This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

Using Free Energy Principle For Blind Image Quality Assessment

Ke Gu, Guangtao Zhai, Xiaokang Yang, *Senior Member, IEEE,* and Wenjun Zhang, *Fellow, IEEE*

***Abstract*—In this paper we propose a new no-reference (NR) image quality assessment (IQA) metric using the recently revealed free energy based brain theory and classical human visual system (HVS) inspired features. The features used can be divided into three groups. The first involves the features inspired by the free energy principle and the structural degradation model. Furthermore, the free energy theory also reveals that the HVS always tries to infer the meaningful part from the visual stimuli. In terms of this finding, we first predict an image that the HVS perceives from a distorted image based on the free energy theory, then the second group of features is composed of some HVS inspired features (such as structural information and gradient magnitude) computed using the distorted and predicted images. The third group of features quantifies the possible losses of "naturalness" in the distorted image by fitting the generalized Gaussian distribution to mean subtracted contrast normalized coefficients. After feature extraction, our algorithm utilizes the support vector machine based regression module to derive the overall quality score. Experiments on LIVE, TID2008, CSIQ, IVC and Toyama databases confirm the effectiveness of our introduced NR IQA metric compared to the state-of-the-art.**

***Index Terms*—Image quality assessment (IQA), no-reference (NR), free energy, structural degradation, human visual system**

I. Introduction

IN THE year 2011, over eighty billion digital photographs were captured throughout the continental United States of America, and this number will be increasing annually. A natural problem is that the visual quality of such a great amount of photographs is hard to guarantee. So the systems to monitor, control and improve the visual quality of digital photographs are highly desirable [1]. Image quality assessment (IQA), due to its capability of simulating human visual perception to image quality, is usually used to solve this problem.

Generally speaking, IQA approaches can be classified into subjective assessment and objective assessment. The role of subjective evaluation is decisive since in most applications it is human viewers who judge the overall visual quality. One important function of subjective assessment is to instruct video coding [2]-[3]. Nonetheless, the subjective IQA is extremely slow, expensive and laborious, and thus is not suitably applied under the condition that hundreds of thousands of images are acquired, compressed and transmitted every moment. Hence, an increasing number of researchers have concentrated on the exploration of objective IQA algorithms.

The largest number of objective metrics are full-reference (FR) methods [4]-[14], which assume that the original image signal is completely known. The mean-squared error (MSE) and its relevant peak signal-to-noise ratio (PSNR) were very popular owing to their low computational cost, high portability and clear physical meaning, but they were found to poorly correlate with human judgments of image quality, i.e. the mean opinion score (MOS) [15]. To this end, the structural similarity index (SSIM) [4] and its variants [5]-[10] have been developed in the pursuit of higher performance.

However, the application scope of FR IQA is largely limited since the original image is unavailable in most cases. Supposing that partial original references can be made available as side information, reduced-reference (RR) IQA techniques have lately attracted great concerns and acquired fairly well performance for different types of distortions [16]-[20]. Yet RR IQA still requires original information in practice, leading to its incompatibility with most existing image/video processing systems that do not permit extra RR information.

To solve the problem of the dependence of original images, many blind quality measures have been developed for specific distortion types during the last decade [21]-[28]. Wang *et al.* [21] introduced a No-reference JPEG-quality Evaluator (WNJE) based on the estimation of blocking effects and relative blur. Marziliano *et al.* proposed a Blind Blur Metric (MBBM) [22] to measure the spread of image edges from horizontal and vertical directions. Very recently, the topic of noise estimation has obtained intensive researches. One type of methods is scale invariant based noise estimator (SINE) [24] and its variant [25], which suppose that the kurtosis values tend to be invariant across scales for a natural image and this scale invariance will be deteriorated by the added noise. In addition, some recent advances in brain science and neuroscience [29]-[30] motivated us to design the no-reference free energy based quality metric (NFEQM) [17] for predicting the quality of both blurry and noisy images.

Note that those above blind measures are distortion-specific. Therefore, the general-purpose blind/no-reference (NR) IQA methods have been emphatically studied in recent years [31]-[38]. The general-purpose NR IQA can be mainly categorized into two classes. The first is to extract effective features from distorted images followed by training a regression module using those features. Inspired by the natural scene statistics (NSS) model, DIIVINE [31], BLIINDS-II [32] and BRISQUE [33] were respectively proposed to work in DWT, DCT and spatial domains. Besides, our NFSDM was designed with an

This work was supported in part by the National Science Foundation of China under Grant 61025005, Grant 61371146, Grant 61221001, and Grant 61390514, the Foundation for the Author of National Excellent Doctoral Dissertation of PR China under Grant 201339, and the Major State Basic Research Development Program of China 973 Program under Grant 2010CB731401.

The authors are with Institute of Image Communication and Information Processing, Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiao Tong University, Shanghai, 200240, China. (email: guke.doctor@gmail.com; zhaiguangtao/xkyang/zhangwenjun@sjtu.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

alternative way of extracting features [34]. The second class of general-purpose NR IQA metrics operates without human scored images. For instance, natural image quality evaluator (NIQE) [35] was developed to estimate the deviations from statistical regularities observed in natural images without any prior knowledge of image contents or distortion types. And quality-aware clustering (QAC) [36] works by learning a set of quality-aware centroids to act as a codebook to compute the quality levels of image patches and infer the quality score of the overall image.

In this article we modify NFSDM to design *NR Free Energy based Robust Metric (NFERM)* by adding HVS inspired features to improve prediction performance, and reducing the total number of features by half. We can divide the used features into three groups. The first one includes 13 features of the free energy and the structural degradation information. The free energy feature comes from the RR free energy based distortion metric (FEDM) [17], which defines the psychovisual quality as the agreement between an input image and the output of internal generative model, while the structural degradation information is computed by the RR structural degradation model (SDM) [19] that amends SSIM with itself. Although the two RR IQA metrics still require partial reference information, the free energy feature and the structural degradation information of original images are found to be of an approximate linear relationship. According to this observation, the dependence of original references can be largely removed. More details can be found in Section II-A.

Furthermore, the free energy theory reveals that the human visual system (HVS) always attempts to reduce the uncertainty based on the internal generative model when perceiving and understanding an input visual stimulus. For example, human brains can automatically restore or denoise a noisy image. We apply the linear autoregressive (AR) model to approximate the generative model to predict an image that the HVS perceives from an input distorted one. Then, six important HVS inspired features (e.g. structural information and gradient magnitude), which are computed from the distorted and predicted images, constitute the second group of features. The third group of four features arises from the NSS model. We estimate the possible losses of 'naturalness' in the distorted image by fitting the generalized Gaussian distribution to mean subtracted contrast normalized coefficients. An important note is that, with free energy principle and image scene statistics, this paper links FR, RR and NR IQA together, and proposes a general model for higher performance via a proper integration of existing FR, RR and NR IQA methods.

The remainder of this paper is organized as follows. Section II first introduces the mainstream scheme of general-purpose NR IQA algorithms, and then describes the proposed NFERM in detail. In Section III, comparative studies of our NFERM with classical FR IQA approaches and state-of-the-art NR IQA metrics are conducted on five popular image databases (LIVE [39], TID2008 [40], CSIQ [41], IVC [42], and Toyama [43]), confirming the effectiveness of the proposed NFERM method. Finally, some concluding remarks are given in Section IV.

II. Proposed NR IQA Metric

Existing distortion-specific blind measures, despite of well performance, are greatly limited by the dependence of the prior knowledge of the distortion category and the dedicated application scenario. The general-purpose NR IQA, which can simultaneously tackle various distortion types, therefore has drawn more attention in recent days.

Broadly speaking, mainstream general-purpose NR image quality metrics operate in three steps.

- First, the features are extracted, e.g. using the classical NSS model.
- Second, the overall distorted images are randomly separated into training and testing groups, and the *model* is then acquired using support vector regressor (SVR) [44] on the extracted features of images in the training group and the corresponding subjective MOS values. Given those features of one image $\mathbf{f} = \{f_1, f_2, ..., f_n\}$ and the training set Φ_1, the *model* is defined as

$$model = \text{SVR\_TRAIN}([\mathbf{f}_i], [q_i], I_i \in \Phi_1) \tag{1}$$

  where q_i is the MOS value of the image I_i.
- Finally, the correlation performance of the NR IQA metric is justified on the testing group with the obtained *model*. The objective quality score s_j of the image I_j is calculated by

$$s_j = \text{SVR\_PREDICT}([\mathbf{f}_j], model, I_j \in \Phi_2) \tag{2}$$

  where Φ_2 is the testing set. Next, the correlation measured between $[q_j]$ and $[s_j]$ in the testing set indicates the performance of the NR IQA method.

We plot the flowchart of the mainstream scheme of NR IQA metrics in Fig. 1. Note that feature extraction is the key point.

A. Feature Group One

The first group, composed of 13 features (f_{01}-f_{13}), comes from two effective RR IQA algorithms (FEDM and SDM) and their underlying connection. Our previous FEDM [17] was realized according to the free energy theory, which was recently revealed by Friston to explain and unify several brain theories in biological and physical sciences about human action, perception and learning [29]-[30]. Similar to the Bayesian brain hypothesis [45], a basic premise of the free energy based brain theory is that the cognitive process is manipulated by an internal generative model. Using this generative model, the human brain can actively infer predictions of the meaningful information of input visual signals and avoid the residual uncertainty in a constructive manner.

This constructive model is a probabilistic model in essential, which can be decomposed into a likelihood term and a prior term. The visual perception is then to infer the posterior possibilities of the given scene by inverting this likelihood term. It is natural that there exists a gap between the real scene and the brain's prediction, in that the generative model cannot be universal. This gap between the external input and its generative-model-explainable part is believed to be very closely related to the quality of human visual perceptions, and even can be use for the quality measure [17].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

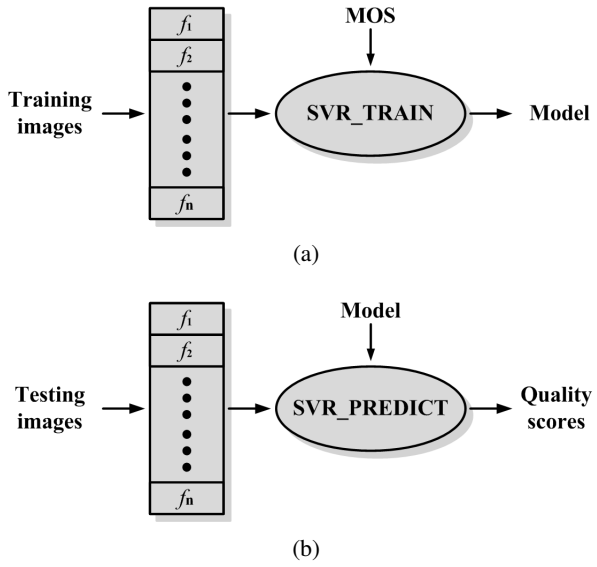


Fig. 1: The flowchart of the mainstream scheme of NR IQA metrics: (a) Using SVR to train some images and associated MOS values to acquire the *model*; (b) Using SVR to predict objective quality scores of the rest images based on the *model* for testifying performance.

For operational amenability, it is assumed that the internal generative model $\mathcal{G}$ for visual perception is parametric, which explains perceived scenes by adjusting the parameter vector $\mathbf{g}$. Given an image I, its 'surprise' (determined by entropy) can be evaluated by integrating the joint distribution $P(I,\mathbf{g})$ over the space of model parameters $\mathbf{g}$

$$-\log P(I) = -\log \int P(I,\mathbf{g}) d\mathbf{g}. \tag{3}$$

We bring a dummy term $Q(\mathbf{g}|I)$ into both the denominator and numerator in Eq. (3) and rewrite it as

$$-\log P(I) = -\log \int Q(\mathbf{g}|I) \frac{P(I,\mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g}. \tag{4}$$

Here $Q(\mathbf{g}|I)$ is an auxiliary posterior distribution of the model parameters given the input image signal I. It can be thought of as an approximate posterior to the true posterior of the model parameters $P(\mathbf{g}|I)$ evaluated by the brain. When perceiving the image I or when adjusting the parameters $\mathbf{g}$ of $Q(\mathbf{g}|I)$ to search for the optimal explanation of I, the brain will minimize the discrepancy between the approximate posterior $Q(\mathbf{g}|I)$ and the true posterior $P(\mathbf{g}|I)$.

Then, we use Jensen's inequality and obtain from Eq. (4):

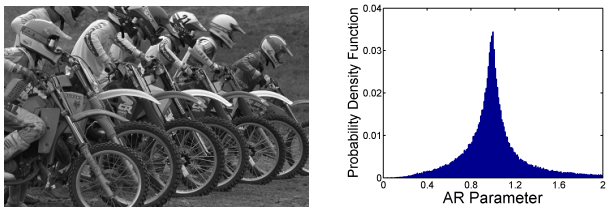
$$-\log P(I) \leq -\int Q(\mathbf{g}|I) \log \frac{P(I,\mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g} \tag{5}$$

and define the right hand side as the free energy:

$$\mathcal{J}(\mathbf{g}) = -\int Q(\mathbf{g}|I) \log \frac{P(I,\mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g}. \tag{6}$$

As a consequence, the free energy estimation of the image I can be expressed by

$$F(I) = \mathcal{J}(\hat{\mathbf{g}}) \quad \text{with} \quad \hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \mathcal{J}(\mathbf{g}). \tag{7}$$



(a) (b)

Fig. 2: Illustration of the posterior distribution of the model parameters $Q(\mathbf{g}|I)$ by: (a) a natural image; (b) the associated distribution of $Q(\mathbf{g}|I)$ computed using the AR model.

A model with higher expressive power approximates the brain better but incurs higher computational complexity. Moreover, a more complex model with a large number of parameters has a higher model cost in the theory of model selection [46], and thus more difficult to estimate from observations. In this paper we choose the generative model to be the linear AR model for its simplicity and ability to well characterize a wide range of natural scenes. The AR model is defined as

$$x_n = \boldsymbol{\chi}^k(x_n)\boldsymbol{\lambda} + \varepsilon_n \tag{8}$$

where x_n is a pixel in question, $\boldsymbol{\chi}^k(x_n)$ is a row-vector that consists of k nearest neighbors of x_n, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_k)^T$ is a vector of AR coefficients, and ε_n is the error term. More details about how to get the $\boldsymbol{\lambda}$ can be found in the Appendix. Next, we can use the input distorted image I_d in a point-wise manner to estimate the predicted version I_p via $\boldsymbol{\chi}^k(x_n) \cdot \boldsymbol{\lambda}_{opt}$, where $\boldsymbol{\lambda}_{opt}$ is the optimal estimate of AR parameters for x_n based on the least square method. In this case, the distribution of the model parameters $Q(\mathbf{g}|I)$ are represented by that of the estimated AR parameters, which exhibits a center-peaked appearance when the sampled data are large enough. In order to illustrate this, a natural image and its auxiliary posterior distribution of the model parameters $Q(\mathbf{g}|I)$ computed using the first-order AR model are shown in Fig. 2.

In reality, the process of free-energy minimization is highly related to the predictive coding, as pointed out in the efficient coding theory [47] and the Infomax theory [48]. As a matter of fact, supposing the internal generative model to be an AR model, the process of free-energy minimization is equivalent to encoding the input visual signal I with the minimum number

TABLE I: Important notations and abbreviations in this paper.

$\mathcal{G}$	The internal generative model
$\mathbf{g}$	The parameter vector
I	The input image
$P(I)$	The distribution of I
$P(I,\mathbf{g})$	The joint distribution of I and $\mathbf{g}$
$P(\mathbf{g}\|I)$	The true posterior distribution of $\mathbf{g}$ given I
$Q(\mathbf{g}\|I)$	The auxiliary posterior distribution of $\mathbf{g}$ given I
$\mathcal{J}(\mathbf{g})$	The free energy
$F(I)$	The free energy estimation of I
x_n	The pixel in question
$\boldsymbol{\chi}^k(x_n)$	A vector including k nearest neighbors of x_n
$\boldsymbol{\lambda}$	A vector of AR coefficients
AR	Autoregressive model
PC	Phase congruency
GM	Gradient magnitude

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

of bits based on the AR model [49]. To achieve the minimum coding length, the piecewise AR model is the best choice, e.g. model selection-based image compression [50]. Precisely, the total description length of I with the kth-order AR model can be expressed by

$$L(\hat{\mathbf{g}}) = -\log P(I|\hat{\mathbf{g}}) + \frac{k}{2}\log N \tag{9}$$

where N is the number of pixels. The model is selected by minimizing $L(\hat{\mathbf{g}})$. As shown in [49], in the large sample limit $N \to \infty$, the free energy is the total description length:

$$\mathcal{J}(\hat{\mathbf{g}}) = -\log P(I|\hat{\mathbf{g}}) + \frac{k}{2}\log N \quad \text{with} \quad N \to \infty. \tag{10}$$

Hence, the free energy of image can be approximated as the total description length of the image data using the AR model, i.e. the entropy (average information amount) of the prediction residuals of I_d and I_p plus the model cost. In this stage, we choose a fixed-model order, and thus the second term $\frac{k}{2}\log N$ is constant and can be ignored in the quality evaluation. We list important notations and abbreviations in Table I.

It was found that, for most images with various distortion types and quality levels, their low-pass filtered versions have different degrees of spatial frequency decrease, which inspires the design of the other RR SDM model [19]. This phenomenon reveals one limitation of SSIM about its inability to distinguish distortion types and quality levels well. The SDM can solve the problem by measuring the similarity between the structural degradation information of original and distorted images, and thus induces performance gain to some extent.

Specifically, following the definition of local statistics in SSIM [4], we first define μ_I and σ_I as the local mean and variance of I with a 2D circularly-symmetric Gaussian weighting function $\mathbf{w} = \{w(k,l)|k = -K, ..., K, l = -L, ..., L\}$, which satisfies sum($\mathbf{w}$) $= 1$ and var($\mathbf{w}$) $= 1.5$ (sum$(\cdot)$ and var$(\cdot)$ compute sum and variance values). $\bar{\mu}_I$ and $\bar{\sigma}_I$ have the same definitions except using the impulse function instead of the Gaussian weighting function. Then, the structural degradation information is given by

$$S_a(I) = E\left(\frac{\sigma_{(\mu_I\bar{\mu}_I)} + C_1}{\sigma_{(\mu_I)}\sigma_{(\bar{\mu}_I)} + C_1}\right) \tag{11}$$

$$S_b(I) = E\left(\frac{\sigma_{(\sigma_I\bar{\sigma}_I)} + C_1}{\sigma_{(\sigma_I)}\sigma_{(\bar{\sigma}_I)} + C_1}\right) \tag{12}$$

where $E(\cdot)$ is a direct average pooling. $\sigma_{(\mu_I\bar{\mu}_I)}$ and $\sigma_{(\sigma_I\bar{\sigma}_I)}$ represent the local covariance similar to the definition in [4]. C_1 is a small constant to avoid the denominator to be zero.

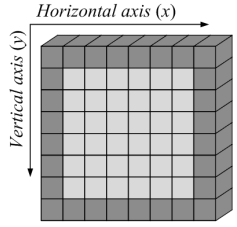


Fig. 3: Illustration of interior parts or exterior parts of blocks. For a block of size 8×8, the dark gray part outside is the exterior part, while the middle part colored with light gray is the interior part.

Fig. 4: The chosen thirty images of a broad range of scenes from the Berkeley image database [51].

This paper picks three pairs of (K, L) as $(1, 1)$, $(3, 3)$ and $(5, 5)$, because Gaussian weighting functions of various sizes introduce different amounts of neighboring pixels' information on one point. Also, note that the relationships between SDM's predictions and subjective scores are quite distinct for images corrupted with white noise and other distortion types. Namely, noise images of poorer quality have larger SDM values whereas images of other distortion types with poorer quality have smaller SDM values. We thereby modify $S_a(I)$ to keep different types of distortions consistent:

$$\tilde{S}_a(I) = \begin{cases} -S_a(I) & \text{if } F(I) > T \\ S_a(I) & \text{otherwise} \end{cases} \tag{13}$$

TABLE II: Definitions of $\tilde{S}_a$ and $\tilde{S}_b$ for interior and exterior parts as well as different (K, L) values.

	Interior parts		Exterior parts	
	$\hat{S}_a$	$\hat{S}_b$	$\check{S}_a$	$\check{S}_b$
$(K, L) = (1, 1)$	$\hat{S}_{a1}$	$\hat{S}_{b1}$	$\check{S}_{a1}$	$\check{S}_{b1}$
$(K, L) = (3, 3)$	$\hat{S}_{a3}$	$\hat{S}_{b3}$	$\check{S}_{b3}$	$\check{S}_{b3}$
$(K, L) = (5, 5)$	$\hat{S}_{a5}$	$\hat{S}_{b5}$	$\check{S}_{b5}$	$\check{S}_{b5}$

TABLE III: The estimates of parameters α_s, β_s, θ_s and ϕ_s for $\hat{S}_s$ and $\check{S}_s$ ($s = \{a1, a3, a5, b1, b3, b5\}$) using the least square method.

	α_s	β_s		θ_s	ϕ_s
$\hat{S}_{a1}$	-13.279	15.194	$\hat{S}_{b1}$	-13.326	15.236
$\hat{S}_{a3}$	-7.9861	8.2961	$\hat{S}_{b3}$	-8.0013	8.3093
$\hat{S}_{a5}$	-13.019	14.988	$\hat{S}_{b5}$	-13.096	15.051
$\check{S}_{a1}$	-7.8427	8.3219	$\check{S}_{b1}$	-7.8451	8.3282
$\check{S}_{a3}$	-12.399	14.808	$\check{S}_{b3}$	-12.378	14.795
$\check{S}_{a5}$	-6.7687	8.1662	$\check{S}_{b5}$	-6.8255	8.1973

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

where T is set as 5 according to the observation. And $S_b(I)$ is modified as $\tilde{S}_b(I)$ similarly.

Since $\tilde{S}_a(I)$ and $\tilde{S}_b(I)$ are not effective quality measures for JPEG compression (i.e. their values for JPEG compressed images near to zero) [19], we use the segmentation of interior and exterior parts in each block. As presented in Fig. 3, for a block of size 8×8, the dark gray part outside corresponds to the exterior part, while the middle light gray part indicates the interior part. Besides, some IQA approaches incorporating the *downsample* strategy have attained better correlation with human perception [9], [11], [14], [18]-[19]. This motivates us to compute $\tilde{S}_a(I)$ and $\tilde{S}_b(I)$ at a reduced resolution (low-pass filtered and downsampled by a factor of 2). We redefine the structural degradation information in Table II.

In [34], we have shown that there exists an approximate linear relationship between the structural degradation information and the free energy feature of original images in the LIVE database. We randomly selected thirty images of different scenes (refer to Fig. 4) from the Berkeley database [51], in order to better validate the generality and database-independency of the NFERM. We use the Berkeley database because existing IQA databases [39]-[43] will be used to testify various NR IQA metrics in later experiments. We then compare the structural degradation information $\hat{S}_s(I_r)$ and $\check{S}_s(I_r)$ ($s = \{a1, a3, a5, b1, b3, b5\}$) with the free energy feature $F(I_r)$ of those thirty images and draw their scatter plots in Fig. 5. The linear dependence between the free energy feature and the structural degradation information provides an opportunity to characterize distorted images without original image information. We fit the linear regression model:

$$F(I_r) = \alpha_s \cdot \hat{S}_s(I_r) + \beta_s \tag{14}$$

$$F(I_r) = \theta_s \cdot \check{S}_s(I_r) + \phi_s \tag{15}$$

where α_s, β_s, θ_s and ϕ_s are obtained based on the least square method, and their values are reported in Table III.

Finally, we utilize $\hat{SS}_s = F(I_d) - (\alpha_s \cdot \hat{S}_s(I_d) + \beta_s)$ and $\check{SS}_s = F(I_d) - (\theta_s \cdot \check{S}_s(I_d) + \phi_s)$ to reduce the dependence of original references, due to the fact that both $\hat{SS}_s$ and $\check{SS}_s$ values of high-quality images (with very few distortions) are quite close to zero, whereas they will be far from zero when distortions become larger. Consequently, we define the first twelve features as follows:

$$\begin{cases} f_{01}\text{-}f_{06} : \hat{SS}_s & s = \{a1, a3, a5, b1, b3, b5\} \\ f_{07}\text{-}f_{12} : \check{SS}_s & s = \{a1, a3, a5, b1, b3, b5\} \end{cases}.$$

Additionally, the NFEQM correlates well with human ratings on noisy and blurred images (as listed in Table IV), so we use it (namely $F(I_d)$) as the feature f_{13} in the first group.

B. Feature Group Two

The second group of 6 features (f_{14}-f_{19}) is also inspired by the free energy theory, which illustrates that the HVS always attempts to perceive and understand an input visual stimulus by reducing the uncertainty based on the internal generative model. When watching a noisy image such as Fig. 6(a), one will be quick to restore or denoise it automatically and catch what that image means. This depends on the above-mentioned

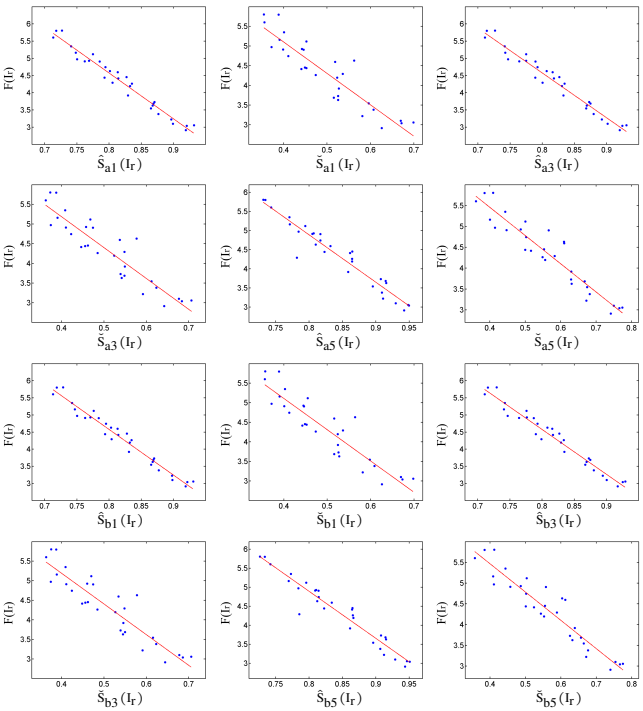


Fig. 5: Scatter plots of the structural degradation information $\hat{S}_s(I_r)$ and $\check{S}_s(I_r)$ ($s = \{a1, a3, a5, b1, b3, b5\}$) vs. the free energy feature $F(I_r)$ on thirty images in the Berkeley database [51]. The straight lines are fitted with the least square method.

internal generative model. Thus, we use the above linear AR model to approximate the generative model, thereby to predict an image that the HVS perceives from the input distorted one, as exhibited in Fig. 6(b).

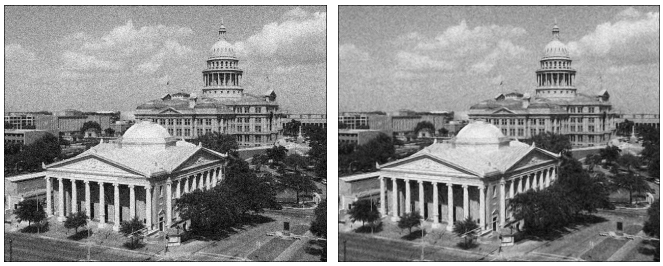
Recalling the approximation of the free energy in Section II-A that the difference between the distorted image and its predicted version is quantified by entropy, we measure that difference in other fashions. As stated in [13], PSNR is good at estimating content-independent distortions such as white noise, while SSIM is suitable for content-dependent distortions such as Gaussian blur, JPEG2000 (JP2K) and JPEG compressions. Performance comparisons in Table IV confirm this conclusion, namely PNSR and SSIM work effectively for white noise and JP2K, JPEG, blur, fastfading. So we in this paper compute PSNR between the distorted image I_d and its predicted version I_p as the feature f_{14}:

$$f_{14} = 10\log_{10}\left(\frac{255^2}{\frac{1}{M}\sum_{i=1}^{M}[I_d(i) - I_p(i)]^2}\right) \tag{16}$$

where M is the number of pixels in the whole image. Next, considering that contrast and structural similarities in SSIM are more valid than the luminance similarity (for example, MS-SSIM mainly focuses on contrast and structural similarities), and furthermore, the luminance similarity is closely related to PSNR, we choose contrast and structural similarities between I_d and I_p to be features f_{15}-f_{16}:

$$f_{15} = E\left(\frac{2\sigma_{(I_d)}\sigma_{(I_p)} + 2C_1}{\sigma^2_{(I_d)} + \sigma^2_{(I_p)} + 2C_1}\right) \tag{17}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia



(a) (b)

Fig. 6: Illustration of the internal generative model of human brains: (a) a noisy image; (b) a predicted image with the linear AR model.

$$\frac{1}{16}\times\begin{bmatrix}3 & 0 & -3\\ 10 & 0 & -10\\ 3 & 0 & -3\end{bmatrix} \qquad \frac{1}{16}\times\begin{bmatrix}3 & 10 & 3\\ 0 & 0 & 0\\ -3 & -10 & -3\end{bmatrix}$$

(a) (b)

Fig. 7: Scharr gradient operator [54].

$$f_{16} = E(\frac{\sigma_{(I_d I_p)} + C_1}{\sigma_{(I_d)}\sigma_{(I_p)} + C_1}). \tag{18}$$

where $E(\cdot)$ is to compute the mean or expectation value.

The HVS is strongly sensitive to phase congruency (PC) and gradient magnitude (GM), which have been shown to be very effective in recent IQA techniques [11]-[13]. Instead of defining features simply at pixels with sharp changes in intensity, the PC model postulates that the HVS perceives features at points, where the Fourier components are maximal in phase. According to the physiological and psychophysical evidences, the PC model provides a simple yet biologically plausible model of how the HVS detects and identifies features in an image [52]-[53]. Hence we set the feature f_{17} as

$$f_{17} = E(\text{PC}_\text{m}) = E\{\max[PC(I_d), PC(I_p)]\} \tag{19}$$

where PC is defined in the widely employed form [53]. On the other hand, image gradient computation, a very classical topic in image processing, is also valid in IQA performance gains [11]-[12]. Gradient operators are inherently expressed by convolution masks. In this study, we utilize the Scharr operator [54], as illustrated in Fig 7. The GM is defined as $GM = \sqrt{GM_x^2 + GM_y^2}$, where GM_x and GM_y are the partial derivatives of the image along horizontal and vertical directions using the Scharr operator. This GM is taken as the feature f_{18}:

$$f_{18} = E(\text{GM}_\text{map}) = E(\frac{2GM(I_d)\cdot GM(I_p) + C_2}{GM(I_d)^2 + GM(I_p)^2 + C_2}). \tag{20}$$

In most cases, salient areas (e.g. PC_m) have a high impact on HVS when evaluating image quality. We thus combine PC and GM components weighted by PC_m to derive the feature f_{19}:

$$f_{19} = \frac{E(\text{GM}_\text{map}\cdot\text{PC}_\text{map}\cdot\text{PC}_\text{m})}{E(\text{PC}_\text{m})} \tag{21}$$

where

$$\text{PC}_\text{map} = \frac{2PC(I_d)\cdot PC(I_p) + C_3}{PC(I_d)^2 + PC(I_p)^2 + C_3} \tag{22}$$

and C_2 and C_3 are two fixed constants similar to C_1.

C. Feature Group Three

The third group has four features (f_{20}-f_{23}). In [55], it was found that the decorrelating function can be acquired by applying a local non-linear operation to log-contrast luminance to remove local mean displacements from zero log-contrast and to normalize the local variance of the log-contrast. Furthermore, these normalized luminance values highly tend towards a unit normal Gaussian characteristic for natural images, which has been employed to model the contrast-gain masking process in early human vision [56]. We therefore first compute mean subtracted contrast normalized coefficients of the distorted image I_d following the method used in [33] and [35].

Next, we suppose that the distribution of above-mentioned coefficients have characteristic statistical properties, which are changed when distortions are exerted. For instance, as found by Ruderman [55], those coefficients of natural images exhibits a Gaussian-like appearance, while the Gaussian blur makes those coefficients a more Laplacian appearance. Besides, it has been found that the generalized Gaussian distribution (GGD) can be used to validly catch a wider spectrum of statistics of distorted images. So this paper estimates the GGD with zero mean using the associated definition provided in [57]:

$$f(x;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})}\exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \tag{23}$$

where

$$\beta = \sigma\sqrt{\frac{\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{3}{\alpha})}} \tag{24}$$

and the gamma function $\Gamma(\cdot)$ is given by:

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt \qquad a > 0. \tag{25}$$

In Eq. (23), the parameter α manipulates the 'shape' of the distribution, while the other parameter σ^2 indicates the variance of the distribution. In this research, the zero mean distribution is selected owing to the generally symmetric distribution of MSCN coefficients. We deploy this parametric model to fit the MSCN empirical distributions from distorted images as well as undistorted ones. For each image, we estimate two pairs of parameters (α, σ^2) from a GGD fit of the MSCN coefficients at two scales - the original scale as well as at a reduced resolution via a low-pass filtering followed by a downsampling with the factor of 2. These constitute the last group of features.

D. Quality Evaluation

After feature extraction, we need to find a suitable mapping that is learnt from the feature space to subjective MOS values using a regression module, and then apply it to yield objective quality scores. Of course, any regressor can be used here. To show the effectiveness of extracted features and make a fair comparison with the state-of-the-art, in this paper we utilize SVR [44] following the method in BRISQUE [33]. Here, the LIBSVM package [58] is adopted to implement SVR with a radial basis function (RBF) kernel. The Matlab code of our metric will be shown at http://multimedia.sjtu.edu.cn/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

III. Experimental Results and Analysis

In this section, we will test the prediction accuracy of the proposed algorithm from three aspects: 1) to demonstrate the effectiveness of our NFERM compared to classical FR IQA and state-of-the-art NR IQA metrics on the LIVE database; 2) to testify the robustness of NFERM through cross-validation experiments on TID2008, CSIQ, IVC and Toyama databases; 3) to compare the performances of three groups of features used in NFERM with each other.

A. Correlation with Human Opinions on LIVE

We first evaluate the performance of the proposed NFERM with a total number of fifteen IQA approaches on the LIVE database [39]: 1) classical FR PSNR, SSIM [4], MS-SSIM [5]; 2) popular blind distortion-specific WNJE [21], MBBM [22], Sheikh [23], SINE [24], JNB [26], CPBD [27], NFEQM [17]; 3) state-of-the-art general-purpose NR DIIVINE [31], BLIINDS-II [32], BRISQUE [33], NIQE [35], QAC [36]. The LIVE database includes 779 distorted images by corrupting 29 pristine versions with five frequently encountered distortion types: JP2K, JPEG, Additive White Gaussian Noise (AWGN), Gaussian blur (Blur), and a Rayleigh fast-fading channel simulation (FF). The subjective test was separately carried out with each distortion type, and the corresponding DMOS value for each distorted image was obtained accordingly.

To account for the correlation performance of our NFERM, a training procedure is required to calibrate the regressor module. Similar to the usual training method, we in this work randomly separate those above 779 distorted images into two subsets. One is the training set which consists of distorted images corresponding to 80% original images, and the other is the testing set containing the rest 20% distorted images. In order to ensure that the proposed NFERM is robust across image contents and is not governed by the specific train-test split, we repeat this random 80% train - 20% test procedure 1000 times, and report the median result of the performance across these 1000 iterations so as to eliminate performance bias as much as possible.

In general, a nonlinear regression suggested by VQEG [59] is first applied to map objective quality scores of testing IQA metrics to subjective human ratings using the four-parameter logistic function:

$$q(\epsilon) = \frac{\xi_1 - \xi_2}{1 + \exp\left(-\frac{\epsilon - \xi_3}{\xi_4}\right)} + \xi_2 \tag{26}$$

where ϵ and $q(\epsilon)$ are respectively the input score and the mapped score, and ξ_j $(j = 1, 2, 3, 4)$ are free parameters to be determined during the curve fitting process. We then compute two commonly used performance measures, Spearman's rank ordered correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC), between the objective quality predictions and subjective DMOS values to evaluate those IQA methods' performances. The SROCC is defined as

$$\text{SROCC} = 1 - \frac{6}{N(N^2 - 1)} \sum_{i=1}^{N} d_i^2 \tag{27}$$

TABLE IV: Correlation performance of FR PSNR, SSIM, MS-SSIM, blind distortion-specific WNJE, MBBM, Sheikh [23], SINE, JNB, CPBD, NFEQM, state-of-the-art general-purpose NR DIIVINE, BLIINDS-II, BRISQUE, NIQE, QAC, and the proposed NFERM (the median value across 1000 times training) on LIVE and its five various distortion types. We bold the best three performed metrics.

SROCC	Type	JP2K (169)	JPEG (175)	AWGN (145)	Blur (145)	FF (145)	All (779)
PSNR	FR	0.8954	0.8809	**0.9854**	0.7823	**0.8907**	0.8756
SSIM	FR	**0.9355**	0.9449	0.9629	0.8944	**0.9413**	0.9104
MS-SSIM	FR	**0.9654**	**0.9794**	0.9745	**0.9587**	**0.9315**	**0.9448**
WNJE	NR	–	**0.9735**	–	–	–	–
MBBM	NR	–	–	–	0.9015	–	–
Sheikh	NR	0.9130	–	–	–	–	–
SINE	NR	–	–	**0.9837**	–	–	–
JNB	NR	–	–	–	0.7871	–	–
CPBD	NR	–	–	–	0.9186	–	–
NFEQM	NR	–	–	0.9682	0.8845	–	–
DIIVINE	NR	0.9123	0.9208	0.9818	**0.9373**	0.8694	0.9250
BLIINDS-II	NR	0.9323	0.9331	0.9463	0.8912	0.8519	0.9250
BRISQUE	NR	0.9139	**0.9647**	0.9786	**0.9511**	0.8768	**0.9395**
NIQE	NR	0.9187	0.9422	0.9718	0.9329	0.8639	0.9086
QAC	NR	0.8621	0.9362	0.9509	0.9134	0.8231	0.8683
NFERM	NR	**0.9415**	0.9645	**0.9838**	0.9219	0.8627	**0.9405**

PLCC	Type	JP2K (169)	JPEG (175)	AWGN (145)	Blur (145)	FF (145)	All (779)
PSNR	FR	0.8996	0.8879	**0.9858**	0.7835	0.8895	0.8701
SSIM	FR	**0.9410**	0.9504	0.9695	0.8743	**0.9428**	0.9014
MS-SSIM	FR	**0.9697**	**0.9814**	0.9724	**0.9530**	**0.9200**	**0.9338**
WNJE	NR	–	**0.9786**	–	–	–	–
MBBM	NR	–	–	–	0.9194	–	–
Sheikh	NR	0.9201	–	–	–	–	–
SINE	NR	–	–	0.9796	–	–	–
JNB	NR	–	–	–	0.8160	–	–
CPBD	NR	–	–	–	0.8953	–	–
NFEQM	NR	–	–	0.9708	0.8921	–	–
DIIVINE	NR	0.9233	0.9347	**0.9867**	0.9370	0.8916	0.9270
BLIINDS-II	NR	0.9386	0.9426	0.9635	0.8994	0.8790	0.9164
BRISQUE	NR	0.9229	0.9734	0.9851	**0.9506**	**0.9030**	**0.9424**
NIQE	NR	0.9262	0.9523	0.9763	**0.9434**	0.8794	0.9054
QAC	NR	0.8648	0.9435	0.9180	0.9105	0.8248	0.8625
NFERM	NR	**0.9548**	**0.9817**	**0.9915**	0.9371	0.8878	**0.9457**

where d_i is the difference between the i-th image's ranks in subjective and objective evaluations, and N indicates the image number in the testing database. The SROCC is a non-parametric rank-based correlation metric, independent of any monotonic nonlinear mapping between subjective ratings and objective scores. The PLCC is calculated by

$$\text{PLCC} = \frac{\sum_i (q_i - \bar{q}) \cdot (o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 \cdot \sum_i (o_i - \bar{o})^2}} \tag{28}$$

where o_i and $\bar{o}$ are the i-th image's subjective rating and the mean of the overall o_i. q_i and $\bar{q}$ are the i-th image's converted objective score after nonlinear regression and their mean value. A value close to 1 for SROCC and PLCC indicates superior performance in terms of correlation between subjective human opinions and objective quality predictions.

We tabulate those performance indices of competitive IQA models in Table IV. It is apparent that the proposed metric highly correlates with human opinion ratings. More concretely, our NFERM outperforms state-of-the-art general-purpose NR IQA methods, especially on all images as well as images of white noise and JP2K compression. In addition, the prediction accuracy of the NFERM is completely higher than popular

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

TABLE V: The SROCC values of FR PSNR, SSIM, MS-SSIM, blind distortion-specific WNJE, MBBM, Sheikh [23], SINE, JNB, CPBD, NFEQM, state-of-the-art general-purpose NR DIIVINE, BLIINDS-II, BRISQUE, NIQE, QAC, our NFERM on TID2008, CSIQ, IVC, Toyama databases and associated various distortion types, and the database size-weighted averages. We bold the top three performed metrics.

Database		TID2008 [40]					CSIQ [41]				
Metrics	Type	JP2K (96)	JPEG (96)	AWGN (96)	Blur (96)	All (384)	JP2K (150)	JPEG (150)	AWGN (150)	Blur (150)	All (600)
PSNR	FR	0.8248	0.8753	**0.9177**	**0.9335**	0.8703	**0.9362**	0.9019	**0.9363**	**0.9291**	**0.9219**
SSIM	FR	0.8785	**0.9248**	0.8110	**0.9444**	0.7678	**0.9207**	0.9222	**0.9255**	**0.9245**	0.8767
MS-SSIM	FR	**0.9727**	**0.9391**	0.8190	**0.9630**	**0.8973**	**0.9707**	**0.9626**	0.9088	**0.9728**	**0.9416**
WNJE	NR	–	0.9212	–	–	–	–	**0.9551**	–	–	–
MBBM	NR	–	–	–	0.7852	–	–	–	–	0.8768	–
Sheikh	NR	0.3093	–	–	–	–	0.5697	–	–	–	–
SINE	NR	–	–	**0.8885**	–	–	–	–	**0.9542**	–	–
JNB	NR	–	–	–	0.7143	–	–	–	–	0.7624	–
CPBD	NR	–	–	–	0.8542	–	–	–	–	0.8853	–
NFEQM	NR	–	–	0.8074	0.7407	–	–	–	0.8380	0.8939	–
DIIVINE	NR	0.8419	0.5805	**0.8322**	0.8150	0.7749	0.8308	0.7996	0.8663	0.8716	0.8284
BLIINDS-II	NR	0.8968	0.8620	0.6062	0.8388	0.7985	0.8951	0.8986	0.7597	0.8766	0.8511
BRISQUE	NR	**0.9037**	0.9101	0.8227	0.8742	**0.8978**	0.8665	0.9040	0.9252	0.9025	0.8990
NIQE	NR	0.8939	0.8756	0.7775	0.8249	0.8006	0.9065	0.8826	0.8098	0.8944	0.8717
QAC	NR	0.8953	0.8773	0.5929	0.8408	0.8538	0.8699	0.9016	0.8222	0.8362	0.8416
NFERM	NR	**0.9474**	**0.9365**	0.8281	0.8436	**0.9156**	0.9051	**0.9223**	0.9220	0.8964	**0.9142**

Database		IVC [42]				Toyama [43]			Average				
Metrics	Type	JP2K (50)	JPEG (50)	Blur (20)	All (120)	JP2K (84)	JPEG (84)	All (168)	JP2K (380)	JPEG (380)	AWGN (246)	Blur (266)	All (1272)
PSNR	FR	0.8500	0.6740	0.8051	0.7708	0.8605	0.2868	0.6132	0.8800	0.7292	**0.9290**	**0.9214**	0.8513
SSIM	FR	0.8501	0.8067	0.8691	**0.8424**	**0.9148**	0.6263	0.7870	**0.8994**	0.8423	0.8808	**0.9275**	0.8287
MS-SSIM	FR	**0.9320**	**0.9221**	**0.9443**	**0.9154**	**0.9470**	0.8360	**0.8870**	**0.9609**	**0.9233**	0.8738	**0.9671**	**0.9185**
WNJE	NR	–	**0.9451**	–	–	–	**0.8829**	–	–	**0.9293**	–	–	–
MBBM	NR	–	–	**0.8758**	–	–	–	–	–	–	–	0.8437	–
Sheikh	NR	0.7759	–	–	–	0.8649	–	–	0.6019	–	–	–	–
SINE	NR	–	–	–	–	–	–	–	–	–	**0.9286**	–	–
JNB	NR	–	–	0.6659	–	–	–	–	–	–	–	0.7378	–
CPBD	NR	–	–	0.7690	–	–	–	–	–	–	–	0.8653	–
NFEQM	NR	–	–	0.0158	–	–	–	–	–	–	0.8261	0.7726	–
DIIVINE	NR	0.6535	0.3528	0.5185	0.3300	0.6114	0.7023	0.6416	0.7618	0.6640	0.8530	0.8246	0.7406
BLIINDS-II	NR	0.7495	0.7705	0.5262	0.5481	0.7222	**0.8678**	0.7967	0.8382	0.8657	0.6998	0.8366	0.7995
BRISQUE	NR	0.8331	0.8020	0.8239	0.8155	0.7970	**0.8690**	**0.8572**	0.8561	0.8844	0.8852	0.8864	**0.8852**
NIQE	NR	**0.8507**	0.8451	0.8638	0.7915	**0.8762**	0.8378	0.8128	0.8893	0.8660	0.7972	0.8670	0.8349
QAC	NR	0.8022	0.9135	0.8405	0.7676	0.5629	0.6714	0.5189	0.7995	0.8461	0.7327	0.8382	0.7957
NFERM	NR	**0.9177**	**0.9395**	**0.9120**	**0.8871**	0.8741	0.8638	**0.8497**	**0.9106**	**0.9152**	**0.8854**	0.8785	**0.9035**

blind distortion-specific measures used in this paper. And furthermore, although FR IQA approaches are considered hardly matchable with NR IQA metrics owing to the unavailability of original images, our NFERM technique is still better than the benchmark PSNR and SSIM, and is comparable to MS-SSIM.

Besides direct comparisons with numerous IQA metrics, we further evaluate the statistical significance using the t-test [60], which is used to determine if two sets of data are significantly different from each other and is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known, on those IQA methods' SROCC values obtained from the 1000 train-test trials. The null hypothesis is that the mean correlation for our NFERM is equal to mean correlation for the column algorithm with a confidence of 95%. A value of '1' in the table indicates that NFERM is statically superior to the column algorithm, whereas a '-1' indicates that NFERM is statistically worse than the column. A value of '0' indicates that NFERM and the column algorithm are statistically indistinguishable (or equivalent), i.e., we could not reject the null hypothesis at the 95% confidence level. Table VI provides the statistical results of performance between the NFERM and each competing IQA approaches considered. From Table VI we can conclude that NFERM is statistically better than all of state-of-the-art NR IQA algirhtms tested and FR PSNR and SSIM, as well as on par with MS-SSIM. It is worth noting that, in addition to the advanced performance, the proposed NFERM only adopts 23 features, much less than the 36 features used in the currently best-performed NR BRISQUE.

B. Cross-Validation on Other Databases

Having testified our technique on the LIVE database, we want to prove that the proposed NFERM is not limited to LIVE. To show this, we use the whole images in LIVE to train NFERM, and then apply it to other four image quality databases as follows:

- The TID2008 database [40] is composed of 25 original images and totally 1700 distorted images over 17 distortion types at 4 distortion levels. Those distortion categories include: AWGN (#01), additive noise in color components is more intensive than additive noise in the luminance component (#02), spatially correlated noise (#03), masked noise (#04), high frequency noise (#05), impulse noise (#06), quantization noise (#07), Blur (#08), image denoising (#09),

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

TABLE VI: Results of one-sided t-test performed between SROCC values of the proposed NFERM and various IQA metrics on LIVE. A value of "1" indicates that the NFERM is statically superior to the column algorithm; "-1" indicates that the NFERM is worse than the column; a value of "0" gives indicates that the two algorithms are statically indistinguishable.

t-test	PSNR	SSIM	MS-SSIM	DIIVINE	BLIINDS-II	BRISQUE	NIQE	QAC
LIVE	1	1	0	1	1	1	1	1

TABLE VII: Performance comparison between our NFERM and other IQA methods with f-test and t-test. The symbol "1", "0" or "-1" means that the proposed NFERM is statistically (with 95% confidence) better, undistinguishable, or worse than the corresponding algorithms.

f-test	PSNR	SSIM	MS-SSIM	DIIVINE	BLIINDS-II	BRISQUE	NIQE	QAC
TID2008	1	1	0	1	1	0	1	1
CSIQ	1	1	0	1	1	1	1	1
IVC	1	1	0	1	1	1	1	1
Toyama	1	0	-1	1	0	0	0	1

t-test	PSNR	SSIM	MS-SSIM	DIIVINE	BLIINDS-II	BRISQUE	NIQE	QAC
TID2008	1	1	0	1	1	0	1	1
CSIQ	1	1	0	1	1	1	1	1
IVC	1	1	0	1	1	1	1	1
Toyama	1	0	-1	1	0	0	0	1

TABLE VIII: The SROCC results of NFERM and state-of-the-art NR IQA metrics (BRISQUE, NIQE and QAC) on each distortion type in TID2008, CSIQ, IVC and Toyama databases. We emphasize the best performed NR IQA algorithm in each type.

	TID2008																
SROCC	# 01	# 02	# 03	# 04	# 05	# 06	# 07	# 08	# 09	# 10	# 11	# 12	# 13	# 14	# 15	# 16	# 17
NFERM	**0.8281**	**0.8389**	0.2126	0.1446	**0.9125**	0.0541	0.6655	0.8436	**0.6589**	**0.9365**	**0.9474**	0.1174	0.1817	0.0691	0.0777	0.0524	0.2419
BRISQUE	0.8227	0.7468	0.5691	0.6227	0.6285	0.6070	0.7399	**0.8742**	0.6354	0.9101	0.9037	**0.3457**	0.3156	0.0858	0.1703	0.1111	0.0585
NIQE	0.7775	0.6853	**0.7447**	**0.7562**	0.8632	0.7133	**0.8010**	0.8249	0.6260	0.8756	0.8939	0.1618	**0.5853**	**0.1090**	0.1795	0.1376	0.0405
QAC	0.5929	0.6911	0.1162	0.7294	0.8004	**0.8603**	0.5592	0.8408	0.4533	0.8773	0.8953	0.0537	0.4612	0.0956	**0.3483**	**0.3094**	**0.2588**

	CSIQ						IVC					Toyama	
SROCC	JP2K	JPEG	AWGN	Blur	APGN	CC	JP2K	JPEG	Blur	JPEG_LC	LAR	JP2K	JPEG
NFERM	0.9051	**0.9223**	0.9220	0.8964	**0.6264**	**0.3774**	**0.9177**	**0.9395**	**0.9120**	0.7943	0.8855	0.8741	0.8638
BRISQUE	0.8665	0.9040	**0.9252**	**0.9025**	0.2529	0.0473	0.8331	0.8020	0.8239	0.6830	0.7539	0.8706	**0.8690**
NIQE	**0.9065**	0.8826	0.8098	0.8944	0.2993	0.2292	0.8507	0.8451	0.8638	0.5532	0.7283	**0.8762**	0.8378
QAC	0.8699	0.9016	0.8222	0.8362	0.0019	0.2446	0.8022	0.9135	0.8405	**0.8771**	**0.9266**	0.5629	0.6714

JPEG (#10), JP2K (#11), JPEG transmission errors (#12), JP2K transmission errors (#13), non-eccentricity pattern noise (#14), local block-wise distortion of different intensity (#15), mean shift (#16), and contrast change (#17). There exists one artificial image in source images, so we selected the rest 24 natural images and their corresponding 1632 counterparts as the testing bed, because the features used in state-of-the-art NR IQA algorithms and NFERM mainly rely on natural images.

- The CSIQ database [41] uses six distortions types (JP2K, JPEG, AWGN, Blur, Additive Pink Gaussian Noise (APGN) and contrast change (CC)) at four to five distortion levels to produce 866 distorted images from 30 original ones.
- The IVC database [42] covers 185 images created from 10 pristine images. Those distortion types are as follows: 1) JP2K (50 images); 2) JPEG (50 images); 3) Blur (20 images); 4) JPEG_LUMICHR (25 images); 5) Local adaptive resolution (LAR) coding (40 images).
- The Toyama database [43] includes the frequently used JP2K and JPEG compressions, each of which consists of 84 distorted images generated from 12 source versions.

The SROCC is an important criteria in IQA performance measures, which illustrates the monotonicity and convergency between the objective quality metric and the subjective human perception. Also, SROCC has been widely used to search for the suitable parameters in some existing IQA approaches [6], [14], [61]. So we apply SROCC in this cross-validation test to measure and compare various IQA methods.

Table V presents the performance evaluations of NFERM on TID2008, CSIQ, IVC and Toyama databases, and also reports those testing IQA metrics' SROCC results. For a more direct and clear comparison, we compute the average values across four image databases above, which is defined by

$$\bar{\delta} = \frac{\sum_i \delta_i \cdot \pi_i}{\sum_i \pi_i} \tag{29}$$

where δ_i $(i = 1, 2, 3, 4)$ indicates the correlation measure for each database, and π is the number of images in each database (i.e. 384 for TID2008, 600 for CSIQ, 120 for IVC, 168 for Toyama) or each associated subset. We tabulate these average results in Table V. It is very clear that the proposed NFERM shows better correlation with human opinions on the overall images and each distortion type.

We further measure the statistical significance of NFERM using the f-test, which is most often used when comparing statistical models that have been fitted to a data set to identify the model that best fits the population from which the data

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

TABLE IX: SROCC and PLCC values (after nonlinear regression) of each group of features in the proposed NFERM (the median value across 1000 times training) on the whole 779 images in LIVE and the five various distortion categories.

SROCC	JP2K	JPEG	AWGN	Blur	FF	All
NFERM (f_{01}-f_{13})	0.9034	0.9532	0.9754	0.9146	0.7746	0.8854
NFERM (f_{14}-f_{19})	0.7715	0.8397	0.9677	0.8062	0.7923	0.8047
NFERM (f_{20}-f_{23})	0.8294	0.8823	0.9631	0.8904	0.7954	0.8429
NFERM (f_{01}-f_{23})	0.9415	0.9645	0.9838	0.9219	0.8627	0.9405

PLCC	JP2K	JPEG	AWGN	Blur	FF	All
NFERM (f_{01}-f_{13})	0.9160	0.9691	0.9824	0.9137	0.8216	0.8901
NFERM (f_{14}-f_{19})	0.7933	0.9016	0.9784	0.8417	0.8314	0.8236
NFERM (f_{20}-f_{23})	0.8454	0.8880	0.9782	0.8961	0.8451	0.8466
NFERM (f_{01}-f_{23})	0.9548	0.9817	0.9915	0.9371	0.8878	0.9457

TABLE X: SROCC and PLCC values of MS-SSIM, BRISQUE, and our NFERM and its three groups of features (the median value across 1000 times training) on four commonly encountered distortion types in LIVE (JP2K, JPEG, AWGN, Blur) and their overall 634 images.

SROCC	Type	JP2K	JPEG	AWGN	Blur	All
NFERM (f_{01}-f_{13})	NR	0.9430	0.9642	0.9831	0.9077	0.9535
NFERM (f_{14}-f_{19})	NR	0.8631	0.9445	0.9538	0.4377	0.8329
NFERM (f_{20}-f_{23})	NR	0.8061	0.9313	0.9546	0.2862	0.8064
NFERM (f_{01}-f_{23})	NR	0.9408	0.9632	0.9838	0.9196	0.9597
BRISQUE	NR	0.9196	0.9622	0.9769	0.9569	0.9583
MS-SSIM	FR	0.9654	0.9794	0.9745	0.9587	0.9510

PLCC	Type	JP2K	JPEG	AWGN	Blur	All
NFERM (f_{01}-f_{13})	NR	0.9533	0.9810	0.9902	0.9243	0.9576
NFERM (f_{14}-f_{19})	NR	0.8915	0.9600	0.9601	0.5970	0.8357
NFERM (f_{20}-f_{23})	NR	0.8240	0.9540	0.9643	0.3527	0.8205
NFERM (f_{01}-f_{23})	NR	0.9544	0.9812	0.9918	0.9378	0.9632
BRISQUE	NR	0.9370	0.9767	0.9877	0.9639	0.9613
MS-SSIM	FR	0.9697	0.9814	0.9724	0.9530	0.9383

are sampled, to compute the prediction residuals between converted objective predictions (after the nonlinear mapping) and subjective scores. We list the statistical significance between our algorithm and other competing IQA metrics in comparison in Table VII, where the symbol "1", "0" or "-1" means that the proposed metric is statistically (with 95% confidence) better, indistinguishable, or worse than the corresponding IQA approach. The t-test is also used here. It is easy to find that, in each subset, NFERM is comparable to FR MS-SSIM and a few distortion-specific blind measures, while superior to other competitors. In the meantime, our algorithm definitely outperforms mainstream FR IQA methods and state-of-the-art NR IQA algorithms, yet slightly inferior to the powerful MS-SSIM on average. It is worth mentioning that, on frequently encountered distortion types, our NFERM works with higher accuracy than state-of-the-art NR IQA metrics and less features (only 23 features) than the currently best-performed SVM-based BRISQUE of 36 features.

Those performance indices above have confirmed the validness of the proposed model across a broad range of image scenes, since original images in CSIQ and IVC are greatly distinct from those in LIVE which our NFERM is trained on. We further compare the correlation accuracy of NFERM and state-of-the-art NR IQA metrics (BRISQUE, NIQE and QAC) on each distortion category in TID2008, CSIQ, IVC and Toyama databases. We report the SROCC values in Table VIII, and label the best performed metric in each type. Our NFERM has won twelve times the first place, whereas BRISQUE, NIQE and QAC independently have five, seven and six times. This also validates the effectiveness of the proposed NFERM metric across various distortion types compared to state-of-the-art NR IQA algorithms.

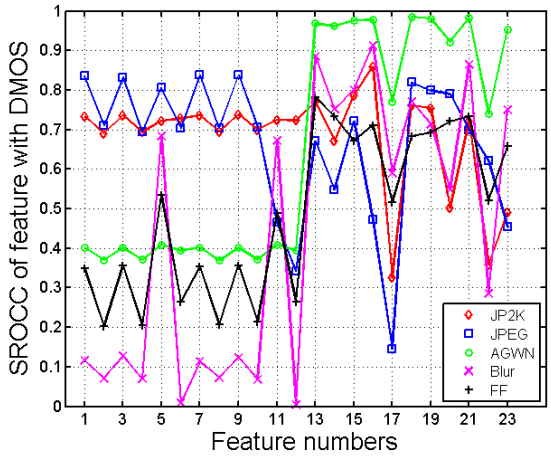


Fig. 8: Correlation of features with human judgments of quality (DMOS) for different distortion categories.

C. Analysis of NFERM's Components

Considering that the proposed NFERM is constituted by three groups of features, it is natural to compare the performance of each group of features. The first group includes f_{01}-f_{13}, the second group includes f_{14}-f_{19}, and the last one includes f_{20}-f_{23}. We first plot in Fig. 8 the SROCC results between each of those extracted features and DMOS values on each distortion type in the LIVE database, so as to ascertain how well the features correlate with human judgments of quality. We then compute the median values of PLCC and SROCC across the 1000 times random 80% train - 20% iterations following the method in Section III-A, and report the results in Table IX.

We have three important findings from above performance comparisons. First, the first group of features is more effective than the other two. This finding may be explained by the fact that the human visual perception to image quality mainly depend upon two strategies: the decomposition and the synthesization. The DCT and DWT decompositions have been widely applied in existing IQA models [4], [7], [10], [32]. The synthesization was recently developed with internal generative model to approximate the human visual sensation of image quality, which has brought remarkable IQA performance gain [13], [17], [34], [38]. The first group of features was proposed to fuse the structural degradation model (decomposition) and the free energy feature about the brain theory (synthesization), thereby acquiring considerably high performance.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

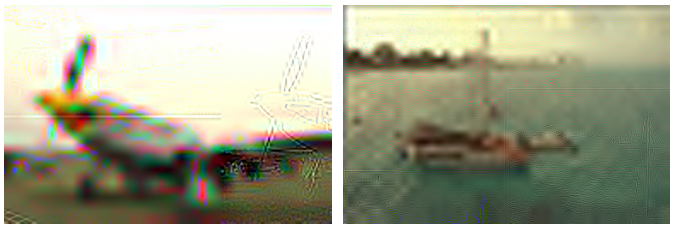


Fig. 9: Sample images of false contours in the LIVE FF subset.

Second, each of three groups of features works ineffectively for the FF distortion. To illustrate this, we also testify the performances of FR MS-SSIM, NR BRISQUE, as well as our NFERM and its three groups of features on four distortion types in LIVE, i.e. JP2K (169 images), JPEG (175 images), AWGN (145 images), Blur (145 images), and their overall 634 images. The performance evaluations are listed in Table X, which once again supports this finding and demonstrates the superior performance of the first group of features in NFERM. In practice, the FF distortion is quite different from our commonly encountered distortion types (e.g. JP2K, JPEG, noise and blur). For instance, We in Fig. 9 present a pair of images of obvious false contours in the LIVE FF subset. Since the features used in NFERM target to characterize natural images, the proposed method cannot work very well on the FF distortion in theory. Of course, there still is some room for performance improvement by considering features that are good at measuring FF distorted images.

Third, it needs to point out that three groups of features in NFERM use various strategies. The first group of features is motivated by a novel strategy of combining two effective RR IQA metrics. The second group is inspired by the HVS. And the third group is to quantify the possible losses of 'naturalness' in distorted images. In fact, from the testing results in Table IV, IX-X, we can easily find that each group of features performs well, and the whole 23 features have even better performance.

IV. Conclusion

In this paper, we have proposed a new NR NFERM quality metric with the recently revealed free energy principle and important HVS inspired features before the SVM-based regression module. A comparison of our NFERM with classical FR IQA methods, popular blind distortion-specific measures, and state-of-the-art general-purpose NR IQA models is conducted on five popular databases (LIVE, TID2008, CSIQ, IVC and Toyama). Experimental results confirm the superior performance of our introduced NR IQA algorithm on LIVE through 1000 times 80 % train - 20% test splits, and on other four databases and each distortion type through across-validation testings. Besides the substantially high prediction accuracy, it is worth emphasizing two points: First, the proposed NFERM only needs 23 features, far less than 36 features used in the currently best-performed SVM-based BRISQUE; Second, a new framework for the design of higher-performance and less-features NR IQA metric is proposed in this work to combine the merits of effective FR, RR and NR IQA approaches.

V. Appendix

For the AR image model $x_n = \boldsymbol{\chi}^k(x_n)\boldsymbol{\lambda}+\varepsilon_n$ with x_n being the pixel and $\boldsymbol{\chi}^k(x_n)$ being the row vector including k nearest neighbors of x_n, consider the data set $I = \{x_n, \boldsymbol{\chi}^k(x_n)\}$, $n = 1, 2, \ldots, N$, containing N pixels and their corresponding support vectors. The likelihood of the data set is given by a quadratic function of $\boldsymbol{\lambda}$ as follows:

$$P(I|\boldsymbol{\lambda}, \beta, \mathcal{H}) = \left(\frac{\beta}{2\pi}\right)^{N/2} \exp\left(-\beta E_I(\boldsymbol{\lambda})\right) \tag{30}$$

with

$$E_I(\boldsymbol{\lambda}) = \frac{1}{2}\sum_{n=1}^{N}(x_n - \boldsymbol{\chi}^k(x_n)\boldsymbol{\lambda})^2 = \frac{1}{2}\left(\boldsymbol{X} - \boldsymbol{\mathcal{X}}\boldsymbol{\lambda}\right)^T\left(\boldsymbol{X} - \boldsymbol{\mathcal{X}}\boldsymbol{\lambda}\right) \tag{31}$$

where $\boldsymbol{X}$ is a column vector with its n-th entry being x_n and $\boldsymbol{\mathcal{X}}$ is a matrix with the n-th row being $\boldsymbol{\chi}^k(x_n)$.

The AR coefficients in $\boldsymbol{\lambda}$ are assumed to be drawn from a zero mean spherical Gaussian distribution with precision α as follows:

$$P(\boldsymbol{\lambda}|\alpha, \mathcal{H}) = \left(\frac{\alpha}{2\pi}\right)^{k/2} \exp\left(-\alpha E_\lambda\right) \tag{32}$$

where $E_\lambda = \frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{\lambda} = \frac{1}{2}\sum_{i=1}^{k}\lambda_i^2$.

Assuming that α and β are drawn from a couple of Gamma priors

$$\begin{aligned} P(\alpha|\mathcal{H}) &= \Gamma(\alpha; b_\alpha, c_\alpha) \\ P(\beta|\mathcal{H}) &= \Gamma(\beta; b_\beta, c_\beta) \end{aligned} \tag{33}$$

where the Gamma distribution, i.e.,

$$\Gamma(x; b, c) = \frac{1}{\Gamma(c)}\frac{x^{c-1}}{b^c}\exp\left(\frac{-x}{b}\right), 0 \le x \le \infty \tag{34}$$

has mean bc and variance b^2c.

If we write all the priors into one parameter vector $\boldsymbol{\theta}$, then

$$P(\boldsymbol{\theta}) = P(\boldsymbol{\lambda}|\alpha)P(\alpha)P(\beta). \tag{35}$$

We let $P(I, \mathbf{g}) = P(I|\mathbf{g})P(\mathbf{g})$ in Eq. (6) and have

$$\begin{aligned} \mathcal{J}(\mathbf{g}) &= \int Q(\mathbf{g}|I)\log\frac{Q(\mathbf{g}|I)}{P(I|\mathbf{g})P(\mathbf{g})}d\mathbf{g} \\ &= \int Q(\mathbf{g}|I)\log\frac{Q(\mathbf{g}|I)}{P(\mathbf{g})}d\mathbf{g} - \int Q(\mathbf{g}|I)\log P(I|\mathbf{g})d\mathbf{g} \\ &= KL(Q(\mathbf{g}|I)||P(\mathbf{g})) + E_Q[\log P(I|\mathbf{g})] \end{aligned} \tag{36}$$

where $KL(Q(\mathbf{g}|I)||P(\mathbf{g}))$ measures the distance between the recognition density and the true prior density of the model parameters, and it attains zeros only when $Q(\mathbf{g}|I) = P(\mathbf{g})$. $E_Q[\log P(I|\mathbf{g})]$ measures the average likelihood of the data over the approximating posterior density. The KL divergence term splits into 3 sub-divergences terms over $\boldsymbol{\lambda}, \alpha$ and β:

$$KL\left(Q(\mathbf{g}|I)\|P(\mathbf{g})\right) = KL(\boldsymbol{\lambda}) + KL(\alpha) + KL(\beta). \tag{37}$$

Specifically, the KL divergence between Normal densities $q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$ and $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ is

$$\begin{aligned} KL(q\|p) = \frac{1}{2}\Bigg(&\log\frac{|\boldsymbol{\Sigma}_p|}{|\boldsymbol{\Sigma}_q|} + Tr\left(\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\Sigma}_q\right) \\ &+ (\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)^T\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_q - \boldsymbol{\mu}_p) - d\Bigg) \end{aligned} \tag{38}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

where d is the dimension of the variable $\boldsymbol{x}$. The KL divergence between Gamma densities $q(\boldsymbol{x}) = \Gamma(\boldsymbol{x}; b_q, c_q)$ and $p(\boldsymbol{x}) = \Gamma(\boldsymbol{x}; b_p, c_p)$ is

$$KL(q\|p) = (c_q - 1)\Psi(c_q) - \log b_q - \log\Gamma(c_q) + \log\Gamma(c_p) - c_q + c_p \log b_p - (c_p - 1)(\Psi(c_q) + \log b_q) + \frac{b_q c_q}{b_p}. \quad (39)$$

According to the analysis by in [62]-[63], the split optimization process gives the following posterior distributions

$$\begin{aligned} Q(\boldsymbol{\lambda}|I) &= \mathcal{N}(\boldsymbol{\lambda}; \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\Sigma}}) \\ Q(\alpha|I) &= \Gamma(\alpha; b'_\alpha, c'_\alpha) \\ Q(\beta|I) &= \Gamma(\beta; b'_\beta, c'_\beta) \end{aligned} \quad (40)$$

where

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \left(\hat{\beta}\boldsymbol{X}^T\boldsymbol{X} + \hat{\alpha}\boldsymbol{I}\right)^{-1} \\ \hat{\boldsymbol{\lambda}} &= \hat{\boldsymbol{\Sigma}}\boldsymbol{X}^T\hat{\beta}\boldsymbol{Y} \\ b'_\alpha &= \left(\frac{1}{2}\hat{\boldsymbol{\lambda}}^T\hat{\boldsymbol{\lambda}} + \frac{1}{2}Tr(\hat{\boldsymbol{\Sigma}}) + \frac{1}{b_\alpha}\right)^{-1} \\ c'_\alpha &= \frac{k}{2} + c_\alpha \\ \hat{\alpha} &= b'_\alpha c'_\alpha \\ b'_\beta &= \left(E_D(\hat{\boldsymbol{\lambda}}) + \frac{1}{2}Tr(\hat{\boldsymbol{\Sigma}}\boldsymbol{X}^T\boldsymbol{X}) + \frac{1}{b_\beta}\right)^{-1} \\ c'_\beta &= \frac{N}{2} + c_\beta \\ \hat{\beta} &= b'_\beta c'_\beta. \end{aligned} \quad (41)$$

Normally, uninformative priors are used for the Gamma distributions in Eq. (33), e.g., $b = 10^3$, $c = 10^{-3}$. However, informative priors can also be constructed via estimating the mean μ and variance σ^2 of the data (and using equations $\mu = bc$ and $\sigma^2 = b^2c$). The posterior distributions are initialized using the ML estimations $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}_{ML}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{ML}$ with

$$\begin{aligned} \boldsymbol{\lambda}_{ML} &= \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y} \\ \sigma^2_{ML} &= \frac{1}{N-1}\sum_{n=1}^{N}(\boldsymbol{\chi}^k(x_n)\boldsymbol{\lambda}_{ML} - x_n)^2 \\ \boldsymbol{\Sigma}_{ML} &= \sigma^2_{ML}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}. \end{aligned} \quad (42)$$

The free energy term can be written as

$$\begin{aligned} F(\mathbf{g}) =& \frac{N}{2}\left(\Psi(c'_\beta) + \log b'_\beta\right) \\ & - \hat{\beta}\left(E_D(\hat{\boldsymbol{\lambda}}) + \frac{1}{2}Tr(\hat{\boldsymbol{\Sigma}}\boldsymbol{X}^T\boldsymbol{X})\right) - \frac{N}{2}\log 2\pi \end{aligned} \quad (43)$$

with Ψ is the digamma (logarithmic derivative of Γ) function.

The optimization process outlined above can be summarized as a two-step EM-like approach as follows.

1) E-$Step$: With model parameter fixed at $\boldsymbol{\lambda}_{t-1}$, update hyper-parameters α and β to minimize $F(\mathbf{g})$.
2) M-$Step$: With hyper-parameters fixed at α_t and β_t, update model parameter $\boldsymbol{\lambda}$ to minimize $F(\mathbf{g})$.

This process has been shown to be a general case of the EM algorithm. This process is then iterated till the objective function of free energy converges. Further details about the implementation of the aforementioned iterative optimization algorithm can be found in [62]-[63].

REFERENCES

[1] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2008-2024, Sept. 2013,

[2] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, "Three dimensional scalable video adaptation via user-end perceptual quality assessment," *IEEE Trans. Broadcasting*, vol. 54, no. 3, pp. 719-727, Sept. 2008.

[3] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bitrate videos," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316-1324, Nov. 2008.

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

[5] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *in Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, pp. 1398-1402, Nov. 2003.

[6] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193-201, Apr. 2009.

[7] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185-1198, May 2011.

[8] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Self-adaptive scale transform for IQA metric," *in Proc. IEEE Int. Symp. Circuits and Syst.*, pp. 2365-2368, May 2013.

[9] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Structural similarity weighting for image quality assessment," *in Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, pp. 1-6, Jul. 2013.

[10] K. Gu, G. Zhai, M. Liu, Q. Xu, X. Yang, and W. Zhang, "Adaptive high-frequency clipping for improved image quality assessment," *in Proc. IEEE Vis. Commun. Image Process.*, pp. 1-5, Nov. 2013.

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

[12] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500-1512, Apr. 2012.

[13] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43-54, Jan. 2013.

[14] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new psychovisual paradigm for image quality assessment: From differentiating distortion types to discriminating quality conditions," *Signal, Image and Video Processing*, vol. 7, no. 3, pp. 423-436, May 2013.

[15] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it?-A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[16] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824-829, Aug. 2011.

[17] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41-52, Jan. 2012.

[18] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378-3389, Aug. 2012.

[19] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," *in Proc. IEEE Int. Symp. Circuits and Syst.*, pp. 1095-1098, May 2013.

[20] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Subjective and objective quality assessment for images with contrast change," *in Proc. IEEE Int. Conf. Image Process.*, pp. 383-387, Sept. 2013.

[21] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," *in Proc. IEEE Int. Conf. Image Process.*, pp. 477-480, Sept. 2002.

[22] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," *in Proc. IEEE Int. Conf. Image Process.*, pp. 57-60, Sept. 2002.

[23] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 1918-1927, Dec. 2005.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

[24] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," *in Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2209-2216, Sept. 2009.

[25] G. Zhai and X. Wu, "Noise estimation using statistics of natural images," *in Proc. IEEE Int. Conf. Image Process.*, pp. 1857-1860, Sept. 2011.

[26] R. Ferzli and L. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717-728, 2009.

[27] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678-2683, 2011.

[28] K. Gu, G. Zhai, M. Liu, X. Yang, and W. Zhang, "Details preservation inspired blind quality metric of tone mapping methods," *in Proc. IEEE Int. Symp. Circuits and Syst.*, June 2014.

[29] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology Paris*, vol. 100, pp. 70-87, 2006.

[30] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, pp. 127-138, 2010.

[31] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From scene statistics to perceptual quality," *IEEE Trans. Image Process.*, pp. 3350-3364, vol. 20, no. 12, Dec. 2011.

[32] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, pp. 3339-3352, vol. 21, no. 8, Aug. 2012.

[33] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, pp. 4695-4708, vol. 21, no. 12, Dec. 2012.

[34] K. Gu, G. Zhai, X. Yang, W. Zhang, and L. Liang, "No-reference image quality assessment metric by combining free energy theory and structural degradation model," *in Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 1-6, Jul. 2013.

[35] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Letters*, pp. 209-212, vol. 22, no. 3, Mar. 2013.

[36] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," *in Proc. IEEE Int. Conf. Comput. Vis. and Pattern Recognition*, pp. 995-1002, Jun. 2013.

[37] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 24, no. 12, pp. 2013-2026, Dec. 2013.

[38] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcasting*, vol. 60, no. 3, pp. 555-567, Sept. 2014.

[39] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment Database Release 2," [Online]. Available: http://live.ece.utexas.edu/research/quality

[40] 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, pp. 30-45, 2009.

[41] E. C. Larson and D. M. Chandler, "Categorical image quality (CSIQ) database," [Online], Available: http://vision.okstate.edu/csiq

[42] A. Ninassi, P. Le Callet, and F. Autrusseau, "Subjective quality assessment-IVC database," [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb

[43] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, "MICT image quality evaluation database," [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html

[44] B. Schökopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207-1245, 2000.

[45] D. C. Knill and A. Pouget, "The Bayesian brain: The role of uncertainty in neural coding and computation," *Trends Neurosci.*, vol. 27, no. 12, pp. 712-719, 2004.

[46] J. Rissanen and Jr. G. G. Langdon, "Universal modeling and coding," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 12-23, Jan. 1981.

[47] H. Barlow, *Cognitive Psychology*, W. Rosenblith, Ed. Cambridge, MA: MIT Press, 1961.

[48] R. Linsker, "Perceptual neural organisation: Some approaches based on network models and information theory," *Annu. Rev. Neurosci.*, vol. 13, pp. 257-281, 1990.

[49] H. Attias, "A variational bayesian framework for graphical models," *in Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 209-215, 2000.

[50] X. Wu, G. Zhai, X. Yang, and W. Zhang, "Adaptive sequential prediction of multidimensional signals with applications to lossless image coding," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 36-42, Jan. 2011.

[51] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *in Proc. IEEE Int. Conf. Comput. Vis.*, pp. 416-423, 2001.

[52] M. C. Morrone, J. Ross, D. C. Burr, and R. Owens, "Mach bands are phase dependent," *Nature*, vol. 324, pp. 250-253, Nov. 1986.

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

[54] B. Jähne, H. Haubecker, and P. Geibler," *Handbook of Computer Vision and Applications.* New York: Academic, 1999.

[55] D. L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517-548, 1994.

[56] M. Carandini, D. J. Heeger, and J. A. Movshon, "Linearity and normalization in simple cells of the macaque primary visual cortex," *J. Neurosci.*, vol. 17, no. 21, pp. 8621-8644, 1997.

[57] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52-56, Feb. 1995.

[58] C-C. Chang and C-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intelligent Symp. Technol.*, vol. 2, no. 3, 2011. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.

[59] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar. 2000, http://www.vqeg.org/.

[60] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures.* London, U.K.: Chapman & Hall, 2004.

[61] K. Gu, G. Zhai, X. Yang, and W. Zhang, "No-reference stereoscopic IQA approach: From nonlinear effect to parallax compensation," *Journal of Electrical and Computer Engineering*, vol. 2012, Sept. 2012.

[62] D. Mackay, "Ensemble learning and evidence maximization," *in Proc. Adv. Neural Inf. Process. Syst.*, 1995.

[63] S. Roberts and W. Penny, "Variational Bayes for generalised autoregressive models," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2245-2257, Sept. 2002.

PLACE PHOTO HERE

Ke Gu received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009. He is currently working toward the Ph.D. degree. He is the reviewer for several IEEE Transactions and Journals, including IEEE Transactions on Cybernetics, IEEE Signal Processing Letters, Journal of Visual Communication and Image Representation, and Signal, Image and Video Processing.

From July to November 2014, he was a visiting student at the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include image quality assessment, contrast enhancement and visual saliency detection.

PLACE PHOTO HERE

Guangtao Zhai (M'10) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently a Research Professor with the Institute of Image Communication and Information Processing.

From 2006 to 2007, he was a Student Intern with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen-Nuremberg, Germany. He received the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012. His research interests include multimedia signal processing and perceptual signal processing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2373812, IEEE Transactions on Multimedia

PLACE PHOTO HERE

Xiaokang Yang (M'00-SM'04) received the B. S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, in 2000. He is currently a Full Professor and Deputy Director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University.

From September 2000 to March 2002, he was a Research Fellow in Centre for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. He has published over 80 refereed papers, and has filed six patents. His current research interests include video processing and communication, media analysis and retrieval, perceptual visual processing, and pattern recognition. He actively participates in the International Standards such as MPEG-4, JVT, and MPEG-21. He received the Microsoft Young Professorship Award 2006, the Best Young Investigator Paper Award at IS&T/SPIE International Conference on Video Communication and Image Processing (VCIP2003), and awards from A-STAR and Tan Kah Kee foundations. He is a member of Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society. He was the Special Session Chair of Perceptual Visual Processing of IEEE ICME2006. He is the local co-chair of ChinaCom2007 and the technical program co-chair of IEEE SiPS2007.

PLACE PHOTO HERE

Wenjun Zhang (M'02-SM'10-F'11) received the B.S., M.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987 and 1989, respectively.

From 1990 to 1993, He worked as a post-doctoral fellow at Philips Kommunikation Industrie AG in Nuremberg, Germany, where he was actively involved in developing HD-MAC system. He joined the Faculty of Shanghai Jiao Tong University in 1993 and became a full professor in the Department of Electronic Engineering in 1995. As the national HDTV TEEG project leader, he successfully developed the first Chinese HDTV prototype system in 1998. He was one of the main contributors to the Chinese Digital Television Terrestrial Broadcasting Standard issued in 2006 and is leading team in designing the next generation of broadcast television system in China from 2011. He holds more than 40 patents and published more than 90 papers in international journals and conferences. Prof. Zhang's main research interests include digital video coding and transmission, multimedia semantic processing and intelligent video surveillance. He is a Chief Scