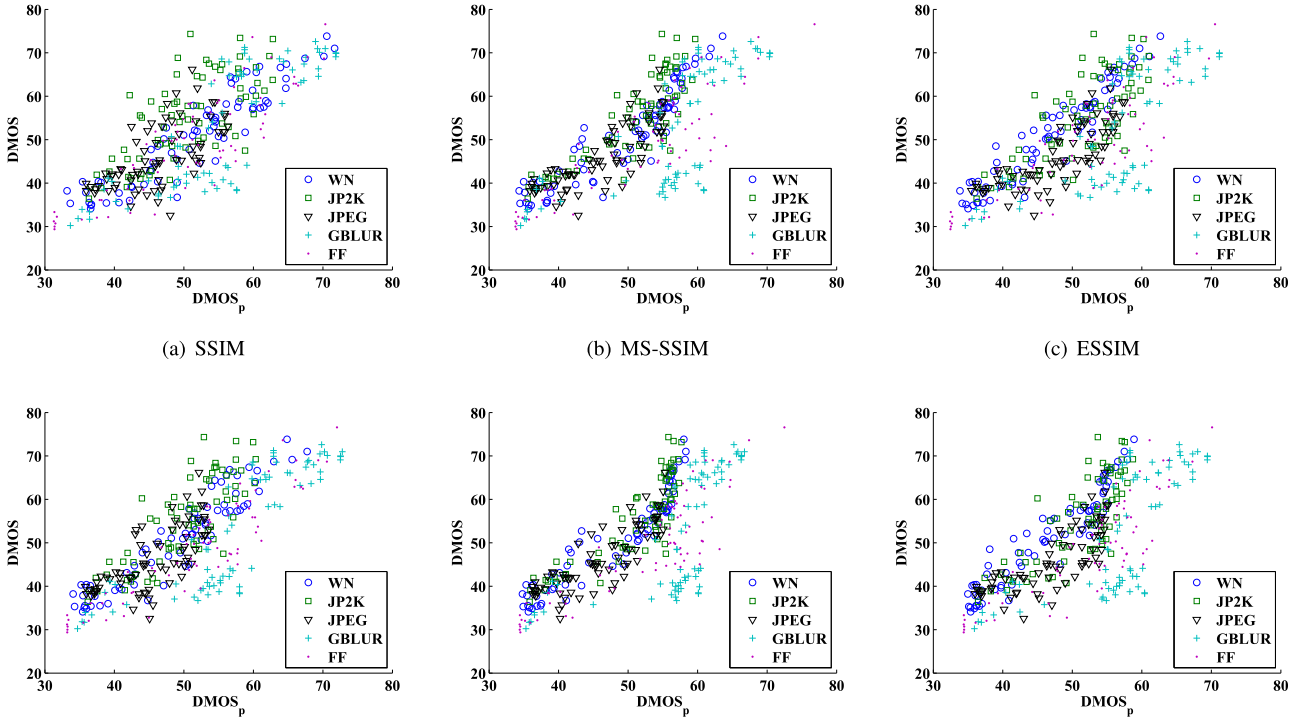
In this section, we compare the performances of different visual saliency models on SIQA, specifically the SR approach [17], SD approach [11], and 3DSD [12] approach. SR analyzed the log-spectrum of an input image, where the spectral residual of an



(a) SSIM (b) MS-SSIM (c) ESSIM

(d) SSIM+QRSIVS (e) MS-SSIM+QRSIVS (f) ESSIM+QRSIVS

Fig. 4. Scatter plots of subjective *DMOS* vs. predicted $DMOS_p$ of SIQA metrics on the LIVE-Phase-I database.



(a) SSIM (b) MS-SSIM (c) ESSIM

(d) SSIM+QRSIVS (e) MS-SSIM+QRSIVS (f) ESSIM+QRSIVS

Fig. 5. Scatter plots of subjective *DMOS* vs. predicted $DMOS_p$ of SIQA metrics on the LIVE-Phase-II database.

Table 1
Performance of the SIQA metrics on LIVE-Phase-I database in terms of CC, SROCC, and RMSE.

Criterion	Metric	JP2K	JPEG	WN	GBLUR	FF	ALL
CC	PSNR	0.7879	0.1191	0.9352	0.7701	0.6948	0.8384
	FI-PSNR	0.8575	0.3266	0.9289	0.8191	0.7096	0.8733
	MJ3DQA	0.9285	0.6575	0.9580	0.9413	0.7489	0.9212
	SSIM	0.8752	0.4883	0.9437	0.9192	0.7243	0.8770
	+SR	0.9015	0.5221	0.9468	0.9395	0.8002	0.9093
	+SD	0.8825	0.4940	0.9416	0.9289	0.7487	0.8893
	+3DSD	0.8787	0.4823	0.9398	0.9306	0.7457	0.8882
	+QRSIVS	**0.9039**	**0.5356**	**0.9478**	**0.9442**	**0.8004**	**0.9180**
	MS-SSIM	0.9335	0.6663	**0.9522**	0.9449	0.8083	0.9297
	+SR	0.9328	0.6613	0.9459	0.9502	0.8280	0.9362
	+SD	0.9321	**0.6866**	0.9451	0.9474	0.8167	0.9315
	+3DSD	0.9287	0.6416	0.9412	0.9481	0.8186	0.9301
	+QRSIVS	**0.9346**	0.6741	0.9448	**0.9512**	**0.8335**	**0.9373**
	ESSIM	0.9011	**0.6713**	**0.9526**	0.9313	0.7468	0.9145
	+SR	0.8905	0.6239	0.9525	0.9448	0.7579	0.9201
	+SD	0.9040	0.6565	0.9511	0.9374	0.7566	0.9176
	+3DSD	0.9024	0.6647	0.9476	0.9379	0.7573	0.9165
	+QRSIVS	**0.9087**	0.6171	0.9488	**0.9509**	**0.7781**	**0.9240**
SROCC	PSNR	0.7993	0.1212	0.9316	0.9020	0.5873	0.8365
	FI-PSNR	0.8522	0.2568	0.9297	0.9394	0.6599	0.8644
	MJ3DQA	0.8938	0.5612	0.9502	0.9223	0.6685	0.9134
	SSIM	0.8581	0.4347	0.9387	0.8793	0.5871	0.8767
	+SR	0.8752	0.4658	0.9381	0.9126	0.6787	0.9045
	+SD	0.8609	0.4568	0.9345	0.8984	0.6195	0.8880
	+3DSD	0.8586	0.4420	0.9330	0.9062	0.6151	0.8873
	+QRSIVS	**0.8776**	**0.4893**	**0.9406**	**0.9236**	**0.7178**	**0.9120**
	MS-SSIM	0.8978	**0.5985**	**0.9423**	0.9282	0.7349	0.9225
	+SR	0.8971	0.5731	0.9423	0.9298	0.7687	0.9273
	+SD	0.8958	0.5954	0.9403	0.9253	0.7559	0.9243
	+3DSD	0.8934	0.5680	0.9347	0.9292	0.7604	0.9233
	+QRSIVS	**0.8989**	0.5855	0.9394	**0.9335**	**0.7877**	**0.9287**
	ESSIM	0.8752	**0.5504**	**0.9498**	0.9026	0.6310	0.9073
	+SR	0.8726	0.4969	0.9471	0.9235	0.6572	0.9112
	+SD	0.8754	0.5444	0.9457	0.9163	0.6518	0.9102
	+3DSD	0.8737	0.5452	0.9440	0.9169	0.6512	0.9103
	+QRSIVS	**0.8810**	0.5129	0.9441	**0.9306**	**0.6864**	**0.9162**
RMSE	PSNR	7.9752	6.5098	5.8916	9.2335	8.9365	8.9363
	FI-PSNR	6.6623	6.2512	6.1612	8.3026	8.7545	7.9886
	MJ3DQA	4.8090	4.9267	4.7713	4.8870	8.2336	6.3810
	SSIM	6.2643	5.7066	5.5008	5.7011	8.5670	7.8794
	+SR	5.6060	5.5771	5.3523	4.9567	7.4516	6.8235
	+SD	6.0907	5.6855	5.6002	5.3598	8.2376	7.4984
	+3DSD	6.1818	5.7282	5.6843	5.2964	8.2785	7.5325
	+QRSIVS	**5.5398**	**5.5221**	**5.3065**	**4.7692**	**7.4488**	**6.5041**
	MS-SSIM	4.6426	4.8762	**5.0822**	4.7386	7.3159	6.0398
	+SR	4.6689	4.9050	5.3964	4.5092	6.9674	5.7644
	+SD	4.6900	**4.7552**	5.4342	4.6323	7.1697	5.9660
	+3DSD	4.8038	5.0159	5.6217	4.6006	7.1374	6.0210
	+QRSIVS	**4.6081**	4.8302	5.4529	**4.4676**	**6.8656**	**5.7156**
	ESSIM	5.6166	**4.8478**	**5.0677**	5.2737	8.2631	6.6355
	+SR	5.8935	5.1105	5.0691	4.7421	8.1055	6.4235
	+SD	5.5369	4.9330	5.1413	5.0403	8.1244	6.5194
	+3DSD	5.5804	4.8856	5.3169	5.0196	8.1143	6.5592
	+QRSIVS	**5.4072**	5.1456	5.2537	**4.4806**	**7.8043**	**6.2721**

image in spectral domain is extracted to construct the corresponding saliency map of the image. SD constructed the saliency map in the compressed domain. The intensity, color, and texture features of the image are extracted from discrete cosine transform (DCT) coefficients in the JPEG bit-stream. The saliency value of each DCT block is obtained based on the Hausdorff distance calculation and feature map fusion. Both the SR and SD approaches focused on the saliency detection of 2D images. For 3D images, specifically the stereoscopic image, the depth information needs to be considered for the saliency detection. 3DSD extracted four types of features, namely color, luminance, texture, and depth, from DCT coefficients for feature contrast calculation. A Gaussian model of the spatial distance between image patches is adopted for consideration of local and global contrast calculation. Moreover, the center bias factor and human visual acuity, the important characteristics of the human visual system, are further employed to enhance the saliency map for stereoscopic images.

From Tables 1 and 2, it can be observed that the SIQA based on our proposed QRSIVS outperforms the other saliency map based approaches on both the LIVE-Phase-I and Ningbo-Phase-II 3D IQDs. For example, for the saliency map weighed SSIM metrics, the CC value of our proposed metric SSIM+QRSIVS on Ningbo-Phase-II is 0.8845, where the original metric SSIM is only 0.8094 and the best competitor SSIM+SR is 0.8621. The reason that our proposed QRSIVS outperforms SR and SD is that QRSIVS explicitly considers the depth information of the stereoscopic image in terms of the disparity map and difference image. SR and SD only focus on the 2D natural image, which only consider the image content cues, such as luminance, contrast, color information, and so on. However, for the stereoscopic image perception, the depth

Table 2
Performance of the SIQA metrics on Ningbo-Phase-II database in terms of CC, SROCC, and RMSE.

Criterion	Metric	JP2K	JPEG	WN	GBLUR	H264	ALL
CC	PSNR	0.9598	0.6949	0.8311	0.9606	0.9079	0.8852
	FI-PSNR	0.9579	0.9468	0.9381	0.9148	0.9627	0.9037
	MJ3DQA	0.9561	0.9010	0.9090	0.9565	0.9298	0.9073
	SSIM	0.9015	0.8273	0.7566	0.9369	0.8524	0.8094
	+SR	0.9408	0.8766	0.8681	0.9334	0.9063	0.8621
	+SD	0.9218	0.8637	0.8218	0.9406	0.8823	0.8347
	+3DSD	0.9107	0.8606	0.7720	0.9394	0.8710	0.8248
	+QRSIVS	**0.9617**	**0.8814**	**0.8889**	**0.9417**	**0.9236**	**0.8845**
	MS-SSIM	0.9687	0.9327	0.9495	0.9379	0.9516	0.9186
	+SR	0.9800	**0.9417**	0.9604	0.9322	**0.9697**	0.9438
	+SD	0.9761	0.9414	0.9538	0.9422	0.9606	0.9338
	+3DSD	0.9755	0.9385	0.9495	0.9373	0.9596	0.9305
	+QRSIVS	**0.9806**	0.9319	**0.9623**	**0.9451**	0.9691	**0.9473**
	ESSIM	0.9121	0.8829	0.7435	0.9460	0.8519	0.8723
	+SR	0.9363	**0.9099**	0.8533	0.9439	0.9067	0.9167
	+SD	0.9225	0.9005	0.8168	0.9474	0.8776	0.8943
	+3DSD	0.9120	0.8932	0.7641	0.9417	0.8616	0.8824
	+QRSIVS	**0.9525**	0.9091	**0.8908**	**0.9521**	**0.9211**	**0.9247**
SROCC	PSNR	0.9529	0.8496	0.8628	0.9499	0.9049	0.9032
	FI-PSNR	0.9501	0.9436	0.9483	0.8552	0.9558	0.8841
	MJ3DQA	0.9517	0.9214	0.9185	0.9274	0.8919	0.9031
	SSIM	0.9132	0.8494	0.8085	0.8808	0.8462	0.8413
	+SR	0.9373	0.8897	0.8889	0.8732	0.9036	0.8853
	+SD	0.9260	0.8861	0.8570	0.8913	0.8797	0.8631
	+3DSD	0.9168	0.8764	0.8236	**0.8916**	0.8785	0.8556
	+QRSIVS	**0.9579**	**0.8962**	**0.9051**	0.8909	**0.9136**	**0.9026**
	MS-SSIM	0.9690	0.9365	0.9349	0.8879	0.9397	0.9214
	+SR	0.9761	0.9437	**0.9527**	0.8774	0.9475	0.9370
	+SD	0.9741	**0.9442**	0.9489	0.8955	0.9505	0.9320
	+3DSD	**0.9763**	0.9415	0.9483	0.8911	0.9487	0.9306
	+QRSIVS	0.9754	0.9341	0.9510	**0.8991**	**0.9523**	**0.9397**
	ESSIM	0.9199	0.8967	0.7831	0.9174	0.8524	0.8847
	+SR	0.9301	**0.9201**	0.8807	0.9041	0.9009	0.9195
	+SD	0.9265	0.9125	0.8377	0.9135	0.8786	0.9028
	+3DSD	0.9172	0.9071	0.8087	0.9089	0.8730	0.8934
	+QRSIVS	**0.9484**	0.9192	**0.9073**	**0.9194**	**0.9139**	**0.9279**
RMSE	PSNR	5.9207	10.2422	6.6656	4.3438	5.8820	7.9931
	FI-PSNR	6.0554	4.5850	4.1502	6.3105	3.7991	7.3548
	MJ3DQA	6.1783	6.1780	4.9965	4.5592	5.1652	7.2240
	SSIM	9.1283	8.0017	7.8368	5.4636	7.3380	10.0887
	+SR	7.1487	6.8553	5.9491	5.6080	5.9298	8.7041
	+SD	8.1777	7.1791	6.8291	5.3031	6.6057	9.4595
	+3DSD	8.7092	7.2530	7.6186	5.3584	6.8936	9.7139
	+QRSIVS	**5.7791**	**6.7274**	**5.4901**	**5.2578**	**5.3781**	**8.0135**
	MS-SSIM	5.2326	5.1366	3.7596	5.4207	4.3126	6.7882
	+SR	4.1935	**4.7920**	3.3404	5.6552	**3.4299**	5.6756
	+SD	4.5811	4.8061	3.6008	5.2364	3.9022	6.1481
	+3DSD	4.6377	4.9182	3.7611	5.4448	3.9482	6.2937
	+QRSIVS	**4.1346**	5.1680	**3.2613**	**5.1047**	3.4619	**5.5020**
	ESSIM	8.6436	6.6875	8.0155	5.0634	7.3490	8.4017
	+SR	7.4080	**5.9089**	6.2490	5.1595	5.9179	6.8663
	+SD	8.1404	6.1941	6.9364	4.9996	6.7274	7.6872
	+3DSD	8.6526	6.4047	7.7316	5.2587	7.1226	8.0834
	+QRSIVS	**6.4209**	5.9337	**5.4466**	**4.7783**	**5.4644**	**6.5414**

