

# Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment



Lin Ma<sup>a</sup>, Xu Wang<sup>b,c,\*</sup>, Qiong Liu<sup>d</sup>, King Ngai Ngan<sup>e</sup>

<sup>a</sup> Huawei Noah's Ark Lab, Hong Kong

<sup>b</sup> College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

<sup>c</sup> Nanjing University of Information Science & Technology, Nanjing 210044, China

<sup>d</sup> Department of Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China

<sup>e</sup> Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong

## ARTICLE INFO

### Article history:

Received 28 January 2015

Received in revised form

17 June 2015

Accepted 20 June 2015

Available online 10 June 2016

### Keywords:

Stereoscopic image quality assessment (SIQA)

Reduced reference (RR)

Reorganized discrete cosine transform (RDCT)

Human visual system (HVS)

## ABSTRACT

In this paper, a novel reduced reference (RR) stereoscopic image quality assessment (SIQA) is proposed by characterizing the statistical properties of the stereoscopic image in the reorganized discrete cosine transform (RDCT) domain. Firstly, the difference image between the left and right view images is computed. Afterwards, the left and right view images, as well as the difference image, are decomposed by block-based discrete cosine transform (DCT). The DCT coefficients are further reorganized into a three-level coefficient tree, resulting in ten RDCT subbands. For each RDCT subband, the statistical property of the coefficient distribution is modeled by the generalized Gaussian density (GGD) function. And the mutual information (MI) and energy distribution ratio (EDR) are employed to depict the statistical properties across different RDCT subbands. Moreover, EDR can further model the mutual masking property of the human visual system (HVS). By considering the GGD modeling behavior within each RDCT subband and MI together EDR characterizing behavior across RDCT subbands, the statistical properties of the stereoscopic image are fully exploited, including the left view, right view, and the difference image. Experimental results demonstrated that the statistical properties of the difference image can well represent the perceptual quality of the stereoscopic image, which outperforms the representative RR quality metrics for stereoscopic image and even some full reference (FR) quality metrics. By considering the left view, right view, and difference image together, the performances of the proposed RR SIQA can be further improved, which presenting a more closely relationship between the quality metric output and human visual perception.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Image perceptual quality assessment plays the essential role in the image processing and communication [1,2], such as image capturing, compression, storage, transmission, displaying, printing, sharing, and so on. Therefore, there are many research works aim to develop image quality metrics for guiding the performance optimization during each step of image processing [3,4] and communication. Human eyes are the ultimate receivers of the images. The subjective test process is regarded as the most reliable way to evaluate the perceptual quality of the image. However, the subjective test process is time-consuming, which is impractical for the optimization process of the online image processing. Therefore, the

objective quality metrics that can automatically evaluate the image perceptual quality and guide the image processing applications are demanded.

Nowadays, with the rapid developments 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 have brought new entertainments and more vivid experiences to the consumers, which attract more and more attentions from not only researchers but also the industries. For these applications, the quality of 3D content is the most critical part to provide the visual QoE. However, in the 3D processing chain including capturing, processing [5–7], coding [8,9], transmitting, reconstruction, retrieving, etc., artifacts are inevitably introduced due to the resource shortage in processing [10]. Therefore, how to automatically evaluate the perceptual quality of 3D content [11] becomes a challenging issue in 3D visual signal processing. Moreover, it is claimed that the artifacts of 3D content

\* Corresponding author at: College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China.

E-mail addresses: [forest.linma@gmail.com](mailto:forest.linma@gmail.com) (L. Ma), [wangxu.cise@gmail.com](mailto:wangxu.cise@gmail.com) (X. Wang), [q.liu@hust.edu.cn](mailto:q.liu@hust.edu.cn) (Q. Liu), [knngan@ee.cuhk.edu.hk](mailto:knngan@ee.cuhk.edu.hk) (K.N. Ngan).

affect more on human visual system (HVS) [12,13], compared with the conventional 2D contents. Therefore, the realization of the HVS properties on 3D content is researched to help more accurately evaluate perceptual quality of the 3D contents.

According to the availability of the reference image, the conventional 2D image quality assessment (IQA) methods can be divided into three categories, specifically the full reference (FR) [14–16], no reference (NR) [17–22], and reduced reference (RR) [22–28], respectively.

FR metrics require the full assess of the original image to evaluate the perceptual quality of the distorted image. The original image is assumed to be artifact free and of perfect quality. Such metrics can be employed to guide the perceptual quality optimization during image/video compression, watermarking, and so on. The most appealing quality metrics are the mean square error (MSE) and its related peak signal-to-noise ratio (PSNR), because of their simplicity, clear physical meaning, and easy optimization. However, MSE and PSNR do not correlate with HVS properties. Therefore, many FR metrics are developed to incorporate the HVS properties and image signal properties. Wang et al. developed the most popular image quality metric structural similarity (SSIM) [14] that captures the structure information loss to depict the perceptual quality of the distorted image. A wavelet-based visual signal-to-noise ratio (VSNR) is developed to capture the distortions in wavelet domain [15]. A simple quality metric considering texture masking and contrast sensitivity function is developed for perceptual image coding [29]. In [16], Ma et al. proposed to incorporate the horizontal effect of HVS into SSIM for a better image quality metric.

However, in real-world applications, we are not able to access the original image for quality analysis in most cases, where only the distorted image is available. The NR quality metrics are thus employed. Many researchers employ the behaviors of specific distortions for the NR quality assessment, such as the blocking artifact of JPEG coded images, ringing artifact of the JPEG 2000 coded images, and so on. As JPEG 2000 employs the wavelet transform to compress the image, the wavelet statistical model is utilized to capture the compression distortion [18]. Liang et al. [19] combined the sharpness, blurring, and ringing measurements together to depict the perceptual quality of the JPEG 2000 coded image. The distribution of the DCT coefficient after quantization is modeled in [20] to predict the PSNR value of the JPEG coded image. Furthermore, Ferzli et al. [21] did the psychophysical experiment to test the blurring tolerance ability of the HVS, based on which the just-noticeable blur (JNB) model is developed. These methods employ the behaviors of specific distortions to predict the degradation level. Therefore, if a new distortion is introduced, these methods can hardly evaluate the perceptual quality of the distorted image. In order to compromise between the FR and NR IQAs, RR IQAs are developed. It is expected that the RR methods can effectively evaluate the image perceptual quality based on a limited number of RR features extracted from the reference image. Only a small number of bits are required for representing the extracted features, which can be efficiently encoded and transmitted for the quality analysis. Consequently, it will be very useful for the quality monitoring during the image transmission and communication.

For RR quality metrics, only partial information of the original image is available for quality analysis, which can be further categorized into the following three classes.

- *Distortion-based RR metrics:* The behaviors of the distortions are modeled for RR quality metric design. Wolf et al. [30,31] proposed to extract a set of spatial and temporal features for measuring the distortions introduced in the standard video compression and communication environment. The features that are associated with the blurring, blocking, and frame differences are extracted in [32] to depict the compression artifacts introduced by MPEG-2. These RR quality metrics are

designed for some specific distortions, which cannot be effectively applied to the other images of different distortions. Therefore, a general RR IQA for evaluating the image perceptual quality of different distortions is required.

- *HVS-based RR metrics:* HVS properties should be considered for quality assessment, as the human eyes are the ultimate receivers. Le Callet et al. [33] employed a neural network to train and evaluate the perceptual qualities of video sequences based on the perceptual-related features extracted from the video frames. In [34,35], the perceptual features motivated from the computational models of low level vision are extracted as the reduced descriptors to represent the image perceptual quality. The merits from the contourlet transform, the contrast sensitivity function, and Webers law of just noticeable difference are incorporated to derive an RR IQA [36], which are employed for evaluating the perceptual qualities of the JPEG and JPEG 2000 coded images. Recently, an RR IQA [37] for wireless imaging is developed by considering different structural information that is observed in the distortion model of wireless link.
- *Statistics-based RR metrics:* The statistical modeling of the image signal has been investigated for the image perceptual quality assessment for RR IQAs. In [38,23], Wang et al. proposed a wavelet-domain natural image statistic metric (WNISM), which models the marginal probability distribution of the wavelet coefficients of a natural image by the generalized Gaussian density (GGD) function. The Kullback–Leibler distance (KLD) is used to depict the distribution difference. To improve the performance and reduce the number of features, the probability distribution was represented by the Gaussian scale mixture (GSM) model in wavelet domain [39]. In [24–26], the RR features are extracted in the reorganized domain. In [24], the DCT coefficients are first reorganized into several representative subbands, whose distributions are modeled by the GGD. The city-block distance (CBD) is utilized to capture the image perceptual quality. In [40], the statistics of image gradient magnitude are modeled by the Weibull distribution to develop an RR image quality metric. Also the statistics of the edge [41] are utilized for developing the RR IQA. In [42], the authors measure the differences between the entropies of wavelet coefficients of the reference and distorted image to quantify the image information change, which can indicate the image perceptual quality.

