Reduced-reference Image Quality Assessment based on Free-Energy Principle with Multi-channel Decomposition

Wenhan Zhu†, Guangtao Zhai†, Yutao Liu♯, Ning Liu† and Xiaokang Yang†
†Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China
♯Department of Computer Science, Harbin Institute of Technology, Harbin, China
Email: {zhuwenhan823,zhaiguangtao}@sjtu.edu.cn,liuyutao2008@gmail.com,{ningliu,xkyang}@sjtu.edu.cn

***Abstract*—The free-energy principle studied in brain theory and neuroscience accounts for the mechanism of perception and understanding in human brain, which is highly adapted for measuring the visual quality of perceptions. On the other hand, psychologists and neurologists report that different frequency and orientation components of one stimulus arouse different neurons in striate cortex. In this paper, a novel reduce-reference (RR) image quality assessment (IQA) metric based on free-energy principle in multi-channel is proposed, which is called MCFEM (Multi-Channel Free-Energy principle Metric). We first decompose the input reference image and distorted image via a two-level discrete Haar wavelet transform (DHWT). Next, the free-energy features of each subband images are computed based on sparse representation. Finally, an overall quality index is received through the support vector regressor (SVR). Extensive experimental comparisons on four (LIVE, CSIQ, TID2008 and TID2013) benchmark image databases demonstrate that the proposed method is highly competitive with the representative RR and no-reference models as well as full-reference ones.**

I. Introduction

With the advent of the information age, visual media have been gradually became an indispensable part of daily life. Most of images in digital applications are meant for human consumers, and images with different visual quality bring viewers great experience difference. It is significant to measure the quality of images for designing better user systems of visual experience. Image quality assessment (IQA) dedicates to figure out the problem of how to create quantitative model to automatically predict image quality in terms of visual perception.

Objective IQA algorithms can be generally classified into full-reference (FR), reduced-reference (RR) and no-reference (NR) algorithms based on the available information of reference image. For FR methods, the pristine images can be referred when evaluating the distorted images. For RR algorithms, partial information of reference images are available, while NR methods assess distorted images without corresponding reference images. FR IQA algorithms have been comprehensively developed over the past two decades. The peak signal-to-noise ratio (PSNR) is a popular and widely used method purposed long before. The structural similarity index (SSIM) [1], the visual information fidelity (VIF) [2], and the feature similarity index (FSIM) [3] are also several successful examples of FR metrics. In addition, the progress on NR IQA is rapid and brillant in recent years. On account of methodology of the measures, the NR IQA algorithms can be divided into three types, which are natural scene statistics (NSS)-based, learning-based, and human visual system (HVS)-based measures. In literature, some effective NR metrics have been created, e.g., distortion identification-based image verity and integrity evaluation (DIIVINE) [4] and natural image quality evaluator (NIQE) [5] in NSS-based methods, Codebook representation for NR IQA (CORNIA) [6] in learning-based methods and NR free-energy based robust metric (NFERM) [7] in HVS-based methods, etc.

Apart from FR and NR IQA, RR IQA is a tradeoff solution, which only employs some amount of information from the pristine image in quality prediction. Generally, the RR algorithms are designed from two categories: spatial domain and transform domain. In spatial domain, Liu et al. [8] composite the bottom-up and top-down features as reference data; Wu et al. [9] decompose images into orderly and disorderly parts to compute information fidelities; Redi et al. [10] extract color correlogram features to predict the image quality. In transform domain, Wang et al. [11] use the features from NSS; Soundararajan et al. [12] utilize the wavelet coefficients of images to estimate their qualities; Golestaneh et al. [13] employ the entropy of DWT coefficients of image gradient map as reference data. A good RR method should keep a balance between the size of features and the accuracy of predictions of image quality. In [14], Zhai et al. propose a new psychovisual image quality metric (FEDM) based on the free-energy principle attached to brain theory and neuroscience, which only need one scalar from the reference image. According to free-energy principle, the human brain adaptively avoids the natural tendency to confusion in a dynamic environment and tries to concisely explain the visual stimuli, such as an image, with an internal generative model [15]. The authors in [14] utilize the linear AR model to mimic the internal generative model. For promoting the efficiency of prediction model, Liu et al. [16] adopt sparse representation to approximate the internal generative model.