information is much more important, which needs to be taken into consideration. The 3DSD model demonstrated high accuracy on predicting the human eye fixation point of the stereoscopic image. However, it is demonstrated that 3DSD cannot well boost the performance on SIQA, compared with SR and SD. In contrary, our proposed QRSIVS targets at the performance enhancement of SIQA, which yields better performances than 3DSD even with the depth information explicitly considered. Based on the observations above, we can make a conclusion that our proposed 3DQRVS based SIQA framework is powerful for predicting the 3D visual quality of stereoscopic images.

5.3. Performances on difference distortions

By breaking down to each distortion type, we can observe that the QRSIVS based quality metrics can mostly outperform other competitor models. Specifically, on the LIVE-Phase-I database, the propose QRSIVS based quality metrics achieve the best performance on the WN, GBLUR, and FF distortion types. On the Ningbo-Phase-II database, the QRSIVS based quality metrics achieve the best performance on the JP2K, WN, and GBLUR distortion types. However, the proposed QRSIVS based SIQAs perform the worst on the type of JPEG of the LIVE-Phase-I database, as shown in Table 1. This is due to that the distortions of JPEG images are less perceptually separated, and thus are more challenging to be assessed [36].

5.4. Generality of the proposed QRSIVS

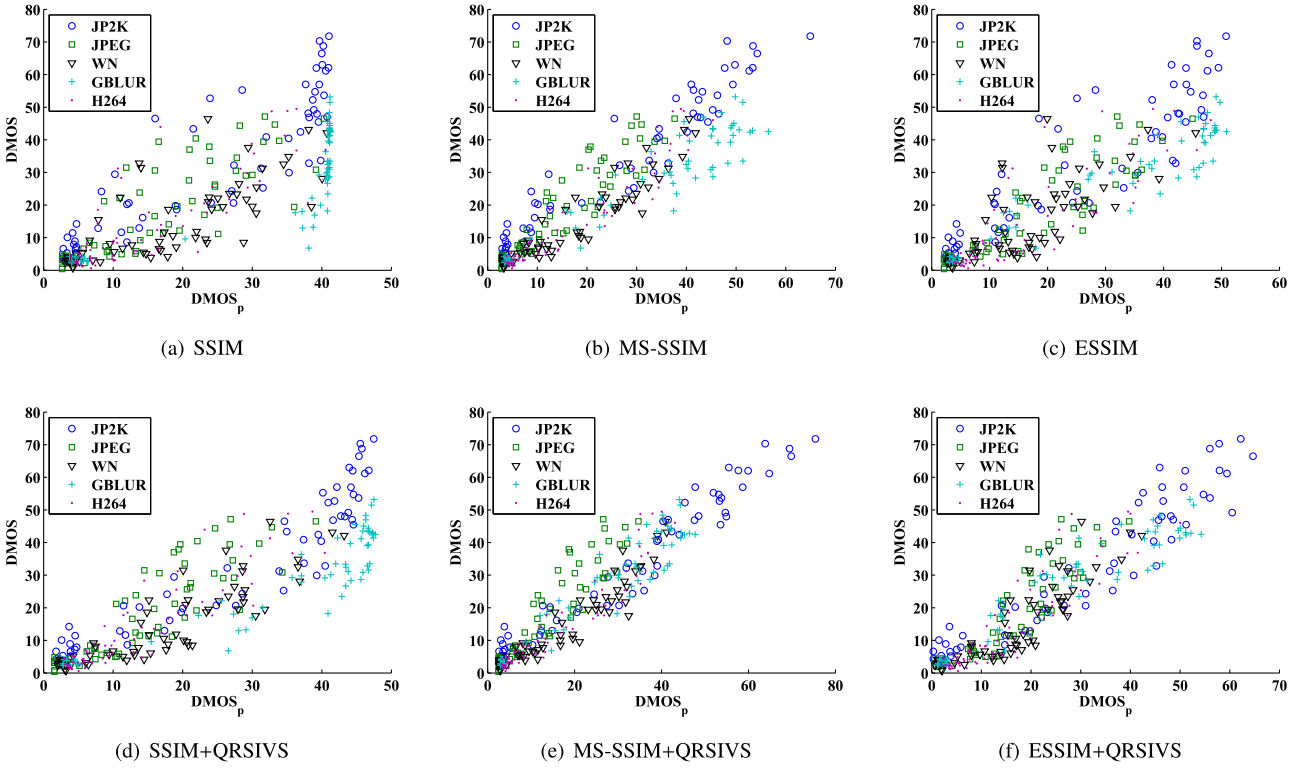
In this section, we test the proposed QRSIVS based quality metrics on LIVE-Phase-II dataset to evaluate its generality on the asymmetric distortions of the stereoscopic image. The results are illustrated in Table 3. It can be observed that the saliency map based

Table 3
Performance of the SIQA metrics on LIVE-Phase-II database in terms of CC, SROCC, and RMSE.

