



# Reduced-reference image quality assessment in reorganized DCT domain

Lin Ma\*, Songnan Li, King Ngi Ngan

Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, NT., Hong Kong

## ARTICLE INFO

Available online 23 August 2012

### Keywords:

Image quality assessment (IQA)  
Reduced-reference (RR)  
Human visual system (HVS)  
Reorganized discrete cosine transform (RDCT)

## ABSTRACT

In this paper, a novel reduced-reference (RR) image quality assessment (IQA) is proposed by depicting the subband statistical characteristics in the reorganized discrete cosine transform (RDCT) domain. First, the block-based DCT coefficients are reorganized into a three-level coefficient tree, resulting in ten RDCT subbands. For the intra RDCT subband characteristic, the coefficient distribution of each RDCT subband is modeled by the generalized Gaussian density (GGD) function. The city-block distance (CBD) is employed to measure the modeling error between the actual distribution and the fitted GGD curve. For the inter RDCT subband characteristic, the mutual information (MI) is utilized to depict the dependencies between coefficient pairs in related RDCT subbands. Moreover, a frequency ratio descriptor (FRD) calculated in the RDCT domain is proposed to depict how the image energy distributes among different frequency components. The FRD values computed from both the reference and distorted images are jointly considered to derive a novel mutual masking strategy for simulating the texture masking property of the human visual system (HVS). By considering the GGD modeling of intra RDCT subband, MI of inter RDCT subbands, and FRD of the image, the proposed RR IQA is developed. Experimental results demonstrate that a small number of RR features is sufficient to represent the reference image for the perceptual quality analysis. The proposed method can outperform the state-of-the-art RR IQAs, and even the full-reference (FR) PSNR and SSIM.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

Image perceptual quality measurement plays a key role in the image processing and communication [1,2] such as image compression, printing, displaying, restoration, enhancement, transmission, and so on. Human eyes are the ultimate receivers of the images. Therefore, the most reliable way for assessing the image perceptual quality is the subjective testing method [3,4]. However, these subjective testing methods require many observers to participate in the experiments and provide their

personal opinions on the image/video perceptual quality, which make them very time consuming and expensive. Therefore, the subjective testing method is not suitable for the practical image applications, such as guiding the image compression, optimization for the image denoising, and so on. The image quality metrics are in demand, which can automatically evaluate the image quality and guide the image processing applications.

According to the availability of the reference image, the quality metrics can be roughly categorized into three classes [5]: full-reference (FR) [7–12], no-reference (NR) [13–19], and reduced-reference (RR) [20–36,58–62,68,69]. In order to evaluate the perceptual quality of the distorted image, the FR metrics require the whole reference image, which is assumed to be artifact free and of perfect quality. These metrics can only be applied to the applications where the

\* Corresponding author. Tel.: +852 39438255.

E-mail addresses: [lma@ee.cuhk.edu.hk](mailto:lma@ee.cuhk.edu.hk) (L. Ma), [snli@ee.cuhk.edu.hk](mailto:snli@ee.cuhk.edu.hk) (S. Li), [knngan@ee.cuhk.edu.hk](mailto:knngan@ee.cuhk.edu.hk) (K.N. Ngan).

reference image is available, such as image compression [10], watermarking [12,55], and so on. The simplest FR metrics are the mean squared error (MSE) and the corresponding peak signal-to-noise ratio (PSNR). MSE and PSNR are widely adopted, because of their clear physical meanings and friendliness for optimization. However, MSE and PSNR only consider the differences in the pixel level, which are not related to the perception property of the human visual system (HVS) [6–33]. They are not reliable for evaluating or even controlling the perceptual quality of the distorted image during the processing stages. Therefore, the structure similarity (SSIM) index [7] is developed by depicting the structural distortions rather than the pixel absolute difference, which is more sensitive to the HVS. In [10], the orientation sensitivity and saliency property of the HVS are incorporated to refine the SSIM distortion map, which can ensure a better performance. Chandler et al. [8] proposed the visual signal-to-noise ratio (VSNR) to evaluate the perceptual quality of the distorted image in the wavelet domain. The disadvantage of these IQAs is the complexity, especially for optimization in the practical image processing applications, such as image coding and watermarking. In [10], Zhang et al. simply considered the contrast sensitivity function and texture masking effect of the HVS to develop a practical image quality metric, which has been proved to be effective for guiding perceptual-based image compression. Subsequently, this practical quality metric is employed for image watermarking [12].

In many real-world applications, we cannot access the reference image for the quality evaluation, such as image denoising, restoration, etc., where only the distorted image is available for analysis. Therefore, the NR IQAs [14–19] are thus needed to evaluate and control the perceptual quality of the processed image. 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 [14]. Liang et al. [15] 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 [16] to predict the PSNR value of the JPEG coded image. Furthermore, Ferzli et al. [18] 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 features extracted from the reference image. Only a small number of bits is 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. The image perceptual quality can be easily analyzed by referring to the extracted

features from the reference image. Therefore, a better quality of user experience can be further provided for the consumers.

For designing an effective RR quality metric, we need to consider not only its performance but also its RR data rate for representing the extracted features. First, the extracted features should be sensitive to a variety of image distortions and relevant to the HVS perception of the image quality. Second, the RR data rates should not be large, as the extracted features need to be embedded or transmitted to the receiver side for the quality analysis. For a larger RR data rate, one may include more information about the reference image. Then a good performance can be obtained. However, it will introduce a heavy burden to the RR feature transmission. The FR IQA can be regarded as an extreme case of RR IQA, with the RR data rate is the whole reference image. For a smaller RR data rate, only a little information of the reference image is available for quality analysis. Therefore, the performance is hard to be ensured. The NR IQA is another extreme case of RR IQA, with no information from the reference image. Therefore, how to balance the RR data rate and the performance is the essential for the RR quality metric development.

In [20–23], the distortions introduced are modeled to derive the RR quality metrics. Wolf et al. [20,22] proposed to extract a set of spatial and temporal features for measuring the distortions, which occur in the standard video compression and communication environment. The features that are associated with the blurring, blocking, and frame differences are extracted in [21] to depict the compression artifacts introduced by MPEG-2. The ratio between the parent coefficient (second DCT coefficient) and the child coefficient (the third and fourth DCT coefficients) is employed to measure the perceptual quality of the MPEG-2 coded video sequence [23]. The amplitude and phase information of the harmonics [60] generated by blocking artifacts are employed to estimate the blockiness of the MPEG-2 coded video sequences. 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.

As the human eyes are the ultimate receivers of the image, the HVS properties need to be considered for designing an effective RR IQA. Le Callet et al. [24] 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 [25,26], 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 Weber's law of just noticeable difference are incorporated to derive an RR IQA [27], which are employed for evaluating the perceptual qualities of the JPEG and JPEG 2000 coded images. Recently, an RR IQA [28] for wireless imaging is developed by considering different structural information that is observed in the distortion model of wireless link. The structural information from the viewing area is trained for the HVS perception. Moreover, as the HVS is sensitive to the degradation around the edges, the RR video quality metric proposed

in [58] mainly measures the edge degradations. The edge degradation is computed by measuring the mean squared error of the edge pixels. Therefore, this method is named as edge PSNR (EPSNR). In [59], the authors employed discriminative local harmonic strength with motion consideration to evaluate the distorted video quality, which we name as RR-LHS. The gradient information of each frame is employed for harmonic and discriminative analysis.

Recently, the statistical modeling of the image signal has been investigated for the image perceptual quality assessment for both RR IQAs [29,36,61,62,68,69] and NR IQAs [14,17,19]. In [32], the divisive normalization is employed to depict the coefficient distributions of the wavelet subbands. The distribution difference between the reference and distorted images is used to depict the image perceptual quality, which we name as RR-DNT. In [61,62], the developed RR image quality metric RR-SSIM extracted the statistical features from a multi-scale, multi-orientation divisive normalization transform. By following the philosophy in the construction of SSIM, a distortion measurement is developed to estimate the SSIM index of the distorted image. In [68], the statistics of image gradient magnitude are modeled by the Weibull distribution to develop an RR image quality metric, which is named as RR-Weibull. Also the statistics of the edge [69] are utilized for developing the RR IQA, which we name as RR-Edge. In [34], 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 [36], the color distribution changes of an image as a consequence of the distortions are employed for depicting the perceptual quality, where the color correlogram is extracted as the RR feature. Wang et al. [29,30] 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. Although WNISM can achieve good performances on image quality assessment, some limitations can still be figured out. First, KLD is asymmetric [39], which is not suitable for the quality analysis, as demonstrated in [33]. The perceptual quality distance from one image to another should be identical no matter how it is measured. Second, as revealed in [32], although WNISM can work quite well on individual distortion types, its performance degrades significantly when image of different distortion types are evaluated together. In order to overcome the aforementioned limitations, the RR quality metric in [33] is proposed by employing the reorganized discrete cosine transform (RDCT) based image representation. 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. However, although a better performance can be achieved as shown in [33], only the identical nature of the coefficient distribution in the same RDCT subband is utilized. As to be demonstrated in the following section, the correlations between related RDCT subbands also exist. However, the RR IQAs [29,33] do not consider the inter RDCT subband relationships. In order to

design an effective RR IQA, both the intra and inter RDCT subband statistical characteristics need to be considered.

As aforementioned, the statistical properties of the image signals which are sensitive to the introduced distortions can be employed to develop IQA methods. DCT has been widely adopted and employed for image/video compression, deblocking, denoising, and so on. Therefore, if we can extract several HVS-sensitive features based on the DCT coefficients, they can be utilized as the RR feature for quality analysis or even guiding the image processing. As shown in [33], the coefficient distribution of the RDCT subband can be accurately modeled by GGD, which is useful for designing the RR IQA. Moreover, the inter RDCT subband relationship needs to be further depicted and modeled. This statistical property will be affected by the introduced distortions. Furthermore, as the image has been represented in the RDCT domain, the relationship between different frequency components should be considered and depicted. First, the energy variations in different frequencies can somewhat represent the distortion level. Second, the energy distributions over different frequencies can help to model the HVS masking properties. As a result, we will manipulate the DCT coefficients in the RDCT domain from the three aspects discussed above to evaluate the perceptual qualities of the distorted images.

In this paper, a novel RR IQA is developed by depicting the intra and inter subband statistical characteristics in the RDCT domain. It is shown that after performing DCT the statistical dependencies between the DCT subbands still exist. Applying the reorganization strategy, the intra RDCT subband statistical characteristic, specifically the identical nature of the coefficient distribution within the RDCT subband, is exploited by the GGD modeling. The inter RDCT subband dependency is captured by the mutual information (MI) between the DCT coefficient pair in corresponding RDCT subbands, such as parent–child pair coefficient, brother–child pair coefficient, and cousin–child pair coefficient. Furthermore, a frequency ratio descriptor (FRD) computed in the RDCT domain is employed to measure the energy distribution among different frequency components. It can be further utilized to simulate the HVS texture masking property. By considering the intra RDCT subband GGD modeling, inter RDCT subband MI values, and the image FRD value, an effective RR IQA is developed.

This paper is organized as follows. The relationship of intra and inter RDCT subbands is presented in Section 2. The detailed algorithms, including the feature extraction in the sender side and quality analysis in the receiver side, will be introduced in Section 3. Section 4 will demonstrate the performance comparisons. Finally, the conclusion will be given in Section 5.