In this paper, we proposed a novel RR SIQA based on the statistical modeling of the stereoscopic image. The difference image is computed by referring to the left view and right view image. After performing the block-based DCT and reorganization process, the coefficients of the images (left view, right view, and the difference image) are reorganized into different RDCT subbands. The statistical property within each RDCT subband is exploited by the GGD modeling of the coefficient distribution. The statistical property across RDCT subbands is modeled by the energy distribution ratio (EDR), which can be further employed for modeling the HVS mutual masking property. The main contributions of our proposed method are listed as follows.

- The statistical property of stereoscopic image is studied for perceptual quality analysis. The statistical properties of the obtained difference image are firstly investigated in the RDCT domain. By considering the difference image, the left and right view images are considered together for perceptual quality analysis, which matches the HVS perception of the stereoscopic image.
- The statistical properties of the difference image are characterized from two perspectives, specifically the within and across RDCT subband statistical properties, respectively. The statistical properties depicted within and across the RDCT

subbands can help capture the distortions introduced to the stereoscopic image.

- The difference image is demonstrated to be most important for SIQA, compared to the sole left view and right view image. Therefore, the proposed RR SIQA metric can be viewed as a scalable approach. If only a few RR features are constrained, we can only extract the RR features from the difference image for quality analysis. Otherwise, all the RR features are extracted for quality analysis from the images, including the left view, right view, and difference image.

The rest of this paper is organized as follows. Section 2 overviews the related work. Section 3 illustrates the proposed RR SIQA model. Experimental results are provided in Section 4. Finally, conclusions are given in Section 5.

## 2. Related work

As 3D images are prevailing, the quality evaluation of 3D images is demanded for the quality control and optimization for better QoE of human viewers. Existing perceptual quality metrics for 3D images can be roughly divided into two categories, named quality model extended from 2D image quality assessment (IQA) metrics and the quality model specifically designed for 3D images.

The first category simply extends the 2D IQA models to the stereoscopic image. Such models treat each view image of the stereoscopic image independently. By fusing the obtained quality index of each view image, the quality value of the stereoscopic image is determined. In [43], Hewage et al. investigated the correlation between subjective quality scores and three quality metrics, including PSNR, Video Quality Model (VQM) [44], and SSIM [14] for 3D video content. It is demonstrated that the VQM metric performs better than the other two metrics for predicting the overall perceptual quality of the stereoscopic image. The quality metric for 3D image in [45] is also extended from 2D IQAs. These extended metrics treat each view image independently, which does not consider the binocular vision properties. However, such binocular vision properties are critical for the 3D image perception, which make the extended quality metrics generate unsatisfactory performances.

The other category considers the specific HVS properties of viewing 3D images, such as the binocular vision and depth perception properties. In [46], Benoit et al. fused the depth (or disparity) information and 2D quality metrics to evaluate the 3D image quality. Ref. [47] investigated the integration of disparity information into 2D quality metrics. Boev et al. [48] combined the monoscopic quality component and the stereoscopic quality component to develop a stereo-video quality metric, where the cyclopean image concept was first introduced by fusing the left and right views. To further improve the performance of 3D IQA, the binocular fusion and rivalry properties are investigated. The binocular spatial sensitivity (BSS) weighted metric is developed [49] based on the binocular just noticeable difference model [50]. Chen et al. [51] developed a framework for assessing the quality of stereoscopic images that have been afflicted by possibly asymmetric distortions. The linear rivalry model was employed to exploit the binocular rivalry property in [52]. However, the features employed in the binocular vision properties based metric are local, which may not work well when the original 3D images are inaccessible.

For 3D RR IQA metrics, image features including the distortion driven features such as binary edge mask [53,54], and HVS features such as contrast sensitivity index [55], are extracted and compared for quality analysis, which correlate with HVS perception and image statistical properties. The marginal distributions of disparity [56] subband coefficients can be well fitted by the GGD model, which is similar to the luminance images. This property was employed as prior

information to improve the performance of Bayesian disparity estimation [57]. In [27], Wang et al. extracted the RR features in the contourlet domain for 3D image quality assessment. After receiving the RR features, the RR metrics evaluate the perceptual quality of the distorted images by referring to the corresponding extracted RR features.

## 3. Proposed reduced reference (RR) stereoscopic image quality assessment (SIQA)

### 3.1. Statistical properties of the stereoscopic image

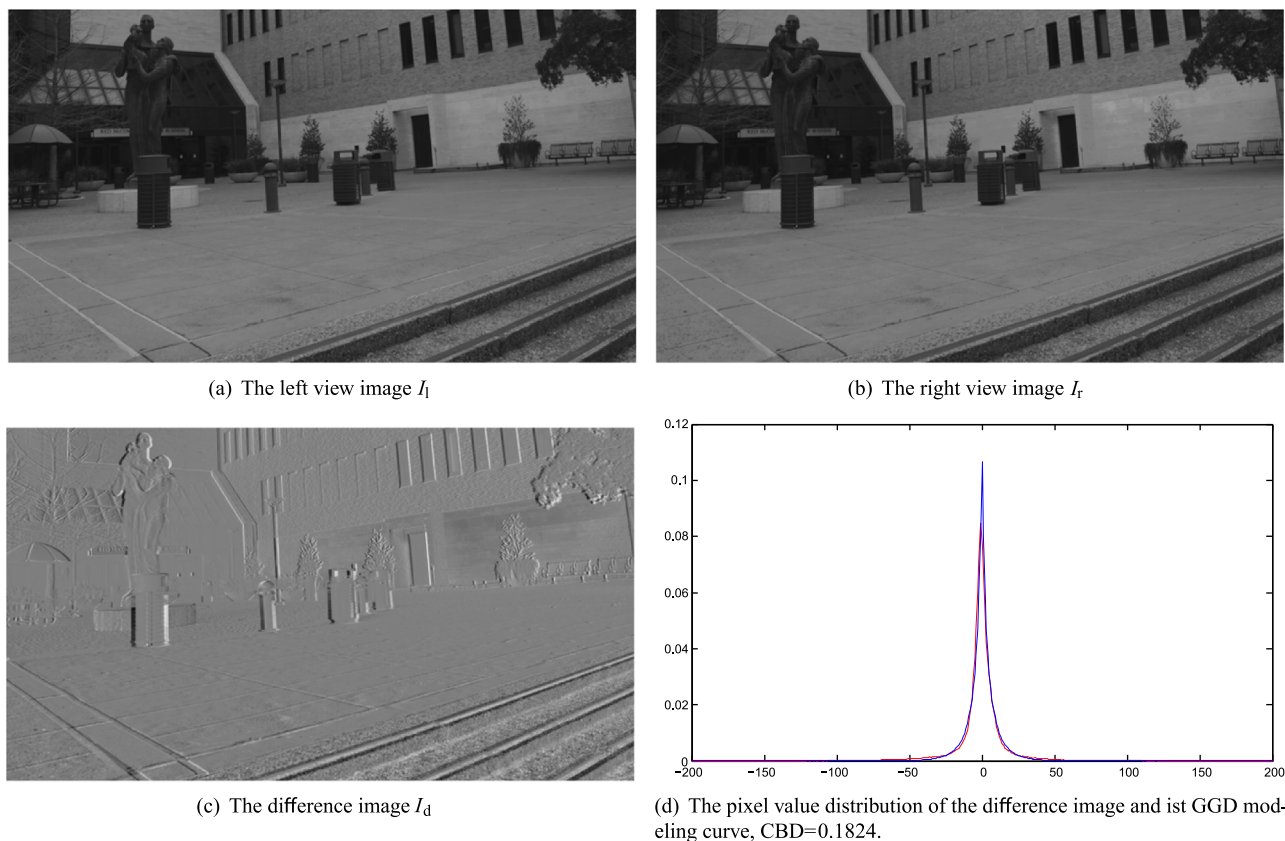
The statistical properties of conventional 2D images have been extensively studied, such as the GGD modeling of the wavelet coefficient in [23], the GSM modeling in wavelet domain [39], GGD modeling in RDCT domain [24]. However, there are no literatures discussing the statistical modeling of the stereoscopic image, specifically jointly considering the left view and right view image together. As mentioned before, HVS perception properties of the stereoscopic images significantly affect the quality analysis of the stereoscopic image. It is expected that an accurate modeling of the statistical properties of the stereoscopic image can help to capture the quality degradation. In this section, we focus on the statistical modeling of the stereoscopic image by jointly considering the left view and right view image.

The correlation between the left view and right view image can be exploited by performing the disparity map estimation between the two images. The disparity map depicts the pixel replacement of one view image referring to the other image. By using the disparity map, we can obtain the view difference image between the two view images. Such difference image introduces the view disparities to the human eyes, which are critical to generate the 3D experiences of the stereoscopic image. Therefore, the statistical properties of such difference image are investigated in this section. However, as we target at an RR SIQA, we need to consider the RR feature extraction in the receiver side, specifically the statistical feature extraction. The features in the receiver side should be easy for computation and extraction following the same procedure as the sender side. Therefore, we cannot perform the disparity map estimation in the sender and receiver side. Firstly, the computational complexity is high. Secondly the disparity map estimated in the reference and distorted stereoscopic images will be different. Therefore, we employ a straightforward and simple way to generate the difference image by subtracting one view image from the other view image. The statistical properties of such difference image are studied. And the experimental results in Section 4 will illustrate the performance improvements by employing the difference image, not only for our proposed RR SIQA, but also for the other RR SIQAs as well as the FR SIQAs.