Criterion	Metric	WN	JP2K	JPEG	GBLUR	FF	ALL
CC	PSNR	0.9174	0.6115	0.4650	0.7133	0.7636	0.6808
	FI-PSNR	0.9247	0.7752	0.6677	0.7384	0.7157	0.6450
	MJ3DQA	0.9641	0.8594	0.8322	0.9617	0.9179	0.9099
	SSIM	0.9311	0.7259	0.6662	0.8491	0.8685	0.8030
	+SR	0.9401	0.7925	0.7290	0.9191	0.9085	**0.8141**
	+SD	0.9386	0.7398	0.6674	0.8784	0.8874	0.8093
	+3DSD	0.9351	0.7509	0.6793	0.8603	0.8842	0.8059
	+QRSIVS	**0.9441**	**0.8077**	**0.7360**	**0.9376**	**0.9161**	0.8110
	MS-SSIM	0.9510	0.8389	0.8324	0.7995	0.8740	**0.7938**
	+SR	**0.9656**	0.8783	**0.8574**	0.8188	0.8837	0.7791
	+SD	0.9604	0.8469	0.8231	0.8126	0.8778	0.7869
	+3DSD	0.9607	0.8571	0.8290	0.7994	0.8790	0.7863
	+QRSIVS	0.9651	**0.8822**	0.8536	**0.8254**	**0.8849**	0.7741
	ESSIM	0.9539	0.7559	0.8296	0.7704	0.8397	**0.7653**
	+SR	0.9556	0.7697	0.8138	0.7726	0.8604	0.7497
	+SD	0.9549	0.7493	0.8263	0.7742	0.8464	0.7575
	+3DSD	**0.9575**	0.7575	**0.8360**	0.7634	0.8410	0.7569
	+QRSIVS	0.9569	**0.7895**	0.8294	**0.7837**	**0.8677**	0.7445
SROCC	PSNR	0.9189	0.5966	0.4909	0.6902	0.7301	0.6651
	FI-PSNR	0.9148	0.7437	0.6681	0.7088	0.6945	0.6456
	MJ3DQA	0.9573	0.8527	0.8314	0.9031	0.8919	0.9051
	SSIM	0.9224	0.7041	0.6777	0.8379	0.8343	0.7919
	+SR	0.9340	0.7755	0.7176	0.8726	0.8896	**0.7994**
	+SD	0.9338	0.7178	0.6777	0.8558	0.8591	0.7965
	+3DSD	0.9293	0.7363	0.6856	0.8445	0.8534	0.7930
	+QRSIVS	**0.9390**	**0.7907**	**0.7222**	**0.8847**	**0.8939**	0.7920
	MS-SSIM	0.9473	0.8172	0.8271	0.8010	0.8304	0.7719
	+SR	0.9639	0.8682	**0.8425**	0.8323	0.8339	0.7411
	+SD	0.9565	0.8265	0.8130	0.8301	0.8302	**0.7596**
	+3DSD	0.9594	0.8383	0.8105	0.8024	0.8327	0.7571
	+QRSIVS	**0.9654**	**0.8694**	0.8378	**0.8483**	**0.8350**	0.7349
	ESSIM	0.9527	0.7248	0.8278	0.7411	0.8016	**0.7466**
	+SR	0.9518	0.7536	0.8251	0.7593	0.8202	0.7246
	+SD	**0.9536**	0.7269	0.8238	0.7476	0.8085	0.7355
	+3DSD	0.9532	0.7351	**0.8373**	0.7403	0.7997	0.7351
	+QRSIVS	0.9533	**0.7734**	0.8259	**0.7703**	**0.8307**	0.7144
RMSE	PSNR	4.2641	8.3015	6.4895	9.7582	7.4294	8.2674
	FI-PSNR	4.0781	6.2014	5.4572	9.3891	8.0362	8.6255
	MJ3DQA	2.8450	5.0193	4.0640	3.8158	4.5659	4.6812
	SSIM	3.9078	6.7514	5.4694	7.3556	5.7038	6.7271
	+SR	3.6516	5.9873	5.0179	5.4976	4.8082	**6.5555**
	+SD	3.6973	6.6046	5.4588	6.6555	5.3054	6.6307
	+3DSD	3.7954	6.4830	5.3790	7.0971	5.3742	6.6829
	+QRSIVS	**3.5323**	**5.7875**	**4.9627**	**4.8456**	**4.6125**	6.6032
	MS-SSIM	3.3132	5.3428	4.0621	8.3631	5.5908	**6.8644**
	+SR	**2.7867**	4.6933	**3.7723**	7.9930	5.3863	7.0755