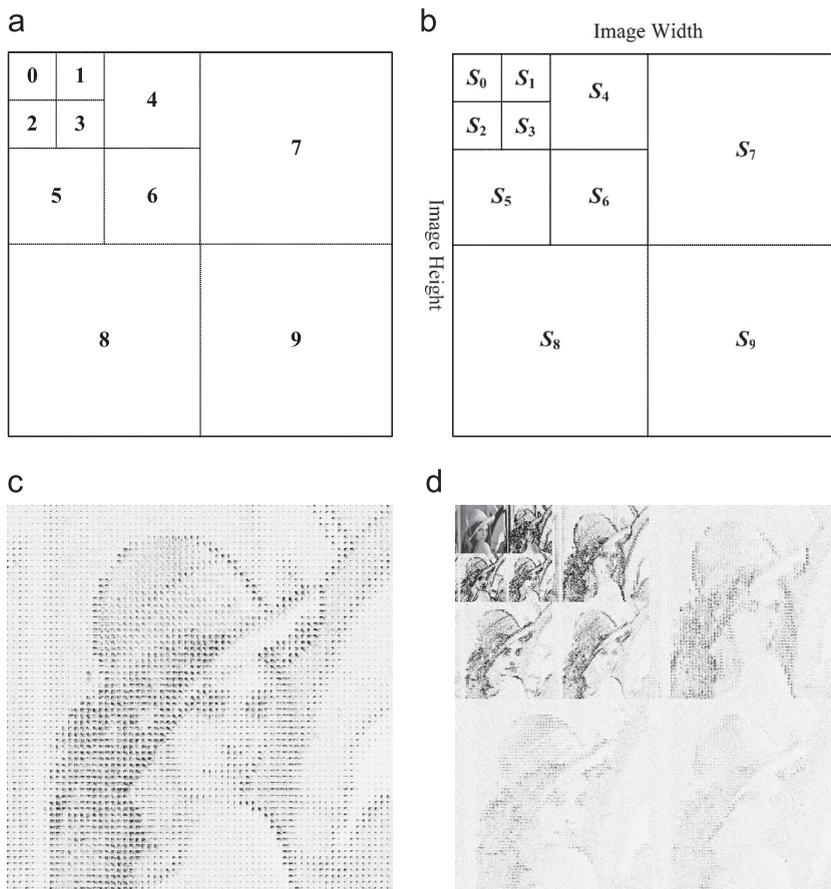
## 2. Relationship analysis of intra and inter RDCT subbands

Since the HVS is more sensitive to luminance than chrominance [63], the proposed image quality metric and the other compared ones work with luminance only. Color inputs will be converted to gray scale before further analysis. The coefficients obtained by the block-based DCT present high correlations, which can be employed for depicting the image degradation level. In [23], after

the 8-tap DCT, the second DCT coefficient and the third/fourth DCT coefficient are related to each other as the parent and child bands of the wavelet transform. In [16], the Laplace probability density function (pdf) is employed to model the coefficient distribution of each DCT subband. The fitted Laplace pdf parameter  $\lambda$  of one DCT subband can be linearly predicted by the  $\lambda$  values of the neighboring upper and left DCT subbands. Therefore, although DCT has decomposed the spatial image block into different frequency components, the relationship between related DCT subbands still exists. In order to utilize the identical nature of the neighboring coefficient distributions, the reorganization strategy [33,37,38] is employed to compose the block-based DCT coefficients into a three-level tree structure, as demonstrated in Fig. 1.

First, the  $8 \times 8$  DCT is performed block by block on the image. For each  $8 \times 8$  DCT block, the DCT coefficients are decomposed into ten subbands, as shown in Fig. 1(a). For the subbands 0, 1, 2, and 3, each subband only contains one DCT coefficient. For the subbands 4, 5, and 6, each subband contains a  $2 \times 2$  DCT coefficient matrix. For the subbands 7, 8, and 9, each subband contains a  $4 \times 4$  DCT coefficient matrix. After the decomposition, the same

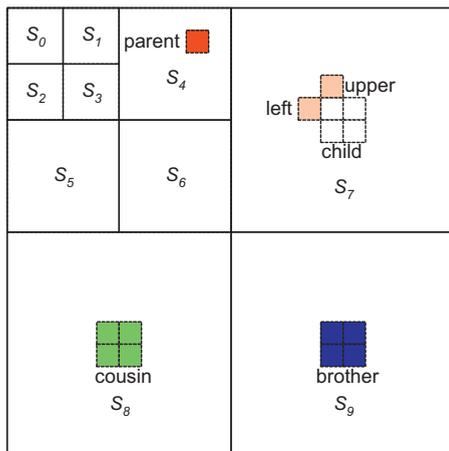
subbands of all the  $8 \times 8$  DCT blocks are grouped and organized together according to their corresponding positions, as shown in Fig. 1(b). In this manner, the block-based DCT coefficients are reorganized into a three-level coefficient tree. In Fig. 1,  $S_n$  denotes the grouped subband of all the DCT coefficients lying on the position denoted by  $n$ . For example,  $S_7$  is the reorganized subband by grouping the  $4 \times 4$  DCT coefficient matrix lying on the position 7 of all the  $8 \times 8$  DCT blocks. An example of the reorganization of the LENA DCT coefficient image is illustrated in Fig. 1. The  $8 \times 8$  DCT representation is obtained by applying the non-overlapped  $8 \times 8$  block based DCT, as shown in Fig. 1(c). The reorganized DCT (RDCT) representation is shown in Fig. 1(d). It can be observed that the RDCT representation appears like a wavelet representation [52], i.e., exhibiting structural similarities between subbands, and coefficient magnitude decaying toward high-frequency subbands. Moreover, the RDCT representation is more efficient for the RR quality metric design than the wavelet representation, such as the steerable pyramid [64,65], even though the wavelet directly have an access to the oriented subbands. As demonstrated in [33], the generalized Gaussian density (GGD) can more accurately depict



**Fig. 1.** Reorganization strategy of DCT coefficients. (a) one  $8 \times 8$  DCT block with ten subband decomposition; (b) the reorganized DCT image representation taken as a three-level coefficient tree; (c)  $8 \times 8$  DCT representation of LENA image; (d) the RDCT representation of LENA image. (For better visualization, the DC components are rescaled to integers between 0 and 255, while the AC coefficients are obtained by  $255 - (5 \times |AC|)$ .)

the coefficient distributions of the RDCT subbands than those of the steerable pyramid ones. Furthermore, the experimental results in [33] have clearly demonstrated that the RDCT representation is more effective than the steerable pyramid to develop the RR IQA.

The statistical relationships between RDCT coefficients are examined in the following ways. First, consider the parent-child coefficient pair representing the information at adjacent scale subbands of the same orientation (e.g.,  $S_4$  and  $S_7$ ). Each parent coefficient in the subband  $S_4$  corresponds to four child coefficients in the subband  $S_7$ , as illustrated in Fig. 2. In order to exploit the underlying statistics, the joint histogram of the coefficient pair (parent, child) is built, which is gathered over the spatial extent of the image. Fig. 3(a) shows the conditional histogram  $h(child|parent)$ , which is simply calculated by counting the child coefficients in the subband  $S_7$  conditioned on the coarser-scale subband  $S_4$ . Several important aspects can be observed from the conditioned histogram. These coefficients are approximately second-order decorrelated, as the value of the child coefficient is always zero when the values of parent coefficients are not large enough. Moreover, the standard deviation of the child coefficients highly depends on the value of the parent coefficient. The larger the parent coefficient value, the larger the standard deviation of the child coefficients tends to be, as illustrated by the blue curve in Fig. 3. In [42,43], it has been demonstrated that the mean and the standard deviation curves of the conditional histogram can be well fitted by a Student's  $t$  model of a cluster of coefficients. Furthermore, although they are decorrelated, the statistical dependency can still be observed between the child and parent coefficients. These dependencies also exist in the wavelet coefficient pairs [40], which cannot be eliminated by the linear transformations. This statistical dependency can be more clearly observed by converting the coefficient value into the log-domain as shown in Fig. 3(b). The left part of the conditional histogram  $h(\log_2(child)|\log_2(parent))$  concentrates on a nearly horizontal line (shown by the green curve), which means that the value



**Fig. 2.** Statistical correlation between inter RDCT subbands. Each parent coefficient in the coarser scale RDCT subband corresponds to four child coefficients in the finer scale subband. Each child coefficient corresponds to one cousin/brother coefficient in the same scale subbands of different orientations.

of  $\log_2(child)$  is independent of  $\log_2(parent)$  in this area. Actually, natural images are composed of smooth regions which are delimited by edge discontinuities. After performing DCT, most of the image energy is compacted to the low-frequency components, which results in a small amount of energy in the high-frequency components. Therefore, the child coefficient values in the finer RDCT subband tend to be small, especially when the parent coefficient values are not large enough. The right part of the conditional histogram in log-domain presents a nearly linear correlation. It implies that the conditional expectation  $e(\log_2(child)|\log_2(parent))$  is approximately proportional to  $\log_2(parent)$ .

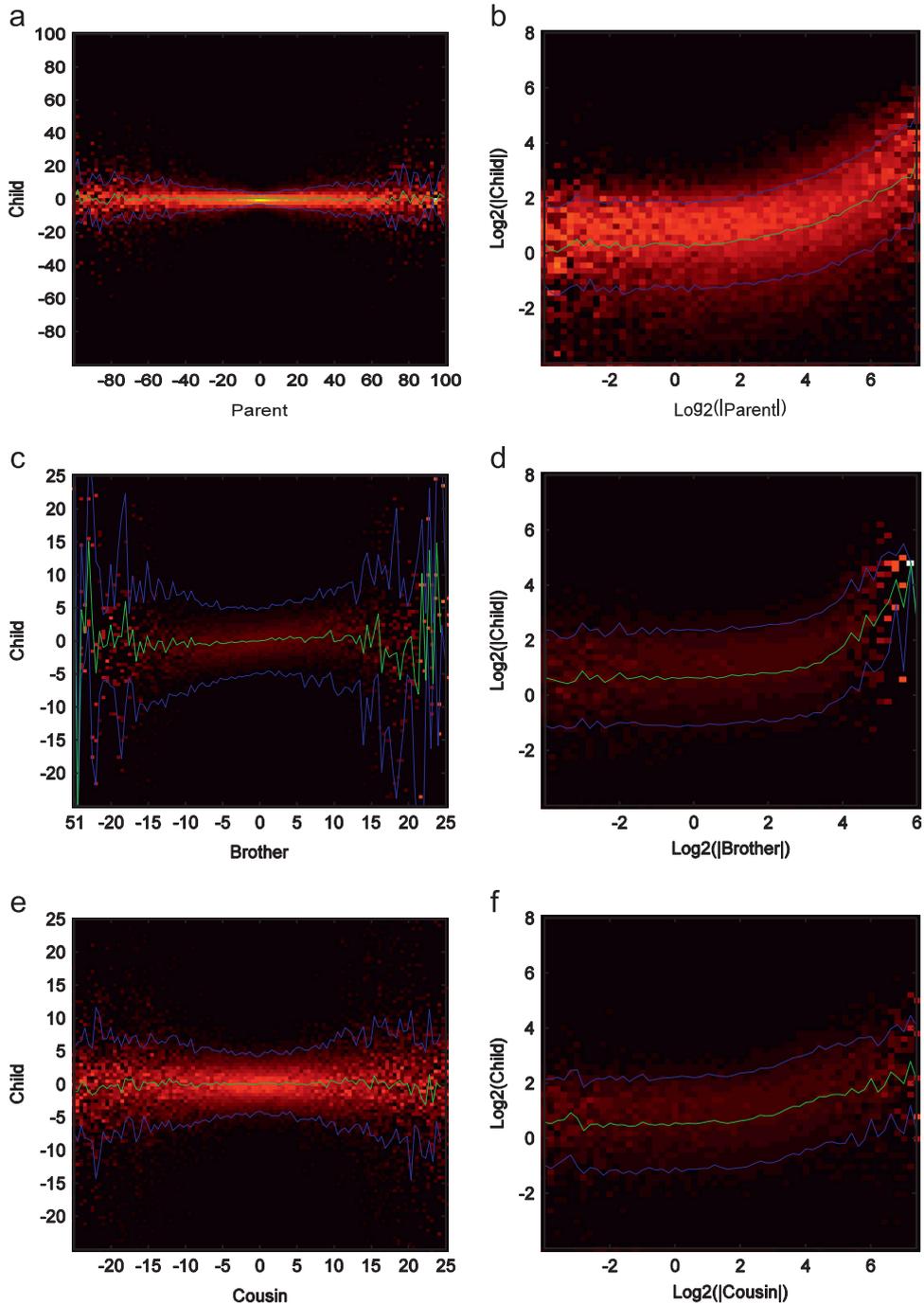
Fig. 3(c) and (e) show the histograms of the child coefficient conditioned on the brother and cousin coefficient, respectively. Compared with the conditional histogram in Fig. 3(a), the child coefficient values vary significantly, which do not present a close scattering around the zero value. When the cousin or brother coefficient value becomes larger, the child value fluctuates more dramatically, which can be observed by the standard deviation values (the blue curve of each figure). The brightness of Fig. 3 corresponds to the probability. The brighter the area, the larger the corresponding probability is. Compared with Fig. 3(a), the brightness of Fig. 3(c) and (e) is not so significant. It means that the child coefficient value depends on the cousin/brother coefficient less than the parent coefficient. Furthermore, it can be observed that Fig. 3(e) is brighter than Fig. 3(c). And the standard deviation curves of Fig. 3(e) appear to be more regular than those of Fig. 3(c). The observations show that the dependency relationship between child-cousin coefficient pair is closer than that between child-brother coefficient pair. After converting the histograms into log-domain, as shown in Fig. 3(d) and (f), the correlations appear much looser. Although the mean value in the log-domain concentrates approximately on a line, the child coefficient values are of great differences. The child coefficients do not present a concentrated distribution (with larger standard deviation values), which makes the probability of each coefficient value to be very small. Therefore, the brightness of Fig. 3(d) and (f) can hardly be detected. Although the relationship between child and brother/cousin appears to be much looser than that between child and parent, it is admitted that the dependency does exist however in a very complex way, which is very hard to depict.