Supposing the left view image is  $I_l$  and the right view image is  $I_r$ , the difference image  $I_d$  is obtained by:

$$I_d = I_r - I_l. \quad (1)$$

$I_l - I_r$  can also be employed to obtain  $I_d$ , which makes no differences to the statistical modeling. The corresponding images of  $I_l$ ,  $I_r$ , and  $I_d$  are illustrated in Fig. 1. Only the luminance component of the image is illustrated, which mostly affects the HVS perception. The difference image is processed as  $I_d + 128$  for better visualization. The pixel value distribution of the difference image is illustrated in Fig. 1(d), where the red line denotes the pixel value distribution of the difference image. It can be observed that the pixel value distribution presents a high kurtosis distribution (with a sharp peak at zero and a fat-tail distribution). As discussed in [24,13], the high kurtosis distributions can be well modeled by GGD, which is defined in:



**Fig. 1.** The stereoscopic image and the generated difference image. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.

$$P_{\alpha,\beta}(x) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x|}{\alpha}\right)^\beta} \quad (2)$$

where  $\beta > 0$  and  $\alpha$  are the two parameters of the GGD function.  $\Gamma$  is the Gamma Function denoted by:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \quad (3)$$

$\alpha$  models the width of the peak (standard deviation), while  $\beta$  is inversely proportional to the decreasing rate of the peak.  $\alpha$  and  $\beta$  also refer to the scale and shape parameters, respectively. The blue line indicates the GGD modeling of the difference image coefficient distribution. It can be observed that the blue and red lines overlap with each other, which demonstrates an accurate GGD modeling of the coefficient distribution. The city-block distance (CBD) [24,26] is employed to capture the distribution difference, which is defined as:

$$d_{CBD}(p, q) = \sum_{i=1}^{h_L} |p(i) - q(i)| \quad (4)$$

where  $p$  and  $q$  are two coefficient distributions, and  $h_L$  denotes the number of histogram bins.

As illustrated in [24,26], the coefficient distribution of the natural image in RDCT domain can be well modeled by GGD. In this paper, we make the first attempt to employ the RDCT decomposition to process the difference image between left and right view images. We investigate whether the RDCT coefficient distribution of the difference image can be accurately modeled by GGD and whether such modeling features can be employed as the stereoscopic image features, which can further help for stereoscopic image quality assessment.

Firstly, the difference image as in Fig. 1(c) is obtained. Afterwards,

the block-based DCT is employed to transform the image pixel values to the DCT domain. Specifically, the  $8 \times 8$  block-based DCT is employed. The obtained image after performing  $8 \times 8$  DCT is illustrated in Fig. 2(a). After obtaining the DCT coefficients, the reorganization process as introduced in [24] is applied, which generate the corresponding RDCT representation as illustrated in Fig. 2(b). For better visualization, the coefficients of the DCT and RDCT representations are rescaled in the range of [0,255]. For the reorganization process, the  $8 \times 8$  DCT coefficients are firstly decomposed into ten groups. Then the coefficients of all the  $8 \times 8$  DCT blocks are grouped and organized together according to their positions. The reorganization process results in a three level coefficient tree with ten RDCT subbands  $S_0, \dots, S_9$ , as shown in Fig. 2(b). The RDCT representation presents a similar appearance like a wavelet representation, which exhibits structural similarity between subbands and coefficient magnitude decaying toward high frequency subbands.

As discussed in [26], the RDCT coefficient distribution of the natural image can be well depicted by GGD. In this paper, we will investigate how the GGD function models the coefficient distribution of the difference image and its RDCT subbands. As illustrated in Fig. 1(d), GGD is demonstrated to accurately model the coefficient distribution of the difference image with the CBD value equals to 0.1824. The GGD modeling of the RDCT subband coefficient distribution is illustrated in Fig. 3. It can be observed that the coefficient distribution of the RDCT subband can be more accurately modeled by GGD, compared with the difference image in Fig. 1(d). We also employ the CBD as defined in Eq. (4) to evaluate the GGD modeling accuracy, which is also illustrated in Fig. 3. It can be observed that only two subbands  $S_1$  and  $S_2$  generate larger CBD values than the CBD value of the difference image. And the average CBD value over all the 9 subbands is 0.129322, which is much smaller than the CBD value of the difference image. It means that the GGD is

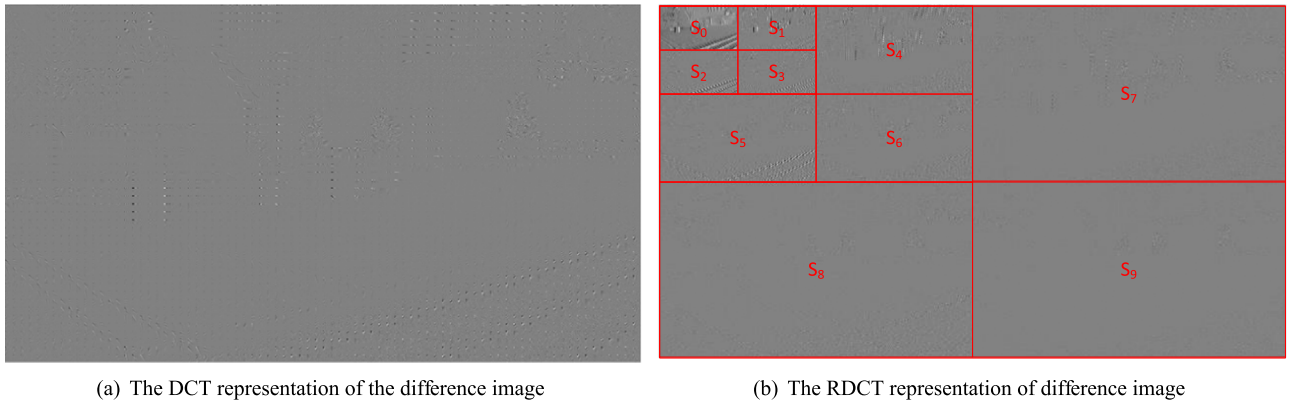


Fig. 2. The DCT and RDCT representation of the difference image.

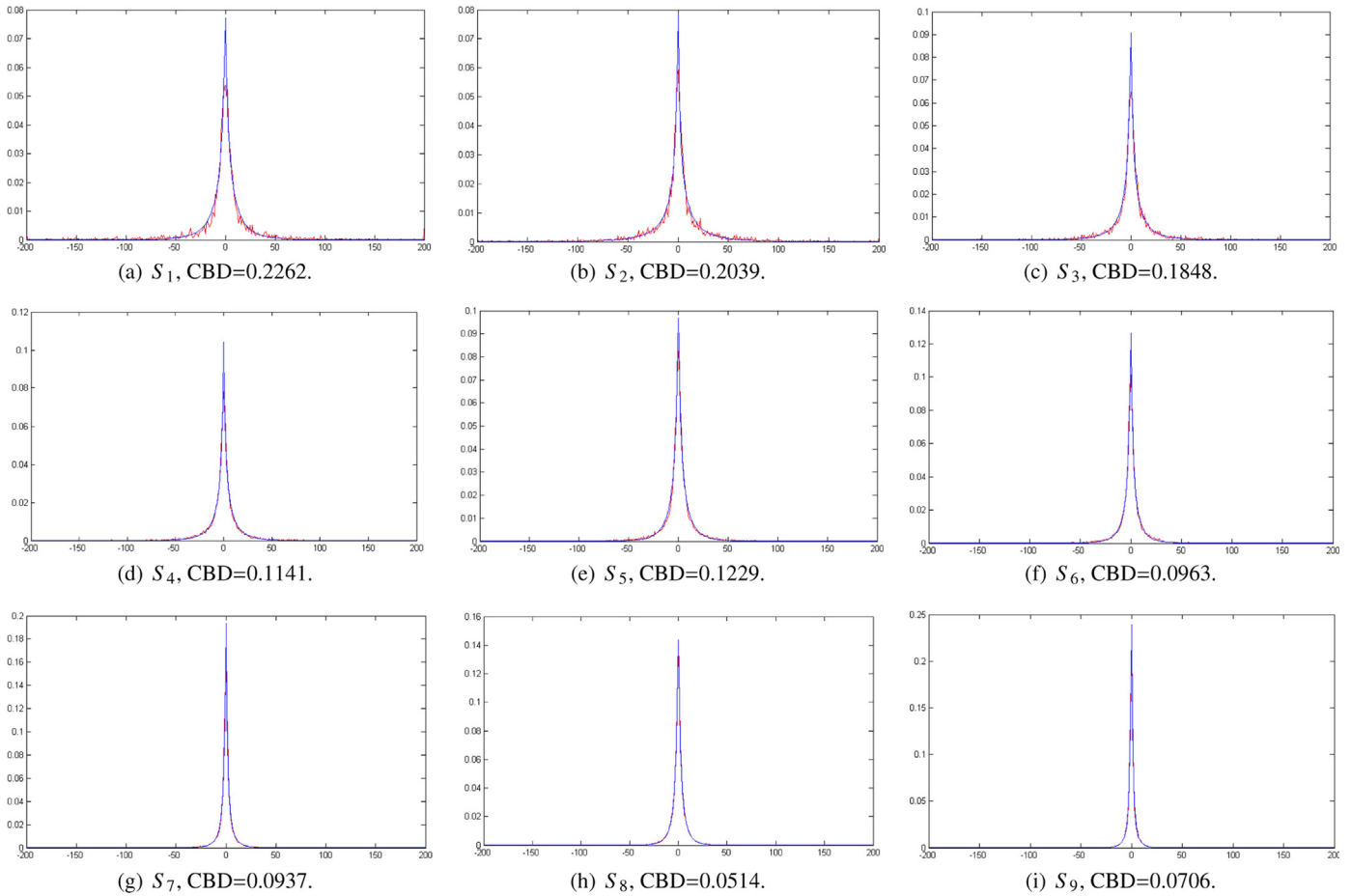


Fig. 3. The coefficient distribution of each RDCT subband and its GGD modeling curve.

more suitable to model the RDCT subband coefficient distribution of the different image.