	+SD	2.9853	5.2197	4.1629	8.1154	5.5121	6.9646
	+3DSD	2.9731	5.0562	4.0996	8.3659	5.4872	6.9745
	+QRSIVS	2.8074	**4.6231**	3.8182	**7.8601**	**5.3594**	7.1452
	ESSIM	3.2167	6.4275	4.0931	8.8766	6.2484	**7.2656**
	+SR	3.1575	6.2670	4.2596	8.8396	5.8648	7.4699
	+SD	3.1803	6.5011	4.1296	8.8130	6.1287	7.3689
	+3DSD	**3.0909**	6.4090	**4.0244**	8.9932	6.2256	7.3762
	+QRSIVS	3.1104	**6.0247**	4.0953	**8.6479**	**5.7208**	7.5356

quality metrics cannot perform very well on LIVE-Phase-II. As introduced in Section 5.3, the distortions of different distortion types and levels also present masking properties of the HVS. Therefore, the asymmetric distortions of the stereoscopic image will inevitably affect the quality perception of HVS. However, the proposed QRSIVS as well as the related visual saliency map, such as SR, SD, and 3DSD, treat the left and right view image equally. That is also the main reason why the saliency map based metrics do not perform very well. Also, the FI-PSNR treats the left and right view equally, which together with PSNR provides an even worse performances, compared with other competitor quality metrics. However, for MJ3DQA, an intermediate image is constructed to have a perceived quality close to that of the cyclopean image. Therefore, the different behaviors of the left and right view images can somewhat be captured, which thereby gives the best performances on LIVE-Phase-II. In the future, we will also consider different behaviors of different view images. Also the different distortions in each view image will be considered to be incorporated into the design of the saliency map, especially for the quality assessment.

Furthermore, we merge the three IQDs together to further test the generality of the proposed QRSIVS based quality metric. From Table 4, it can be observed that the proposed QRSIVS based can achieve the best performances on the merged dataset, compared with 3D quality metrics, such as FI-PSNR and MJ3DQA, and other saliency map based quality metrics. In this case, our proposed QRSIVS are more generally effective to evaluate the stereoscopic images with different distortion types and levels.



(a) SSIM (b) MS-SSIM (c) ESSIM

(d) SSIM+QRSIVS (e) MS-SSIM+QRSIVS (f) ESSIM+QRSIVS

Fig. 6. Scatter plots of subjective *DMOS* vs. predicted $DMOS_p$ of SIQA metrics on the Ningbo-Phase-II database.

Table 4
Performance of the SIQA metrics on the merged dataset in terms of CC, SROCC, and RMSE.

Metric	CC	SROCC	RMSE
PSNR	0.556	0.5368	17.4764
FI-PSNR	0.5439	0.5298	17.6439
MJ3DQA	0.5639	0.5497	17.3646
SSIM	0.5622	0.5371	17.3889
+SR	0.588	0.5662	17.0071
+SD	0.5691	0.5448	17.2891
+3DSD	0.5611	0.5365	17.4053
+QRSIVS	**0.5948**	**0.5749**	**16.9033**
MS-SSIM	0.6322	0.614	16.2912
+SR	0.6311	0.6149	16.3108
+SD	0.6296	0.6128	16.3362
+3DSD	0.6247	0.6078	16.4196
+QRSIVS	**0.6336**	**0.6166**	**16.2673**
ESSIM	0.5847	0.5689	17.0573
+SR	0.5896	0.5737	16.983
+SD	0.5859	0.5712	17.0396
+3DSD	0.5811	0.5662	17.1126
+QRSIVS	**0.5971**	**0.5815**	**16.8664**

6. Conclusion

Stereoscopic image visual saliency map is an effective tool to improve the prediction performance of SIQA metrics. In this paper, we propose a QR based stereoscopic image visual saliency map detection model. The detected stereoscopic image visual saliency map is further incorporated into the SIQA framework. Experimental results show that our proposed QRSIVS based SIQA metric is powerful for predicting the 3D visual quality of stereoscopic images.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grants 61501299, 61672443, 61702336 and 61620106008, in part by the Guangdong Nature Science Foundation under Grant 2016A030310058, in part by Hong Kong RGC General Research Fund (GRF) 9042322 (CityU 11200116) and 9042489 (CityU 11206317), in part by the Shenzhen Emerging Industries of the Strategic Basic Research Project under Grants JCYJ20160226191842793 and JCYJ20170302154254147, in part by Natural Science Foundation of SZU (grant no. 2017031), in part by the Project 2016049 supported by SZU R/D Fund, in part by the Tencent "Rhinoceros Birds"-Scientific Research Foundation for Young Teachers of Shenzhen University, and in part by a grant from the Shenzhen Research Institute, City University of Hong Kong.

References

[1] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, Q. Dai, Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties, IEEE Trans. Image Process. 24 (10) (2015) 2971–2983.
[2] F. Shao, W. Lin, S. Wang, G. Jiang, M. Yu, Q. Dai, Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images, IEEE Trans. Cybern. 46 (3) (2016) 730–743.
[3] L. Ma, X. Wang, Q. Liu, K.-N. Ngan, Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment, Neurocomputing 215 (2016) 21–31.
[4] X. Wang, Q. Liu, R. Wang, Z. Chen, Natural image statistics based 3D reduced reference image quality assessment in contourlet domain, Neurocomputing 151 (2015) 683–691.
[5] C. Hewage, M. Martini, Quality of experience for 3D video streaming, IEEE Commun. Mag. 51 (5) (2013) 101–107.
[6] X. Wang, S. Kwong, H. Yuan, Y. Zhang, Z. Pan, View synthesis distortion model based frame level rate control optimization for multiview depth video coding, Signal Process. 112 (2015) 189–198.
[7] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, Z. Guo, Objective quality assessment of screen content images by uncertainty weighting, IEEE Trans. Image Process. 26 (4) (2017) 2016–2027.
[8] W. Lin, L. Dong, P. Xue, Visual distortion gauge based on discrimination of noticeable contrat changes, IEEE Trans. Circuits Syst. Video Technol. 15 (7) (2005) 900–909.
[9] L. Ma, K.N. Ngan, F. Zhang, S. Li, Adaptive block-size transform based just-noticeable difference model for images/videos, Signal Process. 26 (3) (2011) 162–174.
[10] L. Ma, S. Li, K.N. Ngan, Visual horizontal effect for image quality assessment, IEEE Signal Process. Lett. 17 (7) (2010) 627–630.

[11] Y. Fang, Z. Chen, W. Lin, C.W. Lin, Saliency detection in the compressed domain for adaptive image retargeting, IEEE Trans. Image Process. 21 (9) (2012) 3888–3901.
[12] Y. Fang, J. Wang, M. Narwaria, P.L. Callet, W. Lin, Saliency detection for stereoscpic images, IEEE Trans. Image Process. 23 (6) (2014) 2625–2636.
[13] Y. Fang, C. Zhang, J. Li, J. Lei, M.P.D. Silva, P.L. Callet, Visual attention modeling for stereoscopic video: a benchmark and computational model, IEEE Trans. Image Process. 26 (10) (2017) 4684–4696.
[14] L. Itti, C. Koch, E. Niebur, et al., A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[15] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Advances in Neural Information Processing Systems 19, MIT Press, 2007, pp. 545–552.
[16] J.K. Tsotsos, N.D.B. Bruce, Saliency based on information maximization, in: Advances in Neural Information Processing Systems 18, MIT Press, 2006, pp. 155–162.
[17] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[18] C. Guo, Q. Ma, L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[19] S. Daly, R. Held, D. Hoffman, Perceptual issues in stereoscopic signal processing, IEEE Trans. Broadcast. 57 (2 PART 2) (2011) 347–361.
[20] X. Wang, M. Yu, Y. Yang, G. Jiang, Research on subjective stereoscopic image quality assessment, in: Proc. SPIE, 7255, 2009.
[21] N. Bruce, J. Tsotsos, An attentional framework for stereo vision, in: IEEE Canadian Conference on Computer Robotics Vision, 2005.
[22] Y. Zhang, G. Jiang, M. Yu, K. Chen, Stereoscopic visual attention model for 3D video, in: International Conference on Advances in Mutimedia Model, 2010.
[23] C. Chamaret, S. Godeffroy, P. Lopez, O.L. Meur, Adaptive 3D rendering based on region-of-interest, SPIE Stereoscopic Displays and Applications, 2010.
[24] N. Ouerhani, H. Hugli, Computing visual attention from scene depth, in: International Conference on Pattern Recognition, 2000.
[25] E. Potapova, M. Zillich, M. Vincze, Learning what matters: combining probalilistic models of 2D and 3D saliency cues, International Computer Vision Systems, 2011.
[26] C. Lang, T.V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, S. Yan, Depth matters: influence of depth cues on visual saliency, in: European Conference on Computer Vision, 2012, pp. 101–115.
[27] Y. Niu, Y. Geng, X. Li, F. Liu, Leveraging stereopsis for saliency analysis, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 454–461.
[28] A. Ciptadi, T. Hermans, J.M. Rehg, An in depth view of saliency, in: British Machine Vision Conference, 2013.
[29] J. Wang, M.P.D. Silva, P.L. Callet, V. Ricordel, Computational model of stereoscopic 3D visual saliency, IEEE Trans. Image Process. 22 (6) (2013) 2151–2165.
[30] A. Benoit, P. Callet, P. Campisi, R. Cousseau, Using disparity for quality assessment of stereoscopic images, in: Proceedings - International Conference on Image Processing, ICIP, 2008, pp. 389–392.
[31] J. You, L. Xing, A. Perkis, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Jan, 2010.
[32] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G. Bozdagi Akar, Towards compound stereo-video quality metric: a specific encoder-based framework, in: Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation, 2006, pp. 218–222.
[33] X. Wang, S. Kwong, Y. Zhang, Considering binocular spatial sensitivity in stereoscopic image quality assessment, 2011 IEEE Visual Communications and Image Processing, VCIP 2011, 2011.
[34] 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) (2011) 19–22.
[35] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. Cormack, A. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process. 28 (9) (2013) 1143–1155.
[36] M.-J. Chen, L. Cormack, A. Bovik, No-reference quality assessment of natural stereopairs, IEEE Trans. Image Process. 22 (9) (2013) 3379–3391.
[37] J. Wang, A. Rehman, K. Zeng, S. Wang, Z. Wang, Quality prediction of asymmetrically distorted stereoscopic 3D images, IEEE Trans. Image Process. 24 (11) (2015) 3400–3414.
[38] A. Ogale, Y. Aloimonos, A roadmap to the integration of early visual modules, Int. J. Comput. Vision 72 (1) (2007) 9–25.
[39] T.A. Ell, S.J. Sangwine, Hypercomplex fourier transforms of color images, IEEE Trans. Image Process. 16 (1) (2007) 22–35.
[40] L. Ma, S. Li, K.N. Ngan, Motion trajectory based visual saliency for video quality assessment, in: International Conference on Image Processing, 2011.
[41] A.K. Moorthy, C.-C. Su, A. Bovik, Subjective evaluation of stereoscopic image quality, Signal Process. 28 (8) (2012) 870–883.
[42] J. Zhou, G. Jiang, X. Mao, M. Yu, F. Shao, Z. Peng, Y. Zhang, Subjective quality analyses of stereoscopic images in 3DTV system, in: Proceedings of the IEEE Visual Communications and Image Processing, 2011, pp. 1–4.
[43] Y.-H. Lin, J.-L. Wu, Quality assessment of stereoscopic 3D image compression by binocular integration behaviors, IEEE Trans. Image Process. 23 (4) (2014) 1527–1542.
[44] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, in: IEEE Asilomar Conference on Signals, Systems and Computers, 2003, pp. 1–4.
[45] X. Zhang, X. Feng, W. Wang, W. Xue, Edge strength similarity for image quality assessment, IEEE Signal Process. Lett. 20 (4) (2013) 319–322.