In order to provide a more accurate description about the relationship between RDCT subbands, the mutual information (MI) is employed to describe the dependencies between the child and its condition parent, brother, cousin, upper, and left coefficients, as illustrated in Fig. 2. As introduced in [39], MI admits the direct data compression and classification interpretations. Let  $X$  and  $Y$  be two random variables (or vectors) having a joint pdf  $p(x,y)$ . The MI between  $X$  and  $Y$  is defined as:

$$I(X; Y) = \int_x \int_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dx dy$$

$$\triangleq E_{XY} \left[ \log \frac{p(x,y)}{p(x)p(y)} \right] = D(p(x,y) || p(x)p(y)) \quad (1)$$

where  $D(||)$  is the relative entropy between two distributions, known as the KLD. The MI  $I(X;Y)$  indicates how



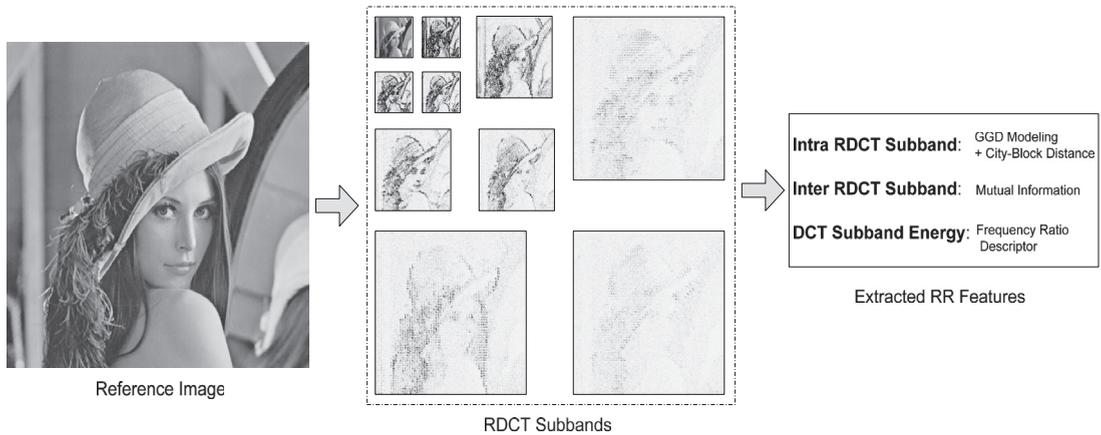
**Fig. 3.** Conditional histogram for the coefficients of the RDCT subbands from the BOAT image. Brightness corresponds to the probability. Each column has been individually rescaled for a better visualization. (a) histogram of the child coefficient conditioned on the parent coefficient; (b) log-domain representation of (a); (c) histogram of the child coefficient conditioned on the brother coefficient; (d) log-domain representation of (c); (e) histogram of the child coefficient conditioned on the cousin coefficient; (f) log-domain representation of (e). The green curve corresponds to  $E(\text{child}|\text{condition})$ , and the blue curves correspond to  $E(\text{child}|\text{condition}) \pm \text{std}(\text{child}|\text{condition})$ , where the *condition* of each figure is the parent, cousin, and brother, respectively.

much information  $Y$  conveys about  $X$ . Therefore, the larger the MI value, the more information is shared by  $X$  and  $Y$ . Hence, the statistical correlation between  $X$  and  $Y$  is stronger.

The MI values between the RDCT subbands are illustrated in Table 1. We have provided the MI values of the inter RDCT subbands, such as parent–child  $S_4$  and  $S_7$ , brother–child  $S_9$  and  $S_7$ , and cousin–child  $S_8$  and  $S_7$ , and

**Table 1**  
Mutual information between the RDCT subbands.

Subband orientation	Inter RDCT subband			Intra RDCT subband	
Horizontal	Parent-child ( $S_4$ and $S_7$ )	Brother-child ( $S_9$ and $S_7$ )	Cousin-child ( $S_8$ and $S_7$ )	Upper-child ( $S_7$ )	Left-Child( $S_7$ )
	0.5496	0.2739	0.2908	0.3892	0.3918
Vertical	Parent-child ( $S_5$ and $S_8$ )	Brother-child ( $S_9$ and $S_8$ )	Cousin-child ( $S_7$ and $S_8$ )	Upper-child ( $S_8$ )	Left-child ( $S_8$ )
	0.5091	0.2685	0.2908	0.3672	0.3508
Diagonal	Parent-child ( $S_6$ and $S_9$ )	Brother-child ( $S_7$ and $S_9$ )	Cousin-child ( $S_8$ and $S_9$ )	Upper-child ( $S_9$ )	Left-child ( $S_9$ )
	0.2974	0.2739	0.2685	0.2165	0.2095



**Fig. 4.** RR feature extraction in the sender side.

the intra RDCT subbands, such as upper-child and left-child. In order to provide a more convincing result, each entry gives the average MI value over all the reference images from the LIVE image quality assessment database [50]. Some interesting findings can be observed. First, no matter what the subband orientation is, the MI value of parent-child is the largest. It means that the parent coefficients in coarser subband affect the child coefficients in the finer subband most, which presents the same property as the wavelet transform. These dependencies have been successfully utilized for the image compression [37,38]. Second, the parent-child MI value of diagonal RDCT subband is much smaller than those of horizontal and vertical ones. The reason is that natural images present much more horizontal and vertical information than the diagonal one. Therefore, most of the DCT coefficients in the diagonal subbands tend to be zero. Furthermore, the MI values somewhat match the HVS property, namely the oblique effect [51], that is, the HVS is more sensitive to the horizontal and vertical frequency components, compared with the diagonal ones. Thirdly, for the horizontal and vertical RDCT subbands, the MI values of intra RDCT subbands are larger than those of brother-child and cousin-child inter RDCT subbands. Therefore, the relationship between neighboring DCT coefficients also exists. This relationship has been further employed for image compression [38] and image quality metric [16,33]. In [16], the authors employ the neighboring DCT subband relationship to improve the modeling accuracy of the DCT coefficient distribution. The adjacent DCT coefficients are reorganized into several representative subbands for designing the RR IQA

[33]. Finally, the dependencies between cousin-child and brother-child RDCT subbands can be observed. Although DCT has decomposed the spatial image content into different components with different orientations and frequencies, the dependencies cannot be removed by the linear transformations. Therefore, the correlations between inter RDCT subbands can be exploited for image processing researches, such as compression [40,41], and so on.

### 3. The proposed reduced-reference image quality metric

As discussed above, the RR IQAs aim at evaluating the image perceptual quality based on some RR features extracted from the reference image. In order to design an effective RR IQA, the features extracted should be sensitive to the distortions, related to the HVS perception property, and efficient for representation. Therefore, the RR features are critical to the RR IQA performances. Based on the analysis in the above section, the dependencies of intra and inter RDCT subbands do exist, which can be depicted and quantified in the receiver side. And these dependencies are expected to be sensitive to the distortions, which can be utilized as the RR features for the quality analysis in the receiver side.

#### 3.1. RR feature extraction in the sender side

Fig. 4 provides the framework of extracting the RR features from the reference image. As shown in Table 1, the intra RDCT subband correlation, which has been proved to be

sensitive to the image distortion, can be utilized for the RR IQA [33]. The identical nature of the coefficient distributions between adjacent DCT subbands in the same RDCT subband is exploited for extracting the RR features. As the coefficient distribution of each RDCT subband is highly kurtotic [29], the generalized Gaussian density (GGD) is employed to model its distribution:

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

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

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

Here  $\alpha$  models the width of the PDF peak (standard deviation), while  $\beta$  is inversely proportional to the decreasing rate of the peak.  $\alpha$  and  $\beta$  are also referred to as the scale and shape parameters, respectively. As demonstrated in [33], GGD is demonstrated to be suitable for modeling the highly kurtotic coefficient distribution in the RDCT domain, which results in a better performance. In order to further improve the modeling accuracy, another parameter except  $(\alpha, \beta)$  is employed which is named as the city-block distance (CBD) [27,33]:

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

where  $p$  is the histogram distribution of the actual RDCT subband,  $p_{\alpha,\beta}$  is the fitted GGD curve, and  $h_L$  is the total number of the histogram bins. Compared with KLD, CBD is symmetrical, which is reasonable for evaluating image perceptual quality [33]. According to the oblique effect [51] of the HVS, human eyes present similar sensitive values to the horizontal and vertical information, while less sensitive to the diagonal information. Therefore, in order to reduce the RR data rates, only three horizontal RDCT subbands, specifically  $S_1$ ,  $S_4$ , and  $S_7$ , are employed for GGD modeling and CBD calculation to extract the RR features.

Referring to Table 1, the dependencies between inter RDCT subbands also exist. MI as defined in (1) is employed to capture the corresponding dependencies, which can be further expressed as:

$$\begin{aligned} I(X; Y) &= h(X) - h(X|Y) \\ &= E_X(-\log_2 p(x)) - E_{X,Y}(-\log_2 p(x|y)) \end{aligned} \quad (5)$$

where  $h(X)$  and  $h(X|Y)$  denote the entropy of  $X$  and  $X$  conditioned on  $Y$ , respectively. As shown in (5), we can observe that the MI is symmetric and non-negative. If  $X$  and  $Y$  are independent, the MI is equal to zero. While if  $X$  is a function of  $Y$ ,  $I(X; Y) = \infty$ . Actually, the MI  $I(X; Y)$  indicates how much information  $Y$  conveys about  $X$ . It admits a well-know data compression interpretation: coding  $X$  to a precision  $\Delta X$  costs  $h(X) - \log_2(\Delta X)$  bits, based on the assumption that  $\Delta X$  is sufficiently small. If  $Y$  is known, by considering the same encoding precision  $\Delta X$ , the total bits cost for encoding  $X$  is equal to  $h(X) - \log_2(\Delta X) - I(X; Y)$  bits [39]. Therefore, the total saving bits by introducing  $Y$  is  $I(X; Y)$ .

The MI value is introduced in the sender side to describe the essential relationship between inter RDCT subbands, which is changed by the introduced distortion. As we employed three horizontal RDCT subbands ( $S_1$ ,  $S_4$ , and  $S_7$ ) for GGD modeling and CBD calculation to depict the intra RDCT relationship, the MI values between these horizontal RDCT subbands and other related ones are computed as the RR features to depict the inter RDCT dependencies. These RR features include two MI values to depict the parent-child correlation between the RDCT subband pairs ( $S_1$ ,  $S_4$ ) and ( $S_4$ ,  $S_7$ ), three MI values to depict the cousin-child correlation between the RDCT subband pairs ( $S_2$ ,  $S_1$ ), ( $S_5$ ,  $S_4$ ), and ( $S_8$ ,  $S_7$ ), and three MI values to depict the brother-child correlation between the RDCT subband pairs, ( $S_3$ ,  $S_1$ ), ( $S_6$ ,  $S_4$ ), and ( $S_9$ ,  $S_7$ ). Therefore, there are 8 MI values in total extracted to capture the inter RDCT subband dependencies.