In these free-energy principle based literatures, the visual stimulus is treated as an entirety. It is not consistent with the primary human vision characteristic. According to many researches on HVS by psychologists and neurologists, the

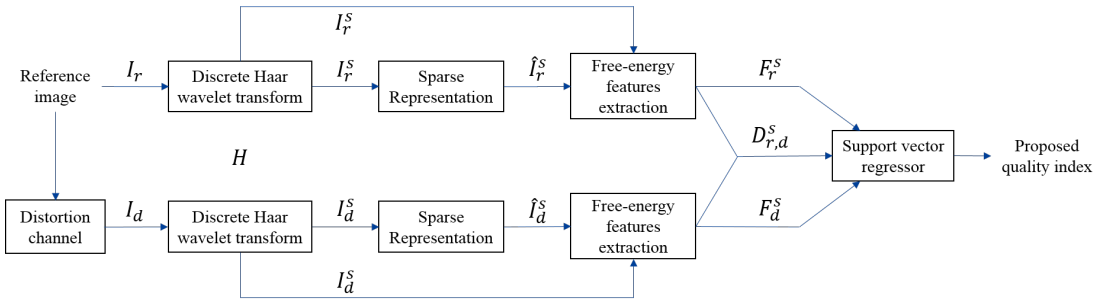

Fig. 1: The block diagram of our proposed MCFEM algorithm, where I^s represents the subband image of image I. $\hat{I}$ denotes the prediction image represented by sparse matrix. F and D are the self feature and combined feature based on free-energy principle, respectively.

vision signal processing in striate cortex attaching to primary visual cortex of human brain is frequency-independent [17]. It means that each neuron in striate cortex is particularly fit for the certain spatial and temporal frequency signal. For example, the simple cortical responds best to bars of a particular orientation, the complex cortical responds best to movement of a correctly oriented bar across the receptive field and end-stopped cortical responds to corners, angles, or bars of a particular length moving in a particular direction [18]. Therefore, the visual signal should be decomposed through an multi-channel model for subsequent processing by the corresponding neurons further. In a number of decomposition algorithms, wavelet decomposition has a superior correlation with HVS [19]. The wavelet decomposition decomposes the visual signal into multi-frequency channels in horizontal and vertical directions. The low frequency subbands of the stimulus reflect the luminance information and the high frequency subbands of the stimulus mainly contain texture and edge features. The pattern of wavelet decomposition conforms to the reaction mechanism of primary visual cortex in human brain.

Inspired by this mechanism, in this paper, we present a novel RR IQA algorithm based on free-energy principle with multi-channel decomposition, called MCFEM (Multi-Channel Free-Energy principle Metric). First, we use a two-level discrete Haar wavelet transform (DHWT) to decompose the images (reference images and distorted images) into four portions, which contain different frequency components in spatial domain. Then, each portion of images is computed the difference between the visual stimulus and its internal generative model in human brain based on free-energy principle, respectively. The sparse representation is chosen to simulate the internal generative model. Next, we calculate the self features and combined features of each pair of partitioned reference and distorted images. Finally, the prediction index of MCFEM is obtained through the support vector regressor (SVR) regressing these features. As we will demonstrate, extensive experiments preformed on four common image databases (LIVE [20], CSIQ [21], TID2008 [22], TID2013 [23]) verify the superior performance of our proposed MCFEM over the mainstream RR methods. It is noteworthy that the proposed metric only needs four scalars extracted from the pristine image, which is less than the majority of other RR algorithms.

The remainder of this paper is organized as follows. Section II introduces details of the MCFEM metric. In Section III, experimental results and comparative analysis employing four image databases are presented and discussed. Finally, general conclusions are drawn in Section IV.

II. The Proposed Algorithm

In this section, we will introduce the proposed metric that assess the visual quality of distorted images in details. The block diagram of the MCFEM method is illustrated in Fig. 1. After decomposing the reference images and distorted images, the free-energy based features are extracted from their subband images. The final index is obtain from the aggregation of these features.

A. Decomposing Images in Multi-Channel based on DHWT

The researches of HVS reveal that the different frequency and orientation components of one stimulus arouse different neurons in striate cortex. Thus, a stimulus or an image should not be treated as an entirety. It should be decomposed by a multi-channel model in human brain where the whole multi-channel subband portions are perceived by internal generative model respectively. In numerous transform algorithms, wavelet transform is one of transforms that has a great correlation with HVS. The wavelet transform can decompose image into multi-frequency channel as well as multi-orientation channel and response the luminance, texture and edge features. Wavelet has many different species, such as Haar wavelet, Daubechies wavelet and Morlet wavelet. Among these wavelets, haar wavelet is a particular case of Daubechies wavelet, which has some advantages: simple, orthogonal and compact. In this paper, we choose two-level discrete Haar wavelet transform (DHWT) to decompose the visual stimulus in multi-channel.

Mathematically, the Haar wavelet can be described as Eq. (1).

$$\begin{aligned}\frac{1}{\sqrt{2}}\psi(\frac{1}{2}) &= \sum_{n=-\infty}^{\infty}(-1)^{1-n}h[1-n]\phi(t-n)\\ &= \frac{1}{\sqrt{2}}\left(\phi(t-1)-\phi(t)\right)\end{aligned} \tag{1}$$

where, $\psi(t) = \begin{cases} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise} \end{cases}$ represents the mother wavelet function, $\varphi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases}$ denotes the scaling function and its filter $h[n]$ is defined as $h[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{if n=0,1} \\ 0 & \text{otherwise} \end{cases}$. The Haar transform cross-multiplies a function against the Haar wavelet with various shifts and stretches. Let $\boldsymbol{H}$ be the DHWT matrix of appropriate scale.

Then, let $I \in \mathcal{R}_+^{M \times N}$ represents a visual stimulus or input image. The two-level DHWT of I can be decomposed with four $\frac{M}{2} \times \frac{N}{2}$ dimensional subband blocks as shown in Eq. (2).

$$\hat{\boldsymbol{I}} = \boldsymbol{H} I \boldsymbol{H}^T = \begin{bmatrix} \hat{I}^{LL} & \hat{I}^{HL} \\ \hat{I}^{LH} & \hat{I}^{HH} \end{bmatrix} \tag{2}$$

where, $\hat{\boldsymbol{I}}$ represents the aggregation of the subband images and $\hat{I}^s, s \in \{LL, HL, LH, HH\}$ denote the LL, HL, LH and HH subbands with low or high frequency component in horizontal or vertical direction.

B. Constructing the Internal Free-Energy Model

After decomposing the image by DHWT, each subband image should be evaluated by the human brain respectively. In this paper, we attempt to use the free-energy principle model to simulate this process. An essential premise in free-energy-based brain principle is that the cognitive process is governed by an internal generative model in human brain. When human brain receive a "surprise", the brain will initiatively predict the meaningful information and remove residual uncertainty for explaining sensations with this model.

For mathematical formulation, we assume that the internal generative model $\mathcal{G}$ for visual perception is parametric, which explains visual scenes by adjusting the vector φ of parameters. Specifically, given the subband image $\hat{I}^s$ in Section II-A, the "surprise" of which can be measured by the integration of the joint distribution $P(\hat{I}^s, \boldsymbol{\varphi}|\mathcal{G})$ on the space of the model parameters vector $\boldsymbol{\varphi}$ as follows:

$$-\log P(\hat{I}^s|\mathcal{G}) = -\log \int P(\hat{I}^s, \boldsymbol{\varphi}|\mathcal{G}) d\boldsymbol{\varphi}. \tag{3}$$

In order to make this mathematical expression easy to comprehend, we introduce a dummy term $Q(\mathbf{g}|I)$ which is inserted into both the numerator and denominator of equation (3) and obtain:

$$-\log P(\hat{I}^s|\mathcal{G}) = -\log \int Q(\boldsymbol{\varphi}|\hat{I}^s) \frac{P(\hat{I}^s, \boldsymbol{\varphi}|\mathcal{G})}{Q(\boldsymbol{\varphi}|\hat{I}^s)} d\boldsymbol{\varphi}. \tag{4}$$

Here, $Q(\boldsymbol{\varphi}|\hat{I}^s, \mathcal{G})$ is an auxiliary posterior distribution of the model parameters given the image. This distribution can be thought of as an approximate posterior to the true posterior of the model parameters $P(\boldsymbol{\varphi}|\hat{I}^s, \mathcal{G})$ which can be calculated by the brain. The brain minimizes the discrepancy between the approximate posterior $Q(\boldsymbol{\varphi}|\hat{I}^s, \mathcal{G})$ and the true posterior $P(\boldsymbol{\varphi}|\hat{I}^s, \mathcal{G})$ when perceiving $\hat{I}^s$ or when adjusting the parameters $\boldsymbol{\varphi}$ of $Q(\boldsymbol{\varphi}|\hat{I}^s, \mathcal{G})$ to best explain $\hat{I}^s$. In order to simplify expression, we drop the generative model $\mathcal{G}$ in subsequent analysis. Using Jensen's inequality, we can alter equation (4) to:

$$-\log P(\hat{I}^s) \le -\int Q(\boldsymbol{\varphi}|\hat{I}^s) \log \frac{P(\hat{I}^s, \boldsymbol{\varphi})}{Q(\boldsymbol{\varphi}|\hat{I}^s)} d\boldsymbol{\varphi}. \tag{5}$$

Afterwards, based on statistical physics and thermodynamics [24], we define the right part of equation (5) as the free energy:

$$F(\boldsymbol{\varphi}) = -\int Q(\boldsymbol{\varphi}|\hat{I}^s) \log \frac{P(\hat{I}^s, \boldsymbol{\varphi})}{Q(\boldsymbol{\varphi}|\hat{I}^s)} d\boldsymbol{\varphi}. \tag{6}$$

Obviously, since $-\log P(\hat{I}^s) \le F(\boldsymbol{\varphi})$, $F(\boldsymbol{\varphi})$ defines an upper bound of 'surprise' for image $\hat{I}^s$. Noting that $P(I, \boldsymbol{\varphi}) = P(\boldsymbol{\varphi}|\hat{I}^s)P(\hat{I}^s)$, we further rearrange equation (6) as:

$$\begin{aligned} F(\boldsymbol{\varphi}) &= \int Q(\boldsymbol{\varphi}|\hat{I}^s) \log \frac{Q(\boldsymbol{\varphi}|\hat{I}^s)}{P(\boldsymbol{\varphi}|\hat{I}^s)P(\hat{I}^s)} d\boldsymbol{\varphi} \\ &= -\log P(\hat{I}^s) + \int Q(\boldsymbol{\varphi}|\hat{I}^s) \log \frac{Q(\boldsymbol{\varphi}|\hat{I}^s)}{P(\boldsymbol{\varphi}|\hat{I}^s)} d\boldsymbol{\varphi} \\ &= -\log P(\hat{I}^s) + \mathbf{KL}(Q(\boldsymbol{\varphi}|\hat{I}^s) \| P(\boldsymbol{\varphi}|\hat{I}^s)), \end{aligned} \tag{7}$$

where $\mathbf{KL}(\cdot)$ represents the Kullback-Leibler divergence between the approximate posterior and the true posterior distributions. More detailed information about free energy in [14].

In order to obtain the approximation $F(\hat{\boldsymbol{\varphi}})$ of the free energy $F(\boldsymbol{\varphi})$, the authors in [14] use the linear auto-regressive (AR) model and the authors in [25] employ sparse representation method. Both of these two methods have been prove that they resemble the pattern of visual cortex for representing images, and the latter performs better in predicting image quality. Thus, in this paper, we adopt the sparse representation method to approximate the internal generative model. Specifically, the subband images $\hat{I}^s$ can be approximately represented as:

$$\hat{I}^s \approx \boldsymbol{D}\boldsymbol{\alpha}_s \quad s.t. \quad \|\hat{I}^s - \boldsymbol{D}\boldsymbol{\alpha}_s\|_p \le \xi \tag{8}$$

where $\boldsymbol{D}$ is the dictionary that can be expressed as $[\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_K]$ and $\boldsymbol{\alpha}_s \in \mathcal{R}^K$ is the representation coefficient vector of subband image $\hat{I}^s$. Moreover, $\|\cdot\|_p$ represents the l^p norm and ξ denotes a positive threshold. The representation coefficient vector that we request for satisfies:

$$\boldsymbol{\alpha}_s^* = \arg\min_{\boldsymbol{\alpha}_s} \|\boldsymbol{\alpha}_s\|_p \quad s.t. \quad \hat{I}^s = \boldsymbol{D}\boldsymbol{\alpha}_s \tag{9}$$

This equation can be converted into an unconstrained optimization problem as shown in Eq. (10):

$$\boldsymbol{\alpha}_s^* = \arg\min_{\boldsymbol{\alpha}_s} \frac{1}{2}\|\hat{I}^s - \boldsymbol{D}\boldsymbol{\alpha}_s\|_2 + \lambda\|\boldsymbol{\alpha}_s\|_p \tag{10}$$

where λ is a positive constant in order to balance the weight of the reconstruction fidelity constraint term and the sparse punishment term. With the $\boldsymbol{D}$ and the $\boldsymbol{\alpha}_s$, the approximation subband image $\hat{I}^s$ can be obtained.

TABLE I: Performance comparison of the proposed metrics and other IQA method on LIVE, CSIQ, TID2008 and TID2013 databases with the whole four or five mainstream distortions. We highlight the top two performed models in each row. The units of features are s: scalars; IS: ImageSize.

Database	Metrics	PSNR	SSIM	NIQE	DIIVINE	$RRED$	$RRED^1$	REDLOG	FEDM	FSI	MCFEM (pro.)
Features		FR	FR	NR	NR	$\frac{IS}{576}$s	1s	6s	1s	1s	4s
LIVE	SRCC	0.8756	**0.9479**	0.9062	0.8560	0.9429	0.7653	**0.9456**	0.7947	0.8950	0.9329
	PLCC	0.8723	0.8723	0.6394	0.8446	**0.9385**	0.6880	0.9373	0.7976	0.8858	**0.9404**
	RMSE	13.360	**8.9455**	21.0067	14.6273	9.4315	19.8278	9.5161	16.4786	12.1461	**5.2107**
CSIQ	SRCC	0.9219	0.9326	0.8717	0.8284	**0.9550**	0.6728	0.9260	0.5889	0.7812	**0.9453**
	PLCC	0.9079	0.9269	0.8886	0.8556	**0.9520**	0.6000	0.9398	0.8197	0.8053	**0.9530**
	RMSE	0.1185	0.1061	0.1296	0.1463	**0.0865**	0.2114	0.0963	0.1619	0.1567	**0.0841**
TID2008	SRCC	0.8512	0.9008	0.7833	0.7630	**0.9346**	0.6728	0.9011	0.6327	0.8593	**0.9399**
	PLCC	0.8250	0.8997	0.7950	0.7563	**0.9162**	0.7381	0.9112	0.7147	0.8704	**0.9520**
	RMSE	0.8945	0.6910	0.9601	1.0356	**0.6344**	1.0680	0.6520	1.1071	0.7794	**0.4603**
TID2013	SRCC	0.9066	0.9187	0.7935	0.7813	**0.9403**	0.7217	0.9091	0.5697	0.8577	**0.9364**
	PLCC	0.8909	0.9252	0.8054	0.7855	**0.9383**	0.7665	0.9224	0.5188	0.8844	**0.9521**
	SRCC	**0.6335**	0.5294	0.8267	0.8632	**0.4823**	0.8959	0.5386	1.1925	0.6511	4.0158

C. The Proposed Image Quality Index

Since free energy represents a discrepancy measure between the image data and the best prediction by the internal generative model, it can be regarded as a natural proxy for perceptual quality of images. Based on the definition of free energy in Eq. (6) and the above-mentioned analysis, we can acquire the self free energy values $F(\hat{\varphi}_r^s)$, $F(\hat{\varphi}_d^s)$ of the reference image I_r^s and its distorted version I_d^s in $s \in \{LL, HL, LH, HH\}$ subband respectively, where

$$\hat{\varphi}_r = \arg\min_{\varphi} F(\varphi|\mathcal{G}, I_r) \tag{11}$$

$$\hat{\varphi}_d = \arg\min_{\varphi} F(\varphi|\mathcal{G}, I_d) \tag{12}$$

Besides the self free-energy feature values of reference images and their distorted versions, we also extract the relative term between them, which can reduce the influence from the image content and complexity. Therefore, we compute the absolute difference between self free-energy feature value of the reference image I_r^s and the value of its corresponding distorted image I_d^s as the combined free-energy feature as follows:

$$D(I_r^s, I_d^s) = |F(\hat{\varphi}_r^s) - F(\hat{\varphi}_d^s)| \tag{13}$$

As mentioned above, a total of twelve features contain self features and combined features based on free energy are extracted from the four reference subband images and their distorted versions. Note that only four scalars (four self free-energy feature values) need to be acquired from the reference image. Finally, the overall quality index MCFEM is obtained via the support vector regressor (SVR) based on these extracted features as illustrated below:

$$MCFEM = SVR([\boldsymbol{\mathcal{F}}_i], [q_i], I_i \in \Phi_1) \tag{14}$$

where Φ_1 denotes the training image set, $\boldsymbol{\mathcal{F}} = \{F(\hat{\varphi}_{ri}^s), F(\hat{\varphi}_{di}^s), D(I_{ri}^s, I_{di}^s)\}, s \in \{LL, HL, LH, HH\}$ means the set of features of image I_i and q_i represents the mean opinion score (MOS) value of the image I_i.

III. Experimental Results and Analysis

In this section, we introduce the experimental configurations and process and examine the prediction accuracy of our proposed method. In order to demonstrate the effectiveness of the MCFEM, we compare it with mainstream IQA metrics on four large-scale image databases.

A. Experimental Configurations and Process

In the process of decomposing image in multi-channel, we select the two-level DHWT. In the part of sparse representation, we divide the subband image into 8×8 non-overlapped patches. The predefined dictionary uses overcomplete DCT dictionary whose size is 64×128 with 128 atoms for representing each patch. The unconstrained optimization problem of sparse representation are solved by orthogonal matching pursuit (OMP) algorithm [26]. In addition, the Gaussian radial basis function (RBF) kernel is utilized in SVR and a training process is needed to adjust the regressor module. For each training stage, we randomly split the images in the database into two subsets. The training set is conisit of 80% distorted images and the rest 20% images compose the testing set. The 80% train-20% test process are repeated one thousand times to ensure the robustness of our proposed metric. Finally, we select the median result of these one thousand iterations' performances in order to eliminate the performance deviation as much as possible.

B. Experimental Results and Comparison

We test our proposed RR IQA metric on four common image databases, which are LIVE [20], CSIQ [21], TID2008

TABLE II: SRCC Values of our MCFEM and other IQA metrics in various individual distortion types on LIVE, CSIQ, TID2008 and TID2013 databases. We bold the top three performed models. The units of features are s: scalars; IS: ImageSize.

Database	Dis. Type	PSNR	SSIM	NIQE	DIIVINE	$RRED$	$RRED^1$	REDLOG	FEDM	FSI	MCFEM (pro.)
Features		FR	FR	NR	NR	$\frac{IS}{576}$s	1s	6s	1s	1s	4s
LIVE	WN	**0.9854**	0.9694	0.9716	**0.9878**	**0.9763**	0.9161	0.9302	0.9153	0.9089	0.9684
	Gblur	0.7823	**0.9517**	0.9329	**0.9584**	0.9221	0.9517	0.9349	0.7594	0.8948	**0.9651**
	JPEG	0.8809	**0.9764**	0.8661	0.7511	**0.9725**	0.8358	**0.9500**	0.8543	0.8875	0.8942
	JP2K	0.8954	**0.9614**	0.8977	0.9025	**0.9536**	0.9234	**0.9523**	0.9025	0.9145	0.9251
	FF	0.8907	**0.9556**	0.8644	0.8592	0.7549	**0.9155**	**0.9638**	0.8230	0.8479	0.8648
CSIQ	WN	**0.9363**	0.8974	0.8097	0.8662	**0.9351**	0.8010	0.8752	0.8246	0.8307	**0.9116**
	Gblur	0.9291	0.9609	0.8945	0.8945	**0.9634**	**0.9649**	0.9346	0.8522	0.8909	**0.9613**
	JPEG	0.8879	**0.9543**	0.8832	0.7998	**0.9523**	0.8220	0.9282	0.9166	0.8886	**0.9315**
	JP2K	0.9361	**0.9605**	0.9062	0.8304	0.9630	0.9387	**0.9479**	0.8945	0.9028	**0.9392**
TID2008	WN	**0.9115**	0.8310	0.7797	0.8085	0.8203	0.6818	**0.8407**	0.6855	0.6552	**0.9063**
	Gblur	0.8682	**0.9596**	0.8165	0.8237	**0.9573**	**0.9565**	0.8775	0.7980	0.9163	0.9356
	JPEG	0.9011	**0.9270**	0.8608	0.6309	**0.9332**	0.6025	0.9117	0.7594	0.8112	**0.9088**
	JP2K	0.8300	**0.9723**	0.8964	0.8964	**0.9681**	0.9396	**0.9562**	0.8162	0.8991	0.9314
TID2013	WN	**0.9225**	0.8656	0.8144	0.8510	0.8461	0.7496	**0.8754**	0.7485	0.6774	**0.9165**
	Gblur	0.9149	**0.9668**	0.7954	0.8344	**0.9666**	**0.9672**	0.8955	0.8891	0.9454	0.9582
	JPEG	0.9189	**0.9200**	0.8423	0.6288	**0.9274**	0.6974	0.8972	0.7482	0.8431	**0.8984**
	JP2K	0.8840	**0.9468**	0.8891	0.8534	**0.9539**	0.8970	**0.9333**	0.8409	0.8904	0.9226

[22], TID2013 [23]. Specifically, we choose the mainstream encountered distortion types, JPEG compression (JPEG), JPEG2000 compression (JP2K), Gaussian white noise (WN), Gaussian blur (Blur) and Fast-fading Rayleigh channel (FF). Before evaluating performance, we employ a five parameter $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ monotonic logistic function suggested by VQEG [27] to map the prediction scores by objective models:

$$Y(x) = \beta_1(0.5 - \frac{1}{1 + e^{\beta_2(x-\beta_3)}}) + \beta_4 x + \beta_5 \tag{15}$$

where x and Y denote the objective scores and mapped scores.

Then, we apply three commonly used evaluation criteria to evaluate the performances of all compared IQA metrics, which are Spearman Rank-Order Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC) and Root Mean Squared Error (RMSE) respectively. The definitions of these indexes are shown as follows:

$$SRCC = 1 - \frac{6\sum_{n=1}^{N} d_i^2}{N(N^2-1)} \tag{16}$$

$$PLCC = \frac{\sum_{i=1}^{N}(p_i - \bar{p})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{N}(p_i - \bar{p})^2 (s_i - \bar{s})^2}} \tag{17}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - s_i)^2} \tag{18}$$

where d_i denotes the difference between the ranks of i-th images in subjective and objective assessments, and N represents the number of testing images. In addition, s_i and p_i are the subjective rating and converted objective score of i-th image after nonlinear regression. $\bar{s}$ and $\bar{p}$ indicate the mean of all s_i and p_i. To specify, SRCC reflects the prediction monotonicity, PLCC indicates the prediction accuracy of the IQA algorithm and RMSE represents the prediction consistency. Moreover, An excellent IQA metric is expected to obtain the value of SRCC and PLCC close to 1, yet the value near 0 for RMSE.

For demonstrating the effectiveness of our proposed method, we compare MCFEM with some representative IQA approaches: 1) classical FR algorithms including PSNR and SSIM [1]; 2) popular NR metrics: DIIVINE [4] and NIQE [5]. 3) competitive RR algorithms containing FEDM [14], FSI [8], RRED [12], REDLOG [13]. For RR algorithms, our MCFEM metric needs the original image to provide four scalars as reference data. On account of the different models of RRED method according to the number of scalars from the reference image, we select the mode of common state (RRED in TABLE I and II), which needs $\frac{ImageSize}{576}$ scalars, and the mode of requiring a single scalar from reference image ($RRED_1$). Besides RRED model, both of FSI and FEDM only transmit one scalar as reference data. Six scalars take charge of reference data in REDLOG algorithm.

The performance indices of all of these IQA models are tabulated in TABLE I, where the top two performed metrics are highlighted. The overall performance is computed with all of distortion types in LIVE and WN, Gblur, JPEG and JP2K in CSIQ, TID2008 and TID2013. It is obvious that our proposed method highly correlates with subjective ratings. Specifically, our proposed MCFEM metrics outperforms the majority of

compared methods and similar with the RRED algorithm in overall performance. MCFEM is superior to RRED in TID2008 while inferior to RRED in LIVE and our metric has similar results with RRED in CSIQ and TID2013. Noted that the number of features in MCFEM (4) is far less than the RRED's requirement ($\frac{ImageSize}{576}$). In addition, though the information of RR methods from reference image are hardly matchable with FR models, our MCFEM algorithm is better than the classical FR methods: PSNR and SSIM.

For more concretely, we tabulate the SRCC values of MCFEM and other IQA algorithms in various individual distortion categories in TABLE II and bold the top three performed metrics. From TABLE II, we can find that SSIM has the best performance, while MCFEM and RRED also have good results. Our method has the best performance in CSIQ among the RR algorithms, while RRED and REDLOG perform better in LIVE. Furthermore, our proposed metric has more stable performance compared with other RR models. For example, RRED model has poor performance with distortion type of FF in LIVE and the result of REDLOG with WN in TID2008 is not good. Our MCFEM metric has no distinct weakness in these common distortion types in four databases.

IV. Conclusion

In this paper, a new perceptual RR IQA model named MCFEM is proposed based on free-energy principle in multi-channel. The proposed model is inspired by the mechanism of information decomposition in HVS and representation in human brain. In MCFEM, the original and distorted images are firstly decomposed into four subband in multi-frequency and multi-orientation based on DHWT. Then the features on the strength of free-energy model using sparse representation are extracted. Finally, the overall quality index is obtained by SVR method. Validation results on four widely used image databases demonstrate that MCFEM has comparative performance with classical FR methods, such as PSNR and SSIM. It is important that our proposed method outperforms the most of state-of-the-art RR metrics. In addition, our metric only needs four scalars from reference image, which is less than the majority of other RR algorithms.

Acknowledgment

This work was supported by the National Science Foundation of China (61521062, 61527804) and Science and Technology Commission of Shanghai Municipality (15DZ0500200).

References

[1] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612.

[2] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444.

[3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386.

[4] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[5] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[6] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1098–1105.

[7] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 50–63, 2015.

[8] M. Liu, K. Gu, G. Zhai, P. Le Callet, and W. Zhang, "Perceptual reduced-reference visual quality assessment for contrast alteration," *IEEE Transactions on Broadcasting*, vol. 63, no. 1, pp. 71–81, 2017.

[9] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, 2013.

[10] J. A. Redi, P. Gastaldo, I. Heynderickx, and R. Zunino, "Color distribution information for the reduced-reference assessment of perceived image quality," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 12, pp. 1757–1769, 2010.

[11] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE transactions on image processing*, vol. 15, no. 6, pp. 1680–1689, 2006.

[12] R. Soundararajan and A. C. Bovik, "Rred indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.

[13] S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DWT coefficients of locally weighted gradient magnitudes," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5293–5303, 2016.

[14] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 41–52, 2012.

[15] G. P. Wagner, "The origins of order: Self-organization and selection in evolution - kauffman,sa," 1993.

[16] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–1, 2017.

[17] E. B. Goldstein and J. Brockmole, *Sensation and perception*. Cengage Learning, 2016.

[18] F. Heitger, L. Rosenthaler, R. Von Der Heydt, E. Peterhans, and O. Kübler, "Simulation of neural contour mechanisms: from simple to end-stopped cells," *Vision research*, vol. 32, no. 5, pp. 963–981, 1992.

[19] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 2091–2110, 1989.

[20] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," 2005. [Online]. Available: http://live.ece.utexas.edu/research/quality/

[21] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 006–011 006, 2010.

[22] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008–A database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.

[23] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[24] R. Feynman, "Statistical mechanics: a set of lectures (advanced book classics)," 1998.

[25] Y. Liu, G. Zhai, X. Liu, and D. Zhao, "Perceptual image quality assessment combining free-energy principle and sparse representation," in *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 1586–1589.

[26] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[27] VQEG, "Final report from the vqeg on the validation of objective models of video quality assessment, Pase II," 2003. [Online]. Available: http://www. vpeg. org