Therefore, in this paper, other than directly modeling the coefficient distribution of the difference image, the coefficient distributions of the RDCT subbands are modeled to extract the corresponding RR features for quality analysis. As the difference image is considered, we explicitly jointly consider the left and right view image, which matches the HVS perception of the stereoscopic image. The prior RR IQAs for stereoscopic image extract the RR features on single view image, which do not consider the

inter-view image statistical properties for stereoscopic image quality assessment.

### 3.2. Proposed reduced reference stereoscopic image quality assessment metric

In this paper, the features extraction and quality analysis schemes in [26] are employed to predict the perceptual quality of degraded stereoscopic image. Fig. 4 provides the framework of our proposed metric.

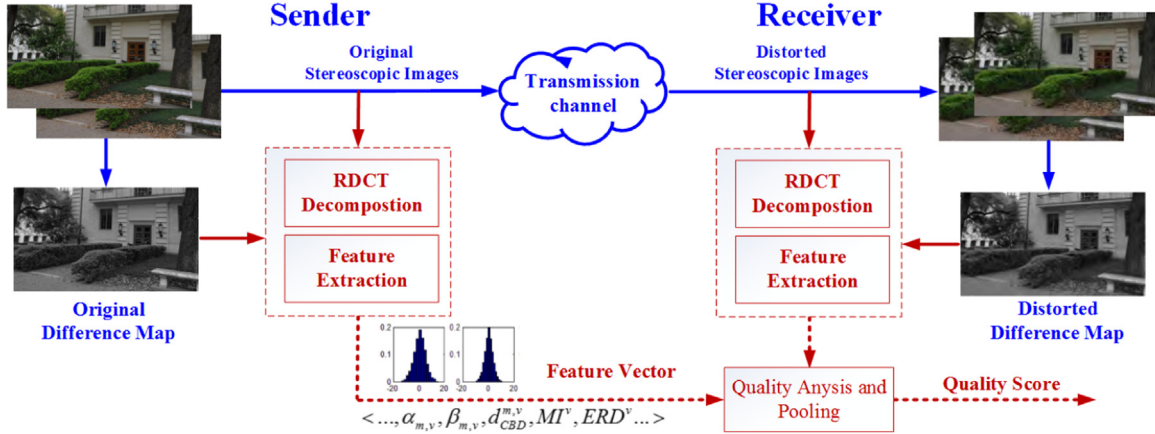


Fig. 4. The framework of the proposed SIQA metric.

### 3.2.1. Feature extraction

In the sender side, the left/right/difference images of original stereoscopic image are decomposed and reorganized into RDCT representation as shown in Fig. 2. Afterwards, the following features are extracted from the RDCT subbands.

- **GGD model parameters:** For the three horizontal RDCT subbands,  $S_1, S_4$  and  $S_7$ , the distribution of subband coefficients is fitted by GGD model. Then, the model parameters  $\alpha, \beta$  and  $d_{CBD}$  are calculated based on Eqs. (2) and (4). Therefore, there are nine GGD model parameters in total extracted.
- **Mutual information (MI):** MI features can describe the essential relationship between the RDCT subbands, which are influenced by introduced distortion. In this paper, two MI values of the parent–child correlation between the subband pairs ( $S_1, S_4$ ) and ( $S_4, S_7$ ), three MI values of the cousin–child correlation between the subband pairs ( $S_1, S_2$ ), ( $S_4, S_5$ ) and ( $S_7, S_8$ ), and three MI values of the brother–child correlation between the subband pairs ( $S_1, S_3$ ), ( $S_4, S_6$ ) and ( $S_7, S_9$ ) are calculated as RR features. Thus, there are eight MI values in total extracted to represent the inter-subband dependencies.
- **Energy distribution ratio (EDR):** EDR is a global feature which can capture the frequency proportion changes caused by introduced distortions. The FRD features are defined as

$$EDR = \frac{M + H}{L}, \quad (5)$$

where  $L, M, H$  are the sums of absolute DCT coefficient values in the low frequency ( $S_0, S_1, S_2, S_3$ ), medium frequency ( $S_4, S_5, S_6$ ) and high frequency ( $S_7, S_8, S_9$ ) RDCT subbands, respectively.

For each image component (left/right/difference image), we employed three horizontal RDCT subbands for GGD modeling and CBD calculation. Each RDCT subband requires two GGD modeling parameters and one CBD parameter. Moreover, eight MI features are extracted to depict the relationship between horizontal RDCT subbands and other RDCT subbands. Additionally, one EDR feature is extracted to globally depict the frequency proportion of the image. Therefore, in total, we have  $3 \times 3 + 8 + 1 = 18$  RR features extracted for each image. For the method with only difference image, we have 18 RR features. For the method with left view, right view, and difference image, we have  $18 \times 3 = 54$  RR features.

### 3.2.2. Quality analysis and pooling

In the receiver side, the extracted features of the original and distorted stereoscopic image are compared to predict the stereoscopic image perceptual quality. First, the following three distance

indexes will be determined to measure the perceptual quality degradation caused by the introduced distortion.

- **CBD:** The distance between the RDCT subband  $m$  original and distorted stereoscopic image is defined as

$$d_{CBD}^{m,v}(p_o^{m,v}, p_d^{m,v}) = \sum_{i=1}^{h_l} |p_o^{m,v}(i) - p_d^{m,v}(i)|, \quad (6)$$

where  $p_o^{m,v}$  and  $p_d^{m,v}$  are the coefficient distributions of original and distorted stereoscopic images, and  $v \in \{\text{left image, right image, difference image}\}$ . Since  $p_o^{m,v}$  is not available in the receiver side, the distance can be approximated as

$$d_{CBD}^{m,v}(p_o^{m,v}, p_d^{m,v}) = d_{CBD}^{m,v}(p_o^{m,v}, p_{\alpha_{m,v}, \beta_{m,v}}^{m,v}) - d_{CBD}^{m,v}(p_{\alpha_{m,v}, \beta_{m,v}}^{m,v}, p_d^{m,v}), \quad (7)$$

where  $d_{CBD}^{m,v}(p_o^{m,v}, p_{\alpha_{m,v}, \beta_{m,v}}^{m,v})$  is the parameters from the sender side.  $p_{\alpha_{m,v}, \beta_{m,v}}^{m,v}$  is the estimated coefficient distribution of the original stereoscopic image based on the GGD model parameters  $\alpha_{m,v}$  and  $\beta_{m,v}$ .

- **Differences between the MI values:** The differences between the MI values of the original and distorted stereoscopic images are determined by

$$d_{MI}^v(S_{m,v}, S_{n,v}) = MI_o^v(S_{m,v}, S_{n,v}) - MI_d^v(S_{m,v}, S_{n,v}), \quad (8)$$

where  $MI_o^v(S_{m,v}, S_{n,v})$  is the MI value of the RDCT subband  $S_{m,v}$  and  $S_{n,v}$  in the original stereoscopic image, and  $MI_d^v(S_{m,v}, S_{n,v})$  is the MI value of the RDCT subband  $S_{m,v}$  and  $S_{n,v}$  in the distorted stereoscopic image.

- **Similarity between the EDR values:** To measure the influence of the introduced distortion on the HVS-related features. The similarity between the EDR values of the original and distorted stereoscopic images are defined as

$$d_{EDR}^v = \begin{cases} \frac{\xi^v}{\xi^v + EDR_o^v} & EDR_o^v < EDR_d^v \\ \frac{\xi^v}{\xi^v + EDR_d^v} & EDR_o^v \geq EDR_d^v, \end{cases} \quad (9)$$

where

$$\xi^v = |EDR_o^v - EDR_d^v|. \quad (10)$$

$EDR_o^v$  and  $EDR_d^v$  are the EDR values of the original and distorted stereoscopic images, respectively.  $\xi^v$  in the denominator is employed to scale  $d_{EDR}^v$  into the range [0,1]. With such scaling process, the mutual masking properties have been considered [58]. The mutual masking effect is determined by the minimum

value of the thresholds calculated from the original and distorted image. Therefore, the computed  $EDR$  value can depict the texture information of the image. The mutual masking effect of the HVS perception is modeled as Eq. (9), where the smaller value of  $EDR_o^v$  and  $EDR_d^v$  is employed to model the masking effect. In this way, only the image is highly textured in both the reference and distorted images (large  $EDR_o^v$  and  $EDR_d^v$  values) can produce a significant masking effect.

After obtaining the CBD values and the similarity index between the ERD values, the quality values for image component  $v$  is defined as

$$Q^v = c_1 \times \sum_m d_{CBD}^{m,v} + c_2 \times \sum_{(m,n)} d_{MI}^v(S_{m,v}, S_{n,v}) + c_3 \times d_{EDR}^v \quad (11)$$

In this paper, the parameters  $c_1$ ,  $c_2$  and  $c_3$  are set as 0.4883, 0.0313 and 0.6719, respectively. To scale the quality values, the quality values  $Q_v$  are transformed by a logarithm process:

$$Q_s^v = \log_{10} \left( 1 + \frac{Q^v}{D_0} \right) \quad (12)$$

$D_0$  is a scale factor to adjust the variation of  $Q^v$ . In this paper,  $D_0$  is set as 0.0001 for simplicity.

Furthermore, the final perceptual quality of the distorted stereoscopic image is determined as

$$Q_{fin} = \sum_v Q_s^v \quad (13)$$

#### 4. Experimental results

In this section, we implement the proposed RR-SIQA metric and make performance comparison 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 [59] where the left/right views of the stereoscopic image are symmetrically distorted. The other category is the asymmetric IQD [13] where the left/right views of the stereoscopic image are degraded with different distortion types and levels. In this paper, we evaluate the performance on both symmetric and asymmetric IQDs. The detailed information of the selected IQDs is described as follows:

- *LIVE 3D IQD Phase I (LIVE-Phase-I)* [60] 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  $640 \times 360$ . 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 images (80 each for JP2K, JPEG, WN and FF; 45 for GBLUR).
- *LIVE 3D IQD Phase II (LIVE-Phase-II)* [51] aims to build a diverse database that consists of both symmetrically and asymmetrically distorted stereoscopic pairs. The 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 comparison, both 2DIQA extension model and binocular vision inspired metric (denoted as SIQA model) are evaluated in the experiment. Two FR-SIQA models, including FI-PSNR [61] and MJ3DQA [51] are implemented in the experiment. To verify the effectiveness of proposed RR-SIQA metric, three different RR-SIQA metrics, including DNT [39], RDCT [26] and contourlet domain based model (denoted as CT [27]) are employed as the benchmark. We proposed two RR quality metrics, specifically the Proposed-I and Proposed-II. Proposed-I only performs the RR quality analysis of the difference image, while Proposed-II performs the RR quality analysis of left view, right view, and the different image.

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

$$DMOS_p = \frac{p_1}{2} - \frac{p_1}{1 + \exp(p_2 \cdot (q - p_3))} + p_4 \cdot q + p_5, \quad (14)$$

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 mapping 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). The scatter plots of subjective  $DMOS$  against the predicted  $DMOS_p$  of RR-SIQA metrics on the 3D IQDs are provided in Fig. 5. It is easy to observe that the our proposed RDCT-based RR-SIQA metrics can achieve the best performance on the entire database for both symmetric and asymmetric IQDs. Detailed experimental results are provided in Table 1.

The performances of different IQAs over different stereoscopic image quality databases are illustrated in Table 1 and Fig. 5. It can be observed that the proposed method can outperform the other RR and even some FR metrics on both LIVE-Phase-I and LIVE-Phase-II, with larger LCC/SROCC and smaller RMSE values. We compared some quality metrics extended from conventional 2D images, such as PSNR, SSIM, VSNR, and so on. From Table 1, experimental results demonstrate that PSNR performs badly, although it requires the whole reference image for perceptual quality analysis. The reason is that PSNR only measures the pixel absolute differences, which does not take the HVS property into consideration. For SSIM, the structural distortions are measured rather than the absolute pixel value differences, which are sensitive to the HVS perception. Therefore, SSIM demonstrates a better performance than PSNR. However, SSIM also requires the whole reference image for quality analysis. For MS-SSIM, a multi-scale SSIM is applied to the stereoscopic image, which generates a better performance than SSIM. However, compared the performance of SSIM and MS-SSIM on LIVE-Phase-I and LIVE-Phase-II, it can be observed that SSIM and MS-SSIM perform worse on LIVE-Phase-II. Therefore, SSIM and MS-SSIM are not able to handle the quality assessment of stereoscopic images with non-symmetric distortions. For other FR quality metrics extended from 2D images, such as VIF and UQI, as the reference stereoscopic image is available and complex HVS and image signal properties have been studied, the performances are much better for both the symmetric and non-symmetric distortions. For the FR quality metrics specifically designed for stereoscopic image, FI-PSNR and MJ3DQA, the performances are not good enough, even worse than the metrics extended from conventional 2D metrics. The reason may 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 stereoscopic images. Consequently, the performances of existing binocular vision inspired metrics are limited, such as FI-PSNR and MJ3DQA.

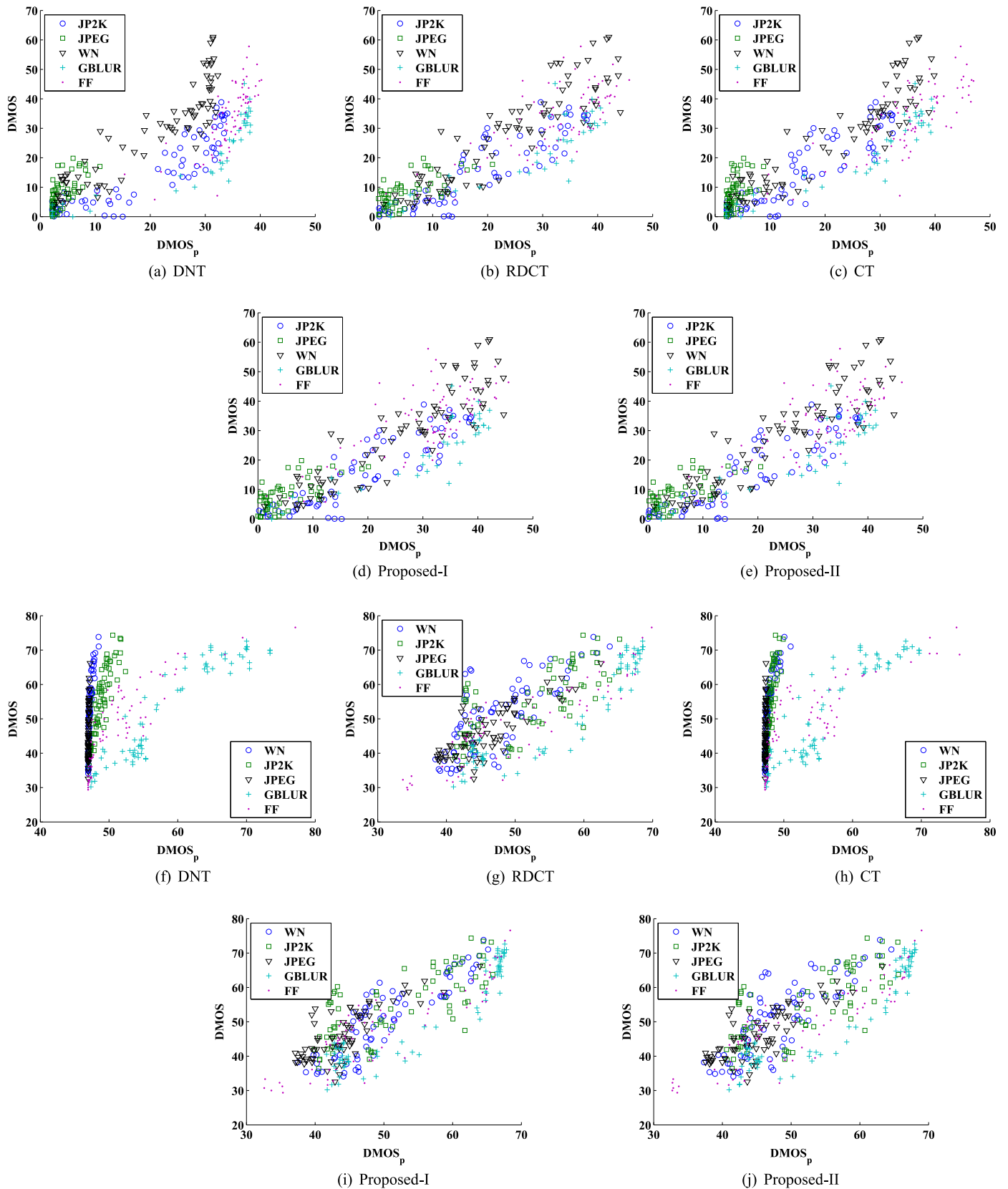


Fig. 5. Scatter plots of subjective  $DMOS$  vs. predicted  $DMOS_p$  of SIQA metrics on the 3D IQDs. (a)–(e) Results on the LIVE-Phase-I database. (f)–(j) Results on the LIVE-Phase-II database.