Furthermore, in order to accurately depict the reference image characteristic, an image-level feature, specifically the frequency ratio descriptor (FRD), is proposed by considering the HVS properties. For our RR feature extraction, after performing  $8 \times 8$  DCT, the coefficients are reorganized into several RDCT subbands, as illustrated in Fig. 1. The frequency  $\omega_{ij}$  of the  $(i, j)$ th subband for each  $8 \times 8$  DCT block can be obtained by [46]:

$$\begin{aligned} \omega_{ij} &= \frac{1}{2N} \sqrt{(i/\theta_x)^2 + (j/\theta_y)^2} \\ \theta_\Delta &= 2 \times \arctan\left(\frac{\aleph_\Delta}{2\Delta}\right), (\Delta = x, y) \end{aligned} \quad (6)$$

where  $N$  is the dimension of the DCT block (in this paper,  $N=8$ ),  $\theta_x$  and  $\theta_y$  are the horizontal and vertical visual angles of a pixel.  $l$  is the viewing distance and  $\aleph$  stands for the display width/length of a pixel on the monitor. According to the international standard ITU-R BT.500-11 [47], the ratio of viewing distance to picture height should be a fixed number between 3 and 6. Moreover, for most of the displays, pixel aspect ratio (PAR) is equal to 1. It means that the horizontal and vertical visual angles ( $\theta_x, \theta_y$ ) are identical:

$$\theta_x = \theta_y = 2 \times \arctan\left(\frac{1}{2 \times R_{vd} \times H_{\text{pic}}}\right), \quad (7)$$

where  $R_{vd}$  is the ratio of viewing distance to picture height.  $H_{\text{pic}}$  is the number of pixels in picture height. The frequency values obtained by (6) and the spatial contrast sensitivity function (CSF) values [67] of the  $8 \times 8$  DCT subbands are illustrated in Fig. 5. It can be observed that the adjacent DCT subbands present similar frequency and CSF values. The lower the frequency component, the larger the CSF value is. After the reorganization process introduced in Section 2, the CSF values of the RDCT subbands  $S_0$ ,  $S_1$ ,  $S_2$ , and  $S_3$  are larger than 0.5, which are the most sensitive components to the HVS. By checking the frequency  $\omega_{ij}$  value, we can find that the frequency values of these RDCT subbands are smaller than 5. Therefore, these RDCT subbands  $S_0$ ,  $S_1$ ,  $S_2$ , and  $S_3$  (denoted by the red box) are simply regarded as the low-frequency (LF) components. For the RDCT subbands  $S_4$ ,  $S_5$ , and  $S_6$ , the CSF values (except the one of  $\omega_{33}$ ) are larger than 0.2 and smaller than 0.5. And the frequency values (except  $\omega_{33}$ ) are larger than 5 and smaller than 12.

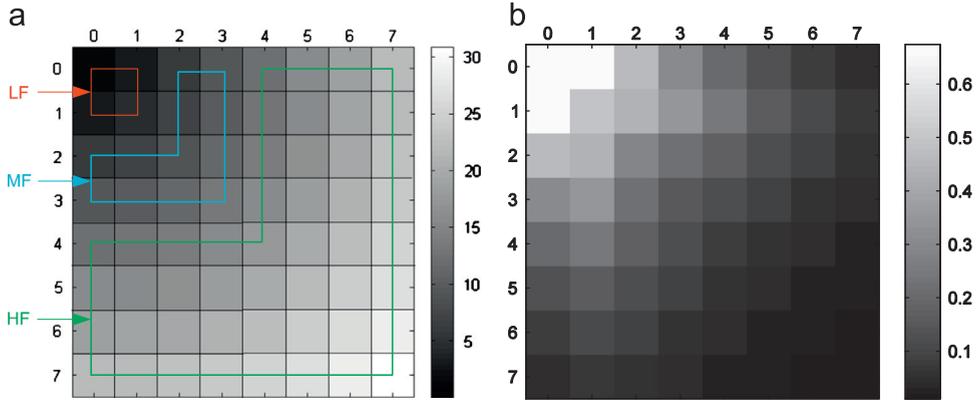


Fig. 5. Frequency  $\omega_{ij}$  and the spatial contrast sensitivity function (CSF) value of each DCT subband. (a) frequency  $\omega_{ij}$  value and (b) spatial CSF value.

These RDCT subbands  $S_4$ ,  $S_5$ , and  $S_6$  (indicated by the blue box) are viewed as the medium-frequency (MF) components with medium sensitivity values. The rest of the RDCT subbands  $S_7$ ,  $S_8$ , and  $S_9$  (denoted by the green box) present the lowest sensitivity values and the highest frequency values larger than 12, which are regarded as the high-frequency (HF) components.

The introduced distortion will not only change the histogram distribution in each RDCT subband and dependencies between adjacent RDCT subbands, but also alter frequency components of the image. For example, if JPEG is utilized to code the reference image, the blocking and ringing artifacts will appear which result from the frequency coefficient truncation. As the quantization steps of the HF components are higher than the LF ones, the HF components will be degraded more seriously than LF ones. In this paper, the image-level FRD is proposed by considering the ratio information between the LF, MF, and HF components. The FRD can be efficiently computed in the RDCT domain, which is defined as:

$$\text{FRD} = \frac{M_{\text{value}} + H_{\text{value}}}{L_{\text{value}}}, \quad (8)$$

where  $L_{\text{value}}$ ,  $H_{\text{value}}$ , and  $M_{\text{value}}$  represent the sums of the absolute DCT coefficient values in the LF ( $S_0$ ,  $S_1$ ,  $S_2$ , and  $S_3$ ), MF ( $S_4$ ,  $S_5$ , and  $S_6$ ) and HF ( $S_7$ ,  $S_8$ , and  $S_9$ ) RDCT subbands, respectively. The FRD can help to capture the frequency proportion changes caused by the distortions. Furthermore, the larger the value of FRD, the more energy the MF and HF components possess. It means that the DCT block is more likely to contain texture information. For the plain block, the energy mostly concentrates in the LF components. For the edge block, there will be only a small number of DCT coefficients in the HF group. Consequently, the texture block will present higher FRD. As discussed in the JND models [44,45,67], the texture block can tolerate more distortions than the plain and edge block, which is interpreted as the texture masking property of the HVS. Therefore, the proposed FRD can be employed to simulate the texture masking property for the derivation of the final image quality metric.

As discussed above, there are total 3 parameters  $\{\alpha, \beta, d_{\text{CBD}}(p, p_{\alpha, \beta})\}$  to depict the histogram distribution of each RDCT subband. Considering the HVS oblique effect,

only the 3 horizontal subbands are included, which results in 9 parameters. For the inter RDCT subband relationship, 8 MI values in total are introduced to depict the parent–child, cousin–child, and brother–child relationships. For the frequency distribution, only one parameter named as FRD is extracted from the reference image. Therefore, the proposed RR method employs  $9 + 8 + 1 = 18$  parameters to represent the reference image. By comparing them with the ones extracted from the distorted image, the perceptual quality can be analyzed.

### 3.2. Visual quality analysis in the receiver side

In the receiver side, we need to compare the extracted features to analyze the perceptual quality of the distorted image. For the parameters extracted from intra RDCT subbands, CBD is employed to depict the distance between the reference image and the distorted one:

$$d_{\text{CBD}}(p, p_d) = \sum_{i=1}^{h_i} |p(i) - p_d(i)|, \quad (9)$$

where  $p$  is the coefficient distribution of the reference image, and  $p_d$  belongs to the corresponding distorted one. However, as the reference image is not available in the receiver side, the distance is approximated by:

$$d_{\text{CBD}}(p, p_d) \triangleq d_{\text{CBD}}(p, p_{\alpha, \beta}) - d_{\text{CBD}}(p_{\alpha, \beta}, p_d), \quad (10)$$

where  $d_{\text{CBD}}(p, p_{\alpha, \beta})$  is the third parameter introduced in the sender side. Therefore, in the receiver side, only  $d_{\text{CBD}}(p_{\alpha, \beta}, p_d)$  needs to be calculated. Their difference will be recorded to represent the statistical feature distance of the intra RDCT subband.

For the inter RDCT subband, the differences between the corresponding MI values of the adjacent RDCT subbands are calculated:

$$d_{\text{MI}}(S_m, S_n) \triangleq I(S_m, S_n) - \hat{I}(S_m, S_n) \quad (11)$$

where  $I(S_m, S_n)$  is the MI of the RDCT subband  $S_m$  and  $S_n$  in the reference image, and  $\hat{I}(S_m, S_n)$  is the MI of  $S_m$  and  $S_n$  in the distorted image.

For the image frequency, as the distortion will degrade the HF, MF, and LF components differently, the FRD distance

can effectively represent the frequency component changes:

$$FL = |\text{FRD}_{\text{ori}} - \text{FRD}_{\text{dis}}|, \quad (12)$$

where  $\text{FRD}_{\text{ori}}$  is the original feature,  $\text{FRD}_{\text{dis}}$  is calculated from the distorted image, and  $FL$  denotes the frequency information change. As discussed before, FRD can represent how much texture information the image contains. Therefore, it can help to simulate the texture masking property of the HVS. Furthermore, as discussed in [48,56], for the content of the original image and the artifacts, one's presence will affect the visibility of the other. Therefore, in this paper, a novel mutual masking strategy is proposed by considering the FRD values of both the original and distorted image:

$$FL_v = \begin{cases} \frac{FL}{FL + \text{FRD}_{\text{ori}}}, & \text{FRD}_{\text{ori}} < \text{FRD}_{\text{dis}} \\ \frac{FL}{FL + \text{FRD}_{\text{dis}}}, & \text{FRD}_{\text{ori}} \geq \text{FRD}_{\text{dis}} \end{cases}, \quad (13)$$

where  $FL_v$  is the final HVS-related features to depict the frequency information change.  $FL$  in the denominator is employed to scale  $FL_v$  into the range [0,1]. When an image containing texture information is smoothed by the distortions, such as JPEG compression and blur, the detailed texture information cannot be perceived by the HVS. Therefore, no visual masking effect should occur. Also if a smooth image is distorted to be highly textured by the distortion, such as additive Gaussian noise and fast-fading in the LIVE database [50], only the noise can be perceived from the degraded image. In this case, there should be no visual masking effect either. This phenomenon is named as the mutual masking [57]. In [56], the mutual masking effect is determined by the minimum value of the thresholds calculated from the original and distorted image. In this paper, as the computed FRD value can depict the texture information of the image, we employ (13) to depict the mutual masking effect of the HVS perception, where the smaller value of  $\text{FRD}_{\text{ori}}$  and  $\text{FRD}_{\text{dis}}$  is employed to model the masking effect. In this way, only the image is highly textured in both the reference and distorted images (large  $\text{FRD}_{\text{ori}}$  and  $\text{FRD}_{\text{dis}}$  values) can produce a significant masking effect. In other cases, an insignificant masking effect will be introduced, as expressed in (13).

Now we have obtained the CBD values of the intra RDCT subbands, MI difference values of the inter RDCT subbands, and the  $FL_v$  value depicting the image frequency information change. How to combine them together for developing an effective RR IQA needs to be considered. In this paper, a simple linear combination method is employed to obtain the final quality values:

$$Q = par_1 \times \sum_{\text{sub}} d_{\text{CBD}}(p^{\text{sub}}, p_d^{\text{sub}}) + par_2 \times \sum_{(m,n)} d_{\text{MI}}(S_m, S_n) + par_3 \times FL_v, \quad (14)$$

where  $(par_1, par_2, par_3)$  are the three weighting parameters to be determined,  $Q$  is the perceptual quality index of the distorted image. First, we sum together the CBD values of intra RDCT subbands, and MI differences of inter RDCT subbands, respectively. Their sum values and the  $FL_v$  value are further linearly combined together according to (14). In order to find the optimal parameters  $(par_1, par_2, par_3)$ , the genetic algorithm [49] is employed to train

them on several distorted images. In this paper, four reference images and their corresponding distorted images in the LIVE image quality assessment database [50] are employed to obtain the three parameter values. The selected four reference images for parameterization are 'rapids', 'paintedhouse', 'plane', and 'building2'. The correlation between the difference mean opinion score (DMOS) values and the calculated  $Q$  values in (14) of the training images is maximized to determine the optimized parameters  $(par_1, par_2, par_3)$ . As there are only 3 parameters to be determined, the number of the genes is equal to 3. Each gene uses 8-bit binary representation. Each gene will be divided by 255 to constrain them within [0,1]. The generation gap is set as 0.9, which means that only  $30 - 30 \times 0.9 = 3$  best fitted genes will be propagated to the successive generation. Therefore, 27 new genes will be produced at each generation. The crossover for creating new genes is a single-point with probability 0.7. And the mutation for creating new genes is with probability 0.0014. The number of generations is set to 100. The initial population is created randomly and uniformly distributed. The fitness assignment is based on ranking instead of raw performance. Selection method is stochastic universal sampling. Reinsertion is fitness-based (instead of uniform random). After performing the genetic algorithm, the parameterization result is  $par_1 = 0.4883$ ,  $par_2 = 0.0313$ , and  $par_3 = 0.6719$ . Furthermore, as in [29,33], a logarithm process is employed to scale the perceptual quality index  $Q$ :

$$Q_{\text{fin}} = \log_{10} \left( 1 + \frac{Q}{D_0} \right), \quad (15)$$

where  $Q_{\text{fin}}$  is the final obtained quality score,  $D_0$  is utilized for scaling the distortion measure to avoid the variation of  $Q$  being too small. It just helps to depict the perceptual quality index clearly, which will not influence the performance of the proposed RR IQA. In this paper,  $D_0$  is set as 0.0001 for simplicity.

## 4. Performances

In this section, the performances of different IQAs are compared to demonstrate the efficiency of the proposed RR IQA for evaluating the image perceptual quality. Subsequently, the statistical significance of our proposed method is discussed. Finally, we will discuss the effectiveness of each component of the proposed RR IQA, specifically, the intra RDCT CBD, the inter RDCT MI differences, and the  $FL_v$  value of the image.

### 4.1. Performance of the proposed RR IQA

We compare the performance of our proposed RR IQA with the representative RR image quality metric: WNISM [29,30], recently developed RR IQA [33], RR-LHS [59], EPSNR [58], RR-DNT [32], RR-SSIM [61,62], RR-Weibull [68], RR-Edge [69], and the FR metrics: PSNR, and SSIM [7]. The LIVE image database [50] (excluding the distorted images generated from the four training reference images), the IRCCyN/IVC image database [4], the MICT image database [66], and the CSIQ image database [70] are employed to compare the performances of these metrics. The most prevailing distortion types have been considered in these image data-

bases, including JPEG, JPEG 2000, blur, white Gaussian noise (WGN), and fast fading (FF). Each distorted image is assigned a subjective score, specifically the DMOS or MOS value, which is obtained from the subjective viewing tests where many observers participated and provided their opinions on the perceptual quality of each distorted image. Therefore, it can be regarded as the ground truth for evaluating the metric performances. We follow the performance evaluation procedure introduced in the video quality experts group (VQEG) HDTV test [53] and that in [54]. Let  $x_j$  represent the perceptual quality index of the  $j$ th distorted image obtained by the corresponding IQA. The five parameter  $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$  monotonic logistic function is employed to map  $x_j$  into  $V_j$ :

$$V_j = \beta_1 \times \left( 0.5 - \frac{1}{1 + e^{\beta_2 \times (x_j - \beta_3)}} \right) + \beta_4 \times x_j + \beta_5. \quad (16)$$

The corresponding five parameters are determined by minimizing the sum squared differences between the mapped objective scores  $V_j$  and the subjective DMOS or MOS values. In order to evaluate the performances, three statistical measurements are employed. Linear correlation coefficient (LCC) defined as:

$$\text{LCC}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (17)$$

measures the correlation.  $x_i$  and  $y_i$  are the sample values, while  $\bar{x}$  and  $\bar{y}$  are the corresponding mean value. The Spearman rank-order correlation coefficient (SROCC) evaluates the prediction monotonicity, which is defined as:

$$\text{SROCC}(X, Y) = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}. \quad (18)$$

$d_i$  is the difference between the  $i$ th image's rank in subjective and the perceptual quality index.  $n$  denotes the number of the total samples. The root mean square prediction error (RMSE) of the fitting process is also introduced to evaluate the efficiencies, which is defined as:

$$\text{RMSE}(X, Y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}. \quad (19)$$

$x_i$  is the subjective value, such as MOS or DMOS value, while  $y_i$  is the objective value generated by each metric after the mapping process defined in (16). Therefore, RMSE measures the difference between the subjective value and the algorithm score after nonlinear regression [54]. After the mapping process, the values obtained from the quality metrics will be in the same range as the subjective quality values. Therefore, the RMSE values will be affected by the ranges of the subjective values. According to the definitions, larger values of LCC and SROCC mean that the objective and subjective scores correlate better, that is to say, a better performance of the IQA. And smaller RMSE values indicate smaller errors between the two groups of scores, therefore a better performance.

The performances of different IQAs over different image subjective quality databases are illustrated in Table 2, where the RR data rate of each IQA is also illustrated. It can be observed that the proposed method can outperform the other RR and FR metrics on the LIVE

[50] and MICT [66] image databases, with larger LCC/SROCC and smaller RMSE value. While for the IRCCyN/IVC [4] image database, only the metrics RR-LHS [59] and RR-SSIM [61,62] can generate better performances. For CSIQ image database [70], only the metric RR-SSIM [61,62] outperforms the proposed RR metric. However, these two RR metrics require a much larger bit rate to represent the RR features than the proposed RR metric. From Table 2, 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 [7], 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 utilizes the whole reference image for quality analysis, which will introduce a heavy burden for the RR feature transmission. For EPSNR [58], in order to reduce the bits to represent the location, the reference and distorted images are first cropped to  $614 \times 454$ , where only the central parts are kept. Therefore, as shown in [58], 19 bits are required to encode the location, while 8 bits are needed to represent the pixel value. In our comparisons, 10 pixels together with their locations are employed as the RR features, which require 270 bits in total for representation. The performances of EPSNR over the three databases seem to be the worst. Although the HVS is sensitive to the edges, 10 edge points are not sufficient to accurately represent the image perceptual quality. If more edge pixels are included, the performance will be better. In that case, a heavy burden for transmitting the RR features will be introduced. For RR-Weibull [68] and RR-Edge [69], as the authors only provide the performance results on the LIVE image database, their performances results on the IRCCyN/IVC, MICT, and CSIQ image databases are not available for comparison. RR-Weibull [68] extracted 6 scalar parameters from each source image to depict the statistics of the image gradient magnitude. It can generate a better performance than WNISM and EPSNR. However, as the number of the RR features is very small, which may not be sufficient to depict the information of the source image, the performance is not good enough, compared with other RR image metrics. RR-Edge [69] further incorporated more RR features to depict the statistics of the edge. In total, 12 RR features are extracted from the source image, which generates a better performance compared with RR-Weibull. However, it is still not good enough. Since RR-LHS [59] considers the motion information to design the RR video quality metric, in our comparisons only the discriminative local harmonic strength in the spatial domain is employed for the RR image quality assessment. Therefore, about 320 elements of each image are extracted as the RR features. If 8 bits are employed to encode each element, the RR data rate is 2560 bits. It is a high burden for the RR data transmission. For RR-DNT [32] and RR-SSIM [61,62], the performance results on the four image databases are illustrated in [62]. All the 779 distorted images in the LIVE image database are employed to demonstrate the performances of RR-

**Table 2**

Performance comparisons of different IQAs over different image subjective quality databases. PSNR, SSIM, WNISM [29,30], EPSNR [58], RR-LHS [59], RR IQA [33], and the proposed method are evaluated on all of these four databases. For RR-Weibull [68] and RR-Edge [69], as the authors only provide the performance results on the LIVE image database, their performances on the IRCCyN/IVC, MICT and CSIQ image databases are not available for comparison. For RR-DNT [32] and RR-SSIM [61,62], the performance results on the four image databases are illustrated in [62]. However, as the proposed method utilized four reference images and the corresponding distorted images for training the parameters, it is not fair to compare RR-DNT and RR-SSIM with the proposed method on the LIVE database. Therefore, only the performance comparisons on IRCCyN/IVC, MICT, and CSIQ image databases are illustrated. (“—” means that the IQA is an FR metric, where the RR feature number is the pixel number of the image, and the RR data rate is also viewed as the whole image. “\*” means that the RR IQA only calculates the number of the features, while the number of the bits for representing the RR parameters cannot be provided.)

Database	Method	LCC	SROCC	RMSE	RR feature number	RR data rate
LIVE [50]	PSNR	0.8759	0.8813	13.157	–	–
	SSIM [7]	0.9041	0.9112	11.653	–	–
	WNISM [29,30]	0.7585	0.7709	17.771	18	162 bits
	EPSNR [58]	0.6571	0.6257	20.559	30	270 bits
	RR-LHS [59]	0.8809	0.8831	12.909	320	2560 bits
	RR-Weibull [68]	0.8567	0.8650	14.475	6	*
	RR-Edge [69]	0.8613	0.8908	14.256	12	96 bits
	RR IQA [33]	0.8984	0.8956	11.975	18	162 bits
	The proposed	<b>0.9309</b>	<b>0.9279</b>	<b>9.965</b>	<b>18</b>	<b>153 bits</b>
IRCCyN/IVC [4]	PSNR	0.7037	0.6791	0.866	–	–
	SSIM [7]	0.7758	0.7778	0.769	–	–
	WNISM [29,30]	0.4525	0.4094	1.087	18	162 bits
	EPSNR [58]	0.3947	0.3958	1.119	30	270 bits
	RR-LHS [59]	0.8078	0.8203	0.718	320	2560 bits
	RR-DNT [32]	0.6316	0.6099	0.9446	48	*
	RR-SSIM [61] [62]	0.8177	0.8156	0.7014	36	*
	RR IQA [33]	0.7057	0.7267	0.863	18	162 bits
	The proposed	<b>0.7712</b>	<b>0.7649</b>	<b>0.776</b>	<b>18</b>	<b>153 bits</b>
MICT [66]	PSNR	0.6154	0.5748	0.987	–	–
	SSIM [7]	0.7174	0.7870	0.872	–	–
	WNISM [29,30]	0.6568	0.6446	0.944	18	162 bits
	EPSNR [58]	0.4016	0.4059	1.146	30	270 bits
	RR-LHS [59]	0.7623	0.7644	0.810	320	2560 bits
	RR-DNT [32]	0.6733	0.6521	0.9253	48	*
	RR-SSIM [61] [62]	0.8051	0.8003	0.7423	36	*
	RR IQA [33]	0.6457	0.6941	0.956	18	162 bits
	The proposed	<b>0.8282</b>	<b>0.8317</b>	<b>0.701</b>	<b>18</b>	<b>153 bits</b>
CSIQ [70]	PSNR	0.7999	0.8049	0.158	–	–
	SSIM [7]	0.8150	0.8368	0.152	–	–
	WNISM [29,30]	0.7302	0.7505	0.179	18	162 bits
	EPSNR [58]	0.6228	0.6537	0.205	30	270 bits
	RR-DNT [32]	0.7009	0.7027	0.1872	48	*
	RR-SSIM [61] [62]	0.8426	0.8527	0.1413	36	*
	RR IQA [33]	0.8135	0.8048	0.153	18	162 bits
	The proposed	<b>0.8197</b>	<b>0.8044</b>	<b>0.150</b>	<b>18</b>	<b>153 bits</b>

DNT and RR-SSIM. However, as the proposed method utilized four reference images and the corresponding distorted images for training the parameters, it is not fair to compare RR-DNT and RR-SSIM with the proposed method on the LIVE database. Therefore, only the performance comparisons on IRCCyN/IVC, MICT, and CSIQ image databases are illustrated in Table 2. RR-DNT [32] employs the divisive normalization to depict the coefficient distributions of the wavelet subbands. The distribution difference between the reference and distorted images is used to depict the image perceptual quality. However, a training process is utilized to determine the 5 parameters in RR-DNT. And the performances of RR-DNT seem to be sensitive to these parameters. That is the reason why RR-DNT performs very well over the LIVE image database [32], while performs poorly over the IRCCyN/IVC, MICT,

and CSIQ image databases. RR-SSIM [61,62] extracted the statistical features from a multi-scale, multi-orientation divisive normalization transform. By following the philosophy in the construction of SSIM, a distortion measurement is developed to estimate the SSIM index of the distorted image. As a linear relationship between the RR-SSIM and SSIM has been discovered, the performances of RR-SSIM are good, which is comparable with the proposed method. RR-SSIM outperforms the proposed metric on the IRCCyN/IVC and CSIQ databases, while its performance is worse than the proposed one on the MICT database. However, RR-SSIM extracted 36 RR features to represent the source image, which is twice of that extracted by the proposed RR metric. WNISM [29,30] is proposed in the wavelet domain by depicting the marginal probability distribution of each wavelet subband. The steerable

pyramid is first employed to decompose the image into several wavelet subbands, whose coefficient distributions are modeled by GGD. As demonstrated in [33], GGD can more accurately model the coefficient distribution of RDCT subband than that of the steerable pyramid. Moreover, KLD is utilized in WNISM to depict the histogram distribution distance. However, KLD is asymmetric, which is not suitable for image quality evaluation, because the visual quality distance from one image to another should be identical no matter how it is measured. Those are the reasons why WNISM performs badly over the four image quality databases, as illustrated in Table 2. For the RR data rate, each wavelet subband needs 3 parameters to describe its distribution, which requires  $8+8+8+3=27$  bits for representing these parameters. In total 6 wavelet subbands are considered to construct WNISM, which results in  $27 \times 6=162$  bits to encode all the RR features of the reference image. In order to handle the drawbacks of WNISM, the RR IQA [33] employed the DCT reorganization strategy and the CBD to depict the histogram distribution difference. As demonstrated in [33] and Table 2, a better performance over LIVE and CSIQ image database can be achieved. However, it performs poorly over the other two image databases. The reason is that only the intra RDCT subband correlation is exploited in [33], but the inter RDCT subband dependencies and the frequency distribution over the image are not considered. Same as WNISM, 162 bits are needed for representing the RR features extracted for the RR IQA [33]. For the proposed method, the inter RDCT subband correlation is depicted by the MI values, and the frequency distribution is captured by FRD, which can further simulate the HVS texture masking property. As the

three horizontal RDCT subbands are employed for depicting the intra RDCT subband distribution,  $(8+8+8+3) \times 3=81$  bits are needed to represent the GGD modeling parameters. For the MI and FRD values, 8-bit representation is employed. Therefore,  $8 \times 9=72$  bits are needed for representing all the MI and FRD values. In total,  $81+72=153$  bits are required to encode all the RR features extracted for the proposed method. The scatter-plots of different IQAs over the LIVE image database are illustrated in Fig. 6. And the scatter plots of the proposed RR metric on the four image subjective quality database are illustrated in Fig. 7. It can be observed that the points of the proposed method scatter more closely to the fitted line, compared with other IQAs. It means that the DMOS or MOS values correlate better with the perceptual quality values obtained by the proposed RR IQA.

Furthermore, we test the proposed RR IQA over individual distortion types from the LIVE image database, which are illustrated in Table 3. It can be observed that PSNR performs well over JPEG 2000 and WGN images, especially for WGN images. However, for the JPEG, Blur, and FF noise images, PSNR performs poorly. EPSNR only employs several edge pixels to measure the corresponding PSNR. Therefore, it presents a performance similar to PSNR. For the WGN images, EPSNR demonstrates a very good performance. It means that the perceptual qualities of WGN images correlate closely with the absolute pixel value differences, in contrast to other noise images. RR-LHS demonstrates good performances on the JPEG 2000, JPEG, WGN, and FF noise images. However, its performance over the Blur noise images is very poor, even a very large number of RR features has been employed. It means

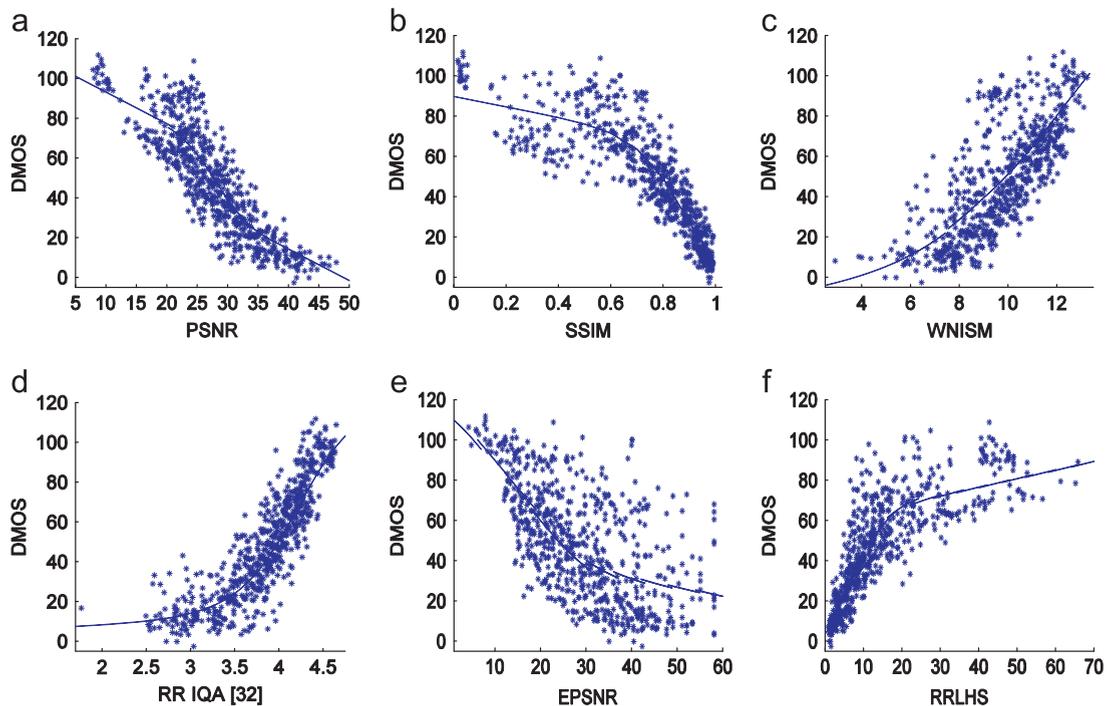


Fig. 6. Scatter plots of the DMOS values versus model predictions on the LIVE image quality assessment database. Each sample point represents one test image. (a) PSNR; (b) SSIM [7]; (c) WNISM [32]; (d) RR IQA [33]; (e) EPSNR [58] and (f) RR-LHS [59].

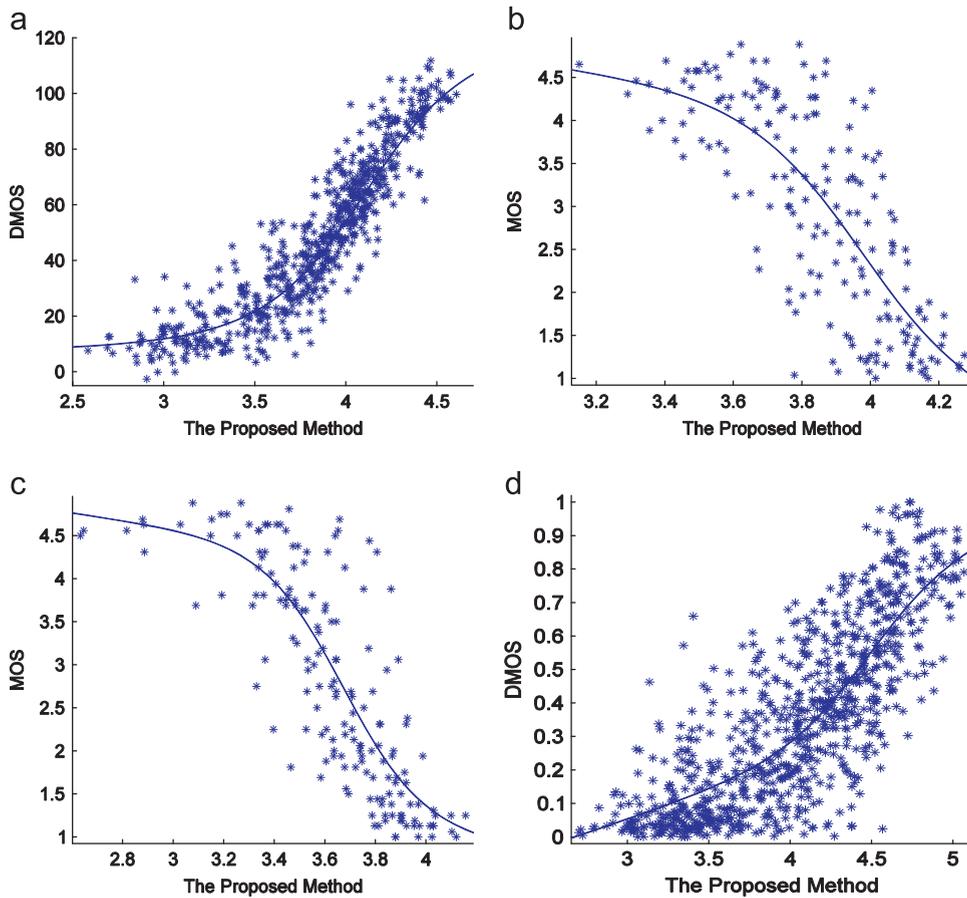


Fig. 7. Scatter plots of the DMOS or MOS values versus model predictions on the three image subjective quality databases. Each sample point represents one test image. (a) LIVE image database; (b) IRCCyN/IVC image database; (c) MICT image database and (d) CSIQ image database.

**Table 3**  
Performances of different IQAs over individual distortion types on the LIVE image database.

		JPEG 2000	JPEG	WGN	Blur	FF
PSNR	LCC	0.9078	0.8942	0.9857	0.7856	0.8880
	SROCC	0.9042	0.8853	0.9850	0.7894	0.8897
	RMSE	10.546	14.406	4.711	11.214	12.898
WNISM [29,30]	LCC	0.9270	0.8629	0.8791	0.9234	0.9422
	SROCC	0.9211	0.8539	0.8572	0.9290	0.9350
	RMSE	9.434	16.258	13.346	6.955	9.399
EPSNR [58]	LCC	0.6773	0.6489	0.9700	0.4890	0.6129
	SROCC	0.6816	0.6400	0.9670	0.3086	0.5612
	RMSE	18.500	24.482	6.776	15.807	22.167
RR-LHS [59]	LCC	0.8861	0.9761	0.9345	0.6051	0.8569
	SROCC	0.8792	0.9557	0.9846	0.6250	0.8575
	RMSE	11.654	6.995	9.970	14.427	14.462
RR-Weibull [68]	LCC	0.9422	0.9493	0.9771	0.9471	0.9234
	SROCC	0.9415	0.9402	0.9749	0.9404	0.9261
	RMSE	7.912	10.115	5.954	5.817	10.741
RR-Edge [69]	LCC	0.9404	0.9383	0.8815	0.9152	0.9421
	SROCC	0.9406	0.9408	0.8654	0.9083	0.9329
	RMSE	8.592	11.128	13.224	7.302	9.400
RR IQA [33]	LCC	0.8335	0.9363	0.9056	0.9062	0.9342
	SROCC	0.8242	0.9239	0.8907	0.9202	0.9148
	RMSE	13.893	11.299	11.879	7.662	10.008
The proposed method	LCC	0.8983	0.9528	0.9275	0.9459	0.9437
	SROCC	0.8912	0.9520	0.9093	0.9525	0.9204
	RMSE	11.051	9.766	10.471	5.880	9.277

that the discriminative local harmonic strength is not suitable for depicting the perceptual quality of the Blur noise images. For WNISM, RR-Weibull, and RR-Edge, the experimental results over the individual distortion types are very good. However, their performances degrade significantly when images with different types of distortions are tested together, as shown in Table 2. As revealed by the previous literature [32], it is also the main drawback of WNISM. In order to handle the drawback, RR IQA [33] is proposed. The experimental results demonstrate that RR IQA [33] outperforms WNISM except for the JPEG 2000 distortion. Actually, JPEG 2000 employed the wavelet transform for compression. Therefore, the steerable pyramid employed in WNISM is more suitable for depicting the coefficient distribution than the DCT. For the proposed method, as new RR features, specifically the inter RDCT subband MI and image FRD  $FL_v$ , have been introduced, the performance over the JPEG 2000 noise images has been greatly improved. Therefore, the proposed method can not only perform very well over individual distortion types, but also provide a good performance across different distortion types. It means that it performs more robustly for evaluating image visual quality. Furthermore, the proposed metric maintains a smaller RR data rate, compared with RR IQA [33], WNISM, and RR-LHS. The improvements have demonstrated that the inter RDCT subband dependencies and the image  $FL_v$  value are helpful for designing an effective RR IQA. It reflects that the MI differences and  $FL_v$  value can help to depict the levels of the introduced distortions. Therefore, for the proposed RR IQA, the RR features for depicting the vertical RDCT subbands [33] are excluded to save some bit rates for the inter RDCT subband MI values and image FRD value.