For the RR quality metrics, we compare our proposed method with DNT [39], RDCT [26], and CT [27], which are extended from the conventional 2D RR metrics. These RR metrics will perform the quality analysis for the left and right view image independently.

Then the average quality index is regarded as the perceptual quality of the stereoscopic image. RDCT tries to depict the coefficient distribution in the RDCT domain as the RR features. As shown in Fig. 2, RDCT presents similar signal decomposition



**Table 1**

Performance of the SIQA metrics for LIVE-Phase-I and LIVE-Phase-II databases in terms of CC, SROCC and RMSE.

Criteria	Type	Metric	LIVE-Phase-I [60]						LIVE-Phase-II [51]					
			JP2K	JPEG	WN	GBLUR	FF	ALL	JP2K	JPEG	WN	GBLUR	FF	ALL
CC	FR	PSNR	0.7878	0.1190	0.9352	0.7694	0.7107	0.8355	0.4726	0.6696	0.4161	0.9280	0.7743	0.7317
		SSIM	0.8753	0.4893	0.9421	0.9180	0.7238	0.8766	0.9301	0.7253	0.6659	0.8421	0.8669	0.8024
		MS-SSIM	0.9335	0.6663	0.9522	0.9449	0.8083	0.9297	0.9510	0.8389	0.8324	0.7995	0.8740	0.7938
		VSNR	0.8905	0.4118	0.9109	0.9015	0.7870	0.8904	0.6942	0.6567	0.4308	0.9710	0.8717	0.7768
		VIF	0.9381	0.6820	0.9318	0.9652	0.8616	0.9253	0.8265	0.8322	0.8237	0.9871	0.9466	0.8408
		UQI	0.9510	0.7737	0.9270	0.9568	0.8791	0.9418	0.8509	0.8345	0.7984	0.9803	0.9615	0.8639
		FI-PSNR [61]	0.8575	0.3266	0.9289	0.8191	0.7096	0.8733	0.9247	0.7751	0.6677	0.7384	0.7157	0.6450
	RR	MJ3DQA [51]	0.8615	0.2914	0.9502	0.9086	0.6783	0.8791	0.9612	0.8218	0.7467	0.9558	0.8888	0.8837
		DNT [39]	0.9380	0.6903	0.9231	0.9636	0.8589	0.8608	0.9393	0.8919	0.8626	0.9701	0.9085	0.5417
		RDCT [26]	0.9157	0.7228	0.9118	0.9204	0.8073	0.9042	0.7468	0.7547	0.8243	0.9694	0.8925	0.8004
		CT [27]	0.9252	0.5574	0.9196	0.9596	0.8339	0.8998	0.9524	0.8971	0.7738	0.9433	0.9030	0.5216
		DWT-I	0.9177	0.6592	0.8927	0.9284	0.7924	0.7061	0.8639	0.7633	0.7826	0.9752	0.9296	0.7161
		DWT-II	0.9187	0.6667	0.888	0.9226	0.7947	0.6629	0.7856	0.7559	0.7993	0.9614	0.9078	0.6743
		Proposed-I	0.9216	0.7179	0.9127	0.9311	0.7966	0.9033	0.8966	0.7478	0.8098	0.9690	0.9100	0.8431
Proposed-II	0.9182	0.7222	0.9131	0.9247	0.8068	0.9056	0.8094	0.7544	0.8220	0.9721	0.9016	0.8179		
SROCC	FR	PSNR	0.7993	0.1212	0.9316	0.9020	0.5875	0.8341	0.7098	0.6557	0.5234	0.8650	0.7660	0.7184
		SSIM	0.8584	0.4361	0.9379	0.8793	0.5861	0.8765	0.9206	0.7027	0.6789	0.8358	0.8348	0.7925
		MS-SSIM	0.8978	0.5985	0.9423	0.9282	0.7349	0.9225	0.9473	0.8172	0.8271	0.8010	0.8304	0.7719
		VSNR	0.8318	0.4063	0.9049	0.8306	0.7295	0.8818	0.6542	0.7577	0.6244	0.8952	0.8348	0.7355
		VIF	0.9022	0.5834	0.9325	0.9312	0.8037	0.9205	0.8196	0.8258	0.7782	0.9501	0.9330	0.8171
		UQI	0.9077	0.7374	0.9255	0.9274	0.8329	0.9373	0.8509	0.8345	0.7984	0.9803	0.9615	0.8639
		FI-PSNR [61]	0.8522	0.2568	0.9297	0.9394	0.6599	0.8644	0.9148	0.7437	0.6681	0.7088	0.6945	0.6456
	RR	MJ3DQA [51]	0.8199	0.2462	0.9327	0.9004	0.6446	0.8603	0.9427	0.8159	0.7583	0.8901	0.8497	0.8775
		DNT [39]	0.9029	0.6597	0.9233	0.9389	0.8224	0.8592	0.9251	0.8859	0.8246	0.9138	0.8622	0.5729
		RDCT [26]	0.8853	0.6217	0.9133	0.8692	0.6943	0.9042	0.7419	0.7291	0.7567	0.8822	0.8845	0.7698
		CT [27]	0.8881	0.5025	0.9149	0.9290	0.7858	0.8922	0.9496	0.8897	0.7709	0.8929	0.8524	0.6123
		DWT-I	0.8783	0.6093	0.8899	0.8867	0.6722	0.7072	0.8532	0.7243	0.6928	0.9059	0.9205	0.7083
		DWT-II	0.8753	0.6198	0.8832	0.8838	0.6722	0.6576	0.7712	0.7164	0.7286	0.895	0.8922	0.6551
		Proposed-I	0.8918	0.6088	0.9132	0.8954	0.6636	0.9034	0.8788	0.7373	0.7319	0.8785	0.8954	0.8093
Proposed-II	0.8866	0.6163	0.9124	0.8791	0.6964	0.9052	0.7939	0.7288	0.7492	0.8868	0.8886	0.7938		
RMSE	FR	PSNR	7.9761	6.4927	5.8916	11.9144	8.7411	9.0104	9.4414	7.2909	8.5860	5.1877	7.2821	7.6936
		SSIM	6.2617	5.7029	5.5775	5.7395	8.5734	7.8913	3.9343	6.7580	5.4725	7.5093	5.7364	6.7366
		MS-SSIM	4.6426	4.8762	5.0822	4.7386	7.3159	6.0398	3.3132	5.3428	4.0621	8.3631	5.5908	6.8644
		VSNR	5.8928	5.9589	6.8631	6.2635	7.6663	7.4643	7.7113	7.4034	6.6151	3.3265	5.6396	7.1083
		VIF	4.4845	4.7822	6.0380	3.7836	6.3072	6.2197	6.0305	5.4424	4.1569	2.2314	3.7104	6.1099
		UQI	4.0033	4.1432	6.2409	4.2087	5.9226	5.5143	5.6283	5.4093	4.4133	2.7495	3.1621	5.6845
		FI-PSNR [61]	6.6623	6.2512	6.1612	8.3026	8.7545	7.9886	4.0781	6.2027	5.4572	9.3891	8.0362	8.6255
	RR	MJ3DQA [51]	6.5771	6.2554	5.1833	6.0463	9.1304	7.8164	2.9538	5.5931	4.8761	4.0949	5.2733	5.2841
		DNT [39]	4.4889	4.7310	6.3973	3.8672	6.3629	8.3455	3.6752	4.4403	3.7077	3.3802	4.8095	9.4876
		RDCT [26]	5.2058	4.5188	6.8307	5.6601	7.3330	7.0050	7.1243	6.4401	4.1501	3.4178	5.1904	6.7659
		CT [27]	4.9133	5.4292	6.5350	4.0748	6.8571	7.1557	3.2670	4.3369	4.6433	4.6228	4.9448	9.6303
		DWT-I	5.1449	4.9174	7.4954	5.3799	7.5856	11.611	5.3949	6.3417	4.5633	3.0791	4.2417	7.8783
		DWT-II	5.1145	4.874	7.6512	5.5823	7.5432	12.277	6.6285	6.4265	4.4052	3.8313	4.8265	8.3353
		Proposed-I	5.0268	4.5525	6.7969	5.2794	7.5119	7.0353	4.7441	6.5176	4.3004	3.4379	4.7700	6.0694
Proposed-II	5.1294	4.5229	6.7843	5.5105	7.3411	6.9542	6.2906	6.4439	4.1743	3.2635	4.9779	6.4939		

property with wavelet transform. Therefore, in order to further demonstrate the effectiveness of the proposed RDCT, we compare the performance of RDCT with discrete wavelet transform (DWT) of the RR IQA for stereoscopic images. For fair performance comparison, only RDCT is replaced by DWT, while the other processes of the proposed RR IQA remain unchanged. The experimental results are also illustrated in Table 1. It can be observed that the proposed method using RDCT can significantly outperform the one using DWT. Therefore, the RDCT is more suitable for GGD modeling and CBD calculation, which can further benefit the performance improvement, compared with DWT. Moreover, it can be observed that RDCT performs better than DNT and CT. DNT and CT are also designed by depicting the coefficient distribution in the wavelet domain. Therefore, the RDCT coefficient distribution is demonstrated to benefit more for the RR quality analysis than the wavelet coefficient distribution.

Moreover, RDCT performs comparatively with the methods of Proposed-I and Proposed-II on LIVE-Phase-I. However, the performance of Proposed-I and Proposed-II on LIVE-Phase-II is much worse than that on LIVE-Phase-I. The difference of the two databases is the symmetric property of the distortions, which introduced different

distortion levels to the left and right view image, respectively. For LIVE-Phase-I, the distortions introduced in left and right view image are the same. No perceptual quality difference exists between the left and right view images. The two view images will be perceived and contributed equally to the final stereoscopic image quality perception. However, for the non-symmetric distortions, the left and right view images present different distortions. The HVS perception of such stereoscopic images will be significantly affected. There exist complicated binocular perception properties of stereoscopic images, such as the binocular masking effect and error tolerance effect between the two view images. As such, the simple averaging process of the quality values of left and right view image cannot well depict the real quality perception behavior of HVS. Therefore, we cannot make sure a good performance of proposed RR IQA on LIVE-Phase-II. It is also the main reason why the other RR and FR metrics perform quite well on LIVE-Phase-I but poorly on LIVE-Phase-II. Therefore, we think that the perceptual qualities of the left and right view image need to be jointly analyzed to match the HVS perception behavior for the non-symmetric distortions. In this paper, we proposed to employ the difference image to depict the stereoscopic image quality. The distortion difference between left and right view image can be more

clearly depicted, which will help more accurately analyze the degradation level of the stereoscopic image. However, this joint consideration is simple and straightforward. In the future, we will consider more accurately modeling the relationship between the left and right view image to derive better RR IQAs for stereoscopic images.

Besides, the performance of Proposed-I is slightly worse than that of Proposed-II on LIVE-Phase-I, while better than that of Proposed-II on LIVE-Phase-II. It means that the difference image is capable to capture the introduced distortions, specifically for the non-symmetric distortions. However, the performance of RDCT demonstrates that the qualities of the left and right view images are also helpful for SIQA. It arises an open question that how we can joint consider the left, right view, and difference images to generate a better RR SIQA. By considering the Proposed-I and Proposed-II, the difference image is most important for quality analysis. The proposed RR SIQA metric can thus be viewed as a scalable approach. If only a few RR features are constrained, we only extract the RR features from the difference image for quality analysis. If more RR features are tolerant, we can extract all the RR features from the images, including the left view, right view, and difference image for quality analysis. Such process can help to improve the performances of the stereoscopic images with symmetric distortions.

## 5. Conclusion

In this paper, we propose a novel RR-SIQA metric based on the natural image statistics in RDCT domain. For the coefficient in the RDCT subband of luminance image and difference image of the stereoscopic images, the marginal distribution of the coefficients is approximately GGD distributed. For each RDCT subband of the luminance image and difference map of the stereoscopic images, the model parameters and fitting error of the GGD distribution, MI and ERD values are extracted as feature parameters. At the receiver side, the 3D visual quality is predicted by measuring the distance between the features. Experiments show that the proposed metric has good consistency with 3D subjective perception of human.

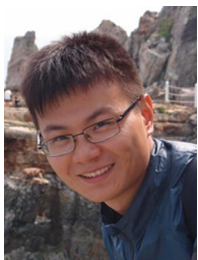
## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61501299 and 61471178, in part by the Guangdong Nature Science Foundation under Grant 2016A030310058, in part by the Shenzhen Emerging Industries of the Strategic Basic Research Project under Grants JCYJ20150525092941043 and JCYJ20130326105637578, in part by the Project 2016049 supported by SZU R/D Fund, in part by the grant from the Research Grants Council of the Hong Kong SAR, China (Project CUHK 415712), and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology (CICAEET) fund.

## References

- [1] W. Lin, C.C.J. Kuo, Perceptual visual quality metrics: a survey, *J. Visual Commun. Image Represent.* 22 (4) (2011) 297–312.
- [2] Z. Wang, H. Sheikh, A.C. Bovik, Objective video quality assessment, in: *The Handbook of Video Databases: Design and Application*, 2003, pp. 1041–1078.
- [3] Y. Zhou, S. Kwong, H. Guo, W. Gao, X. Wang, Bilevel optimization of block compressive sensing with perceptually nonlocal similarity, *Inf. Sci.* 360 (2016) 1–20.
- [4] H. Kong, Z. Lai, X. Wang, F. Liu, Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning, *Neurocomputing* 177 (2016) 198–205.
- [5] J. Li, X. Li, B. Yang, X. Sun, Segmentation-based image copy-move forgery detection scheme, *IEEE Trans. Inf. Forensics Secur.* 10 (3) (2015) 507–518.
- [6] B. Chen, H. Shu, G. Coatrieux, G. Chen, X. Sun, J.-L. Coatrieux, Color image analysis by quaternion-type moments, *J. Math. Imaging Vis.* 51 (1) (2015) 124–144.
- [7] Y. Zheng, B. Jeon, D. Xu, Q.M.J. Wu, H. Zhang, Image segmentation by generalized hierarchical fuzzy C-means algorithm, *J. Intell. Fuzzy Syst.* 28 (2) (2015) 961–973.
- [8] Z. Pan, Y. Zhang, S. Kwong, Efficient motion and disparity estimation optimization for low complexity multiview video coding, *IEEE Trans. Broadcast.* 61 (2) (2015) 166–176.
- [9] H. Yuan, S. Kwong, X. Wang, W. Gao, Y. Zhang, Rate distortion optimized inter-view frame level bit allocation method for MV-HEVC, *IEEE Trans. Multimed.* 17 (12) (2015) 2134–2146.
- [10] C. Hewage, M. Martini, Quality of experience for 3D video streaming, *IEEE Commun. Mag.* 51 (5) (2013) 101–107.
- [11] H. Yuan, S. Kwong, X. Wang, Y. Zhang, F. Li, A virtual view PSNR estimation method for 3-D videos, *IEEE Trans. Broadcast.* 62 (1), 2016, 134–140.
- [12] S. Daly, R. Held, D. Hoffman, Perceptual issues in stereoscopic signal processing, *IEEE Trans. Broadcast.* 57 (2 PART 2) (2011) 347–361.
- [13] X. Wang, M. Yu, Y. Yang, G. Jiang, Research on subjective stereoscopic image quality assessment, in: *Proceedings of the SPIE*, vol. 7255, 2009.
- [14] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.
- [15] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, *IEEE Trans. Image Process.* 16 (9) (2007) 2284–2298.
- [16] L. Ma, S. Li, K.N. Ngan, Visual horizontal effect for image quality assessment, *IEEE Signal Process. Lett.* 17 (7) (2010) 627–630.
- [17] S.S. Hemami, A.R. Reibman, No-reference image and video quality estimation: applications and human-motivated design, *Signal Process.: Image Commun.* 25 (7) (2010) 469–481.
- [18] H.R. Sheikh, A.C. Bovik, L. Cormack, No-reference quality assessment using nature scene statistics: JPEG 2000, *IEEE Trans. Image Process.* 14 (11) (2005) 1918–1927.
- [19] L. Liang, S. Wang, J. Chen, S. Ma, D. Zhao, W. Gao, No-reference perceptual image quality metric using gradient profiles for JPEG 2000, *Signal Process.: Image Commun.* 25 (7) (2010) 502–516.
- [20] T. Brandao, M.P. Queluz, No-reference image quality assessment based on DCT domain statistics, *Signal Process.* 88 (4) (2008) 822–833.
- [21] R. Ferzli, L.J. Karam, A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB), *IEEE Trans. Image Process.* 18 (4) (2009) 717–728.
- [22] Z. Wang, A.C. Bovik, Reduced- and no-reference image quality assessment: the natural scene statistic model approach, *IEEE Signal Process. Mag.* 28 (2011) 29–40.
- [23] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E. Yang, A.C. Bovik, Quality-aware image, *IEEE Trans. Image Process.* 15 (6) (2006) 1680–1689.
- [24] L. Ma, S. Li, F. Zhang, K.N. Ngan, Reduced-reference image quality assessment using reorganized DCT-based image representation, *IEEE Trans. Multimed.* 13 (4) (2011) 824–829.
- [25] L. Ma, S. Li, K.N. Ngan, Reduced-reference video quality assessment of compressed video sequences, *IEEE Trans. Circuits Syst. Video Technol.* 22 (10) (2012) 1441–1456.
- [26] L. Ma, S. Li, K.N. Ngan, Reduced-reference image quality assessment in re-organized dct domain, *Signal Process.: Image Commun.* 28 (8) (2013) 884–902.
- [27] X. Wang, Q. Liu, R. Wang, Z. Chen, Natural image statistics based 3D reduced reference image quality assessment in contourlet domain, *Neurocomputing* 151 (2) (2015) 683–691.
- [28] X. Wang, G. Jiang, M. Yu, Reduced reference image quality assessment based on contourlet domain and natural image statistics, in: *Fifth International Conference on Image and Graphics*, 2009, pp. 45–50.
- [29] F. Zhang, L. Ma, S. Li, K.N. Ngan, Practical image quality metric applied to image coding, *IEEE Trans. Multimed.* 13 (4) (2011) 615–624.
- [30] S. Wolf, M.H. Pinson, Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system, in: *Proceedings of the SPIE*, 1999.
- [31] S. Wolf, M.H. Pinson, Low bandwidth reduced reference video quality monitoring system, in: *Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005.
- [32] P.L. Callet, C.V. Gaudin, D. Barba, Continuous quality assessment of MPEG2 video with reduced reference, in: *Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005.
- [33] P.L. Callet, C.V. Gaudin, D. Barba, A convolutional neural network approach for objective video quality assessment, *IEEE Trans. Neural Netw.* 17 (5) (2006) 1316–1327.
- [34] M. Carnec, P.L. Callet, D. Barba, An image quality assessment method based on perception of structural information, in: *Proceedings of the ICIP*, 2003.
- [35] M. Carnec, P.L. Callet, D. Barba, Visual features for image quality assessment with reduced reference, in: *Proceedings of the ICIP*, 2005.
- [36] D. Tao, X. Li, W. Lu, X. Gao, Reduced-reference IQA in contourlet domain, *IEEE Trans. Syst. Man Cybern. Part B: Cybern.* 19 (6) (2009) 1623–1627.
- [37] U. Engelke, M. Kusuma, H.J. Zepernick, M. Caldera, Reduced-reference metric design for objective perceptual quality assessment in wireless imaging, *Signal Process.: Image Commun.* 24 (7) (2009) 525–547.
- [38] Z. Wang, E.P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, in: *Proceedings of the SPIE, Conference on Human Vision and Electronic Imaging*, X, vol. 5666, March 18, 2005, p. 149, <http://dx.doi.org/10.1117/12.597306Proc>.

- [39] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, *IEEE J. Sel. Top. Signal Process.* 3 (2) (2009) 202–211.
- [40] W. Xue, X. Mou, Reduced reference image quality assessment based on Weibull statistics, in: *Proceedings of the International Workshop on Quality of Multimedia Experience*, 2010.
- [41] M. Zhang, W. Xue, X. Mou, Reduced reference image quality assessment based on statistics of edge, in: *Proceedings of the SPIE*, 2011.
- [42] R. Soundararajan, A.C. Bovik, Rred indices: reduced reference entropic differencing framework for image quality assessment, in: *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2011.
- [43] C. Hewage, S. Worrall, S. Dogan, A. Kondoz, Prediction of stereoscopic video quality using objective quality models of 2-D video, *Electron. Lett.* 44 (2008) 963–965.
- [44] M. Pinson, S. Wolf, A new standardized method for objectively measuring video quality, *IEEE Trans. Broadcast.* 50 (3) (2004) 312–322.
- [45] P. Campisi, P. Le Callet, E. Marini, Stereoscopic images quality assessment, in: *European Signal Processing Conference*, 2007, pp. 2110–2114.
- [46] A. Benoit, P. Callet, P. Campisi, R. Cousseau, Using disparity for quality assessment of stereoscopic images, in: *Proceedings of the International Conference on Image Processing, ICIP*, 2008, pp. 389–392.
- [47] J. You, L. Xing, A. Perks, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, in: *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2010.
- [48] 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*, vol. 2006, 2006, pp. 218–222.
- [49] X. Wang, S. Kwong, Y. Zhang, Considering binocular spatial sensitivity in stereoscopic image quality assessment, in: *2011 IEEE Visual Communications and Image Processing, VCIP 2011*, 2011.
- [50] 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.
- [51] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. Cormack, A. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, *Signal Process.: Image Commun.* 28 (9) (2013) 1143–1155.
- [52] M.-J. Chen, L. Cormack, A. Bovik, No-reference quality assessment of natural stereopairs, *IEEE Trans. Image Process.* 22 (9) (2013) 3379–3391.
- [53] C. Hewage, M. Martini, Reduced-reference quality assessment for 3D video compression and transmission, *IEEE Trans. Consum. Electron.* 57 (3) (2011) 1185–1193.
- [54] G. Nur, G. Akar, An abstraction based reduced reference depth perception metric for 3D video, in: *Proceedings - International Conference on Image Processing, ICIP*, 2012, pp. 625–628.
- [55] A. Maalouf, M.-C. Larabi, CYCLOP: a stereo color image quality assessment metric, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 1161–1164.
- [56] Y. Liu, A.C. Bovik, L.K. Cormack, Disparity statistics in natural scenes, *J. Vis.* 8 (11) (2008) 1–14.
- [57] Y. Liu, L. Cormack, A. Bovik, Statistical modeling of 3-D natural scenes with application to Bayesian stereopsis, *IEEE Trans. Image Process.* 20 (9) (2011) 2515–2530.
- [58] S. Daly, The visible difference predictor: an algorithm for the assessment of image fidelity, in: *Digital Images and Human Vision*, 1993, pp. 179–206.
- [59] 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.
- [60] A.K. Moorthy, C.-C. Su, A. Bovik, Subjective evaluation of stereoscopic image quality, *Signal Process.: Image Commun.* 28 (8) (2012) 870–883.
- [61] 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.



**Lin Ma** received the B.E., and M.E. degrees from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, both in computer science. He received his Ph.D. degree in Department of Electronic Engineering at the Chinese University of Hong Kong (CUHK) in 2013. He is now a Researcher with Huawei Noah's Ark Lab, Hong Kong.

He was a Research Intern in Microsoft Research Asia from October 2007 to March 2008. He was a Research Assistant with the Department of Electronic Engineering, CUHK, from November 2008 to July 2009. He was a Visiting Student with the School of Computer Engineering, Nanyang Technological University (NTU),

from July 2011 to September 2011. His current research interests include multi-modal matching, deep learning, image/video processing, and quality assessment.

He got the best paper award in Pacific-Rim Conference on Multimedia (PCM) 2008. He was awarded the Microsoft Research Asia fellowship in 2011. He was a finalist to HKIS young scientist award in engineering science in 2012.



**Xu Wang** received the B.S. degree from South China Normal University, Guangzhou, China, in 2007, and M.S. degree from Ningbo University, Ningbo, China, in 2010. He received his Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, in 2014. In 2015, he joined the College of Computer Science and Software Engineering, Shenzhen University as an Assistant Professor. His research interests are 3D video coding and 3D image quality assessment.



**Qiong Liu** received her B.S. degree in Computer Science and Technology from Central China Normal University and Ph.D. degree in National Engineering Research Center for Multimedia Software from Wuhan University in 2008. She is now the faculty of Division of Communication and Intelligent Network, WNLO. Her research interests include multi-view video coding, depth estimation for multi-view video, low complexity video coding, region-of-interest video compression, scalable video coding and video surveillance



**King Ngi Ngan** received his B.Sc. (Hons) and Ph.D. degrees, both in Electrical Engineering from Loughborough University, UK, in 1978 and 1982, respectively. He joined the Department of Electronic Engineering, the Chinese University of Hong Kong as Chair Professor in 2003. He has been appointed Chair Professor at the University of Electronic Science and Technology, Chengdu, China, under the National Thousand Talents Program since 2012. Previously he was a Full Professor at the School of Computer Engineering, Nanyang Technological University, Singapore, and the School of Electrical, Electronic and Computer Engineering, University of Western Australia, Australia. He was appointed as Honorary Professor of Tunku Abdul Rahman University in Malaysia, Adjunct/Visiting Professor of Zhejiang University, Shanghai Jiaotong University, Shanghai University, Huazhong University of Science and Technology and Huaqiao University in China, Distinguished Adjunct Professor of the National Sun Yat-sen University in Taiwan, and Adjunct Professor of the University of Western Australia and RMIT University in Australia. He served as the External Examiner for the B.Eng. degree program of Malaysia's Multimedia University, and was appointed as IEEE Distinguished Lecturer of the IEEE Circuits and Systems Society in 2006–2007.

Prof. Ngan is active in many professional activities. He was an associate editor of the *Journal on Visual Communications and Image Representation*, U.S.A., and an area editor of *EURASIP Journal of Image Communications*, and served as an associate editor of *IEEE Transactions on Circuits and Systems for Video Technology* and *Journal of Applied Signal Processing*. He chaired a number of prestigious international conferences on video signal processing and communications and served on the advisory and technical committees of numerous professional organizations. He co-chaired the IEEE International Conference on Image Processing (ICIP) in 2010 in Hong Kong. He has published extensively including 3 authored books, 6 edited volumes and over 300 refereed technical papers in the areas of visual signal processing, image/video coding and multimedia communications, which are his main research interests. He also holds 11 patents on image coding/transmission techniques.