As illustrated in Tables 2 and 3, the effectiveness of our proposed RR quality metric has been clearly demonstrated compared with the other RR metrics or even FR metrics in terms of both performance and required RR data rate. The computational complexities of RR feature extraction and comparison need to be further evaluated. The processing complexity in the sender side is different from that in the receiver side. In the sender side, as introduced in Section 3.1,  $8 \times 8$  block-based DCT is first performed on the source image. After the reorganization strategy, the DCT subbands are grouped into several

representative RDCT subbands. The DCT coefficient distribution of each RDCT subband is modeled by the GGD. MI is employed to depict the relationship between different RDCT subbands. Based on the RDCT subband, the image FRD  $FL_v$  is calculated. We implement the RR feature extraction in Matlab. During our implementation, we do not perform any optimizations. A speed test is performed on our PC with a 3.0 GHz Quad CPU and 1.0 GB memory. For each source image of LIVE image database, it only requires 2.94 s on average to extract the RR features. In the receiver side, as illustrated in Section 3.2, the  $8 \times 8$  block-based DCT and reorganization strategy is also performed. But the fitting process of the GGD does not need to be performed. Only the histogram of each RDCT subband is constructed. And MI values between RDCT subbands, and the image FRD  $FL_v$  are calculated. Therefore, the computation is faster. The speed test is performed on the same PC, which indicates that only 1.93 s per image on average is needed for the image quality analysis. If further optimization is applied, it is believed that the quality analysis in the receiver side can perform even faster.

#### 4.2. Statistical significance

To assess the statistical significance of the performance difference between two metrics, F-test is conducted on the prediction residuals between the metric outputs (after nonlinear mapping) and the subjective ratings. The residuals are supposed to be Gaussian. Smaller residual variance implies more accurate prediction. Let  $F$  denote the ratio between the residual variances of two different metrics (with the larger variance as the numerator). If  $F$  is larger than  $F_{\text{critical}}$  which is calculated based on the number of residuals and a given confidence level, then the difference between the two metrics are considered to be significant at the specified confidence level. Table 4 lists the residual variance of each metric on the four subjective image databases. Notably due to the differences in employed subjective scales, the residual variance varies a lot across different image databases. The  $F_{\text{critical}}$  with 95% confidence is also shown in Table 4 for each database. In Table 5, the proposed metric is compared with the other metrics regarding the statistical significance.

**Table 4**  
Residual variances of the IQAs on the four image subjective quality databases.

	LIVE (672 images) $F_{\text{critical}}=1.1355$	IRCCyN/IVC (185 images) $F_{\text{critical}}=1.275$	MICT (168 images) $F_{\text{critical}}=1.291$	CSIQ (866 images) $F_{\text{critical}}=1.1185$
PSNR	173.3645	0.7534	0.9804	0.0249
SSIM [7]	136.0017	0.5942	0.7647	0.0232
RR-LHS [59]	166.8887	0.5186	0.6600	–
EPSNR [58]	423.3052	1.2599	1.3213	0.0422
RR-Weibull [68]	209.8357	–	–	–
RR-Edge [69]	203.5310	–	–	–
WNSIM [29,30]	333.7304	1.1869	0.8958	0.0322
RR IQA [33]	159.3776	0.7485	0.9186	0.0226
The proposed	99.6236	0.6049	0.4948	0.0233

**Table 5**

Performance comparisons regarding the statistical significance. In each entry, the symbol “1”, “0” or “=” means that on the image database the proposed RR metric is statistically (with 95% confidence) better, worse or indistinguishable in comparison to its competitor. “\*” means that the comparison cannot be performed due to the unavailable result data.

	LIVE (672 images) $F_{\text{critical}}=1.1355$	IRCCyN/IVC (185 images) $F_{\text{critical}}=1.275$	MICT (168 images) $F_{\text{critical}}=1.291$	CSIQ (866 images) $F_{\text{critical}}=1.1185$
PSNR	1	=	1	=
SSIM [7]	1	=	1	=
RR-LHS [59]	1	=	1	*
EPSNR [58]	1	1	1	1
RR-Weibull [68]	1	*	*	*
RR-Edge [69]	1	*	*	*
WNSIM [29,30]	1	1	1	1
RR IQA [33]	1	=	1	=

In each entry, the symbol “1”, “0”, or “=” means that on the image databases indicated by the first row of the table, the proposed metric is statistically (with 95% confidence) better, worse, or indistinguishable, respectively, when compared with its competitors indicated by the first column. “\*” means that the comparison cannot be performed due to the unavailable result data. For the RR metrics RR-Weibull and RR-Edge, the metric outputs of the distorted images on the IRCCyN/IVC, MICT, and CSIQ image databases are not available. Also we cannot fit the RR-LHS outputs to the obtained DMOS value using (16) for the CSIQ image database. Therefore, we cannot compare the statistical significances of these metrics with the proposed method on the corresponding image databases. By referring to the other entry values shown in Table 5, it can be observed that the proposed metric outperforms most of its competitors statistically. Although its performances on IRCCyN/IVC, and CSIQ image databases seem to be equivalent to other IQAs, overall it demonstrates better performances on the other two image databases.

#### 4.3. Performance analysis of each component

As we have mentioned before, the intra RDCT subband correlation, the inter RDCT subband dependency, and the image frequency distribution are utilized to design the RR IQA. In this section, we will try to figure out the contribution of each component to the final performance.

Table 6 illustrates the individual performance of each component of the proposed RR metric over the image databases. For the CBD values of intra RDCT subbands, only three horizontal RDCT subbands are considered. Therefore, as 3 parameters are required to depict the coefficient distribution,  $3 \times 3 = 9$  parameters are extracted for the RR features of intra RDCT subband correlation. It can be observed that the performance is comparable with the RR IQA in [33]. Actually, it matches the HVS oblique effect, which means the HVS presents similar sensitivity to the horizontal and vertical information. Therefore, by considering only the horizontal ones, the visual quality of the distorted image can be accurately depicted. For the MI difference of inter RDCT subbands, 8 MI values are employed to depict the parent–child, cousin–child, and brother–child dependencies. The performance outperforms the WNISM, while it only requires a far smaller number of RR features

**Table 6**

Performance of each component of the proposed RR metric.

		CBD	MI	$FL_v$
LIVE [50]	LCC	0.8983	0.7746	0.8770
	SROCC	0.8943	0.7697	0.8809
	RMSE	11.981	17.248	13.106
IRCCyN/IVC [4]	LCC	0.7004	0.6956	0.7324
	SROCC	0.6906	0.6656	0.7227
	RMSE	0.870	0.875	0.830
MICT [66]	LCC	0.6713	0.4787	0.6259
	SROCC	0.6806	0.4435	0.7120
	RMSE	0.927	1.099	0.976

(8 parameters vs. 18 parameters of WNISM). Additionally, it can be observed that the MI differences perform worse than the CBD values. The reason is that the correlations between inter RDCT subbands have been essentially ensured by the linear transformations. Therefore, compared with the coefficient distribution in each RDCT subband, the MI values between different subbands vary less significantly, thus cannot effectively depict the image distortions. However, the introduced distortion in the image will affect the MI values between RDCT subbands. Therefore, it is necessary to incorporate the inter RDCT subband dependencies in designing the RR IQA, which plays a less but nevertheless an important role in image quality assessment. For the  $FL_v$  of the image, the performance is very good. Even with only one parameter FRD extracted from the reference image, the performance is comparable with the RR IQA [33] and PSNR, and even better than WNISM, RR-Weibull, and RR-Edge. Therefore, if we want to further reduce the RR data rate, we can extract the FRD only and transmit it to the receiver side for perceptual quality analysis. It only requires 8 bits to represent the FRD of the reference image. The good performance may attribute to two reasons. First, the distortions introduced will significantly change the frequency distribution of the image. The larger the FRD changes, the higher the distortion level. For example, the more compression is introduced for JPEG coded image, the more HF and MF components are discarded, compared with the LF ones. The FRD differences as in (12) will become larger, which indicates

**Table 7**

Performances of the combinations of different components of the proposed metric.

		CBD+MI	CBD+ $FL_v$	$FL_v$ +MI
LIVE [50]	LCC	0.9100	0.9206	0.9114
	SROCC	0.9050	0.9194	0.9127
	RMSE	11.309	10.652	11.220
IRCCyN/IVC [4]	LCC	0.7151	0.7632	0.7342
	SROCC	0.7059	0.7550	0.7235
	RMSE	0.852	0.787	0.827
MICT [66]	LCC	0.6764	0.8126	0.7208
	SROCC	0.6874	0.8186	0.7132
	RMSE	0.922	0.729	0.867

**Table 8**

Performances of the combinations of different components of the proposed metric over individual distortion type.

		JPEG 2000	JPEG	WGN	Blur	FF
CBD+MI	LCC	0.8461	0.9347	0.9141	0.9245	0.9394
	SROCC	0.8400	0.9241	0.9010	0.9331	0.9177
	RMSE	13.403	11.438	11.357	6.907	9.617
CBD+ $FL_v$	LCC	0.8908	0.9556	0.9150	0.9320	0.9383
	SROCC	0.8821	0.9540	0.8988	0.9406	0.9170
	RMSE	11.426	9.479	11.296	6.568	9.703
$FL_v$ +MI	LCC	0.9315	0.9586	0.9294	0.9652	0.9288
	SROCC	0.9223	0.9597	0.9144	0.9666	0.9268
	RMSE	9.149	9.160	10.331	4.736	10.395

worse perceptual quality. Therefore, the FRD difference can depict the distortion level. Second, the mutual masking strategy is employed as formulated in (13). As discussed in [48], for the content of the original image and the artifacts, one's presence will affect the visibility of the other. Therefore, by using mutual masking, the texture masking effect of the HVS can be more accurately simulated.

In order to further demonstrate the contribution of each component of the proposed RR metric, different combinations of these components are evaluated on the image databases, as well as over each individual distortion type. The experimental results are illustrated in Tables 7 and 8. It can be observed that the proposed method can outperform all of these different combinations. It means that each component of our proposed RR metric does contribute to the final performance. Comparing the three combinations, we can see that CBD+ $FL_v$  can ensure the best performance. It is also consistent with the performances illustrated in Table 7, where CBD and  $FL_v$  perform better than MI. However, CBD+ $FL_v$  is still not as good as the proposed RR metric. Therefore, the MI is still a necessary component that contributes to the performance improvement of the proposed RR metric. In this case, if a very small RR data rate is required, we can only extract  $FL_v$  and transmit it to the receiver side for perceptual quality analysis. With the increasing of the required RR data rate, we can further transmit the CBD RR features to the receiver side. Finally, if the RR data rate is sufficient,

all the three components, specifically the  $FL_v$ , CBD, and MI, will be extracted and transmitted to the receiver side for a better performance.

## 5. Conclusion

In this paper, we propose a novel RR IQA by considering the intra and inter subband correlations in the RDCT domain. The CBD and MI values are first employed to depict the intra and inter RDCT relationships, respectively. The FRD calculated in RDCT domain depicts the frequency distribution of the images, which can be further employed to simulate the HVS texture masking effect in a mutual masking way. Combining the CBD values, MI differences, and FRD value together, an effective RR IQA is developed. Evaluations on several image quality databases demonstrate that the proposed method outperforms the state-of-the-art RR metrics and even FR metrics PSNR and SSIM. It means that the proposed metric correlates well with the human perception of the image quality. Meanwhile, only a small number of RR features are extracted. Furthermore, we can extend the proposed RR image quality metric to the RR video quality metric. The video sequences are encoded by the block-based DCT, such as MPEG-2 and H.264. Since the proposed metric is based on block-based DCT, it can easily capture the distortion due to the compression. Hence, the proposed RR image can be easily applied to the video sequence and employed for monitoring video perceptual quality during the transmission.

## Acknowledgement

This work was supported in part by a grant from the Research Grants Council of Hong Kong, under Project 415712. The authors are grateful to Prof. Xuanqin Mou and Mr. Wufeng Xue for providing their results for experiment comparisons. Thank Prof. Irwan Prasetya Gunawan and Mr. Kai Zeng for their valuable discussions.

## References

- [1] W. Lin, C.-C.J. Kuo, Perceptual visual quality metrics: a survey, *Journal of Visual Communication and Image Representation* 22 (4) (2011) 297–312.
- [2] Z. Wang, H.R. Sheikh, A.C. Bovik, Objective video quality assessment, in: B. Furht, O. Marqure (Eds.), *The Handbook of Video Databases: Design and Application*, CRC Press, 2003, pp. 1041–1078. Sept.
- [3] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, M. Etoh, Cross-dimensional perceptual quality assessment for low bit-rate videos, *IEEE Transactions on Multimedia* 10 (7) (2008) 1316–1324.
- [4] P. Le Callet, F. Atrousseau, Subjective Quality Assessment IRCCyN/IVC database. [Online]. Available: <<http://www.irccyn.ec-nantes.fr/ivcdb/>>.
- [5] Z. Wang, A.C. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool, New York, 2006.
- [6] B. Girod, What's Wrong With Mean-squared Error, in *Digital Images and Human Vision*, MIT Press, Cambridge, MA, 1993 207–220.
- [7] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing* 13 (4) (2004) 600–612.
- [8] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, *IEEE Transactions on Image Processing* 16 (9) (2007) 2284–2298.

- [9] K. Yang, C.C. Guest, K.E. Maleh, P.K. Das., Perceptual temporal quality metric for compressed video, *IEEE Transactions on Multimedia* 9 (7) (2007) 1528–1535.
- [10] L. Ma, S. Li, K.N. Ngan, Visual horizontal effect for image quality assessment, *IEEE Signal Processing Letters* 17 (7) (2010) 627–630.
- [11] F. Zhang, L. Ma, S. Li, K.N. Ngan, Practical image quality metric applied to image coding, *IEEE Transactions on Multimedia* 13 (4) (2001) 615–624.
- [12] F. Zhang, W. Liu, W. Lin, K.N. Ngan, Spread spectrum image watermarking based on perceptual quality metric, *IEEE Transactions on Image Processing* 20 (11) (2011) 3207–3218.
- [13] S.S. Hemami, A.R. Reibman, No-reference image and video quality estimation: applications and human-motivated design, *Signal Processing: Image Communication* 25 (7) (2010) 469–481.
- [14] H.R. Sheikh, A.C. Bovik, L. Cormack, No-reference quality assessment using nature scene statistics: JPEG 2000, *IEEE Transactions on Image Processing* 14 (11) (2005) 1918–1927.
- [15] 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 Processing: Image Communication* 25 (7) (2010) 502–516.
- [16] T. Brandao, M.P. Queluz, No-reference image quality assessment based on DCT domain statistics, *Signal Processing* 88 (4) (2008) 822–833.
- [17] Z. Wang, H. R. Sheikh, and A. C. Bovik, No-reference perceptual quality assessment of JPEG compressed images, *Proc. IEEE International Conference on Image Processing* Sept. 2002.
- [18] R. Ferzli, L.J. Karam, A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB), *IEEE Transactions on Image Processing* 18 (4) (2009).
- [19] Z. Wang, A.C. Bovik, Reduced- and no-reference image quality assessment: the natural scene statistic model approach, *IEEE Signal Processing Magazine* 28 (2011) 29–40.
- [20] S. Wolf, M. H. Pinson, Low bandwidth reduced reference video quality monitoring system, *Proc. International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, AZ, 2005.
- [21] P. Le Callet, C. V. Gaudin, and D. Barba, Continuous quality assessment of MPEG2 video with reduced reference, *Proc. International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, AZ, Jan. 2005.
- [22] S. Wolf, M.H. Pinson, Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system, *Proceedings of SPIE* 3845 (1999) 266–277.
- [23] S. Yang, Reduced reference MPEG-2 picture quality measure based on ratio of DCT coefficients, *Electronics Letters* 47 (6) (2011) 382–383.
- [24] P. Le Callet, C.V. Gaudin, D. Barba, A convolutional neural network approach for objective video quality assessment, *IEEE Transactions on Neural Networks* 17 (5) (2006) 1316–1327.
- [25] M. Carnec, P. Le Callet, D. Barba, An image quality assessment method based on perception of structural information, *Proc. IEEE International Conference on Image Processing* 3 (2003) 185–188.
- [26] M. Carnec, P. Le Callet, D. Barba, Visual features for image quality assessment with reduced reference, *Proc. International Conference on Image Processing* 1 (2005) 421–424.
- [27] D. Tao, X. Li, W. Lu, X. Gao, Reduced-reference IQA in contourlet domain, *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics* 39 (6) (2009) 1623–1627.
- [28] U. Engelke, M. Kusuma, H.J. Zepernick, M. Caldera, Reduced-reference metric design for objective perceptual quality assessment in wireless imaging, *Signal Processing: Image Communication* 24 (7) (2009) 525–547.
- [29] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E. Yang, A.C. Bovik, Quality-aware images, *IEEE Transactions on Image Processing* 15 (6) (2006) 1680–1689.
- [30] Z. Wang, E.P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, *Proc. SPIE, Human Vision and Electronic Imaging* (2005).
- [31] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference video quality assessment of compressed video sequences", *IEEE Transactions on Circuits and Systems for Video Technology*, accepted.
- [32] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, *IEEE Journal of Selected Topics in Signal Processing* 3 (2) (2009) 201–211.
- [33] L. Ma, S. Li, F. Zhang, K.N. Ngan, Reduced-reference image quality assessment using reorganized DCT-based image representation, *IEEE Transactions on Multimedia* 13 (4) (2011) 824–829.
- [34] R. Soundararajan, A. C. Bovik, RRED indices: reduced reference entropic differencing framework for image quality assessment, *Proc. International Conference on Acoustics, Speech, and Signal Processing*, 2011.
- [35] L. Ma, S. Li, K. N. Ngan, Reduced-reference image quality assessment via intra- and inter-subband statistical characteristics in reorganized DCT domain, *Proc. Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, Oct., 2011.
- [36] J.A. Redi, P. Gastaldo, I. Heynderickx, R. Zunino, Color distribution information for reduced-reference assessment of perceived image quality, *IEEE Transactions on Circuits and Systems for Video Technology* 20 (12) (2010) 1757–1769.
- [37] Z. Xiong, K. Ramchandran, M.T. Orchard, Y.Q. Zhang, A comparative study of DCT- and wavelet-based image coding, *IEEE Transactions on Circuits and Systems for Video Technology* 9 (5) (1999) 692–695.
- [38] D. Zhao, W. Gao, Y.K. Chan, Morphological representation of DCT coefficients for image compression, *IEEE Transactions on Circuits and Systems for Video Technology* 12 (9) (2002) 819–823.
- [39] T.M. Cover, J.A. Thomas, *Element of Information Theory*, Wiley, New York, 1991.
- [40] R.W. Buccigrossi, E.P. Simoncelli, Image compression via joint statistical characterization in the wavelet domain, *IEEE Transactions on Image Processing* 8 (12) (1999) 1688–1701.
- [41] J. Liu, P. Moulin, Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients, *IEEE Transactions on Image Processing* 10 (11) (2001) 1647–1658.
- [42] S. Lyu, E.P. Simoncelli, Nonlinear image representation using divisive normalization, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2008).
- [43] S. Lyu, Divisive normalization: justification and effectiveness as efficient coding transform, *Advances in Neural Information Processing Systems (NIPS)* (2010).
- [44] X. Zhang, W. Lin, P. Xue, Improved estimation for just-noticeable visual distortion, *Signal Processing* 85 (2005) 795–808.
- [45] X. Zhang, W. Lin, P. Xue, Just-noticeable difference estimation with pixels in images, *Journal of Visual Communication and Image Representation* 19 (2008) 30–41.
- [46] Z. Wei, K.N. Ngan, Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain, *IEEE Transactions on Circuits and Systems for Video Technology* 19 (3) (2009) 337–346.
- [47] ITU-R Recommendation BT.500-11, Methodology for the subjective assessment of the quality of television pictures, *International Telecommunications Union*, Tech. Rep., 2002.
- [48] S. Li, F. Zhang, L. Ma, K.N. Ngan, Image quality assessment by separately evaluating detail losses and additive impairments, *IEEE Transactions on Multimedia* 13 (5) (2011) 935–949.
- [49] Genetic Algorithm Toolbox. [Online]. Available: <<http://www.shef.ac.uk/acse/research/ecrg/gat.html>>.
- [50] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, LIVE Image Quality Assessment Database. [Online]. Available: <<http://live.ece.utexas.edu/research/quality>>.
- [51] B. Li, M.R. Peterson, R.D. Freeman, Oblique effect: a neural basis in the visual cortex, *Journal of Neurophysiology* (2003) 204–217.
- [52] M.N. Do, M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback–Leibler distance, *IEEE Transactions on Image Processing* 11 (2) (2002) 146–158.
- [53] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment, 2000. [Online]. Available: <<http://www.vqeg.org>>.
- [54] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on Image Processing* 15 (11) (2006) 3440–3451.
- [55] Y. Niu, M. Kyan, L. Ma, A. Beghdadi, and S. Krishnan, A visual saliency modulated just noticeable distortion profile for image watermarking, *European Signal Processing Conference*, 2011.
- [56] A.P. Bradley, A wavelet visible difference predictor, *IEEE Transactions on Image Processing* 8 (5) (1999) 717–730.
- [57] S. Daly, The visible difference predictor: an algorithm for the assessment of image fidelity, in: A.B. Watson (Ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, MA, 1993, pp. 179–206.
- [58] ITU-T Recommendation J.246, Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference, Aug. 2008. [Online] Available: <<http://www.itu.int/rec/T-REC-J.246/en>>.
- [59] I.P. Gunawan, M. Ghanbari, Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration, *IEEE Transactions on Circuits and Systems for Video Technology* 18 (1) (2010) 71–83.
- [60] K.T. Tan, M. Ghanbari, Blockiness detection for MPEG2-coded video, *IEEE Signal Processing Letters* 7 (8) (2000) 213–215.

- [61] A. Rehman, Z. Wang, Reduced-reference SSIM estimation, in Proc. IEEE International Conference on Image Processing, Sept. 26-29, 2010.
- [62] A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Transactions on Image Processing, accepted, to appear 2012.
- [63] D. Austin, Image compression: seeing what's not there, Feature Column, American Mathematical Society, [Online]. Available: <<http://www.ams.org/samplings/feature-column/fcarc-image-compression>>.
- [64] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, Shiftable multi-scale transforms, IEEE Transactions on Information Theory 38 (2) (1992) 587–607.
- [65] E.P. Simoncelli, W.T. Freeman, The steerable pyramid: a flexible architecture for multi-scale derivative computation, IEEE International Conference on Image Processing (1995) 444–447.
- [66] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. Parvez, MICT Image Quality Evaluation Database. [Online]. Available: <<http://mict.eng.u-toyama.ac.jp/mictdb.html>>.
- [67] L. Ma, K.N. Ngan, F. Zhang, S. Li, Adaptive block-size transform based just-noticeable difference model for images/videos, Signal Processing: Image Communication 26 (3) (2011) 162–174.
- [68] W. Xue, X. Mou, Reduced reference image quality assessment based on weibull statistics, Proc. International Workshop on Quality of Multimedia Experience (QoMEX), Jun. 2010.
- [69] M. Zhang, W. Xue, X. Mou, Reduced reference image quality assessment based on statistics of edge, Proceedings of SPIE 7876 (2011) 787611.
- [70] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, Journal of Electronic Imaging 19 (1) (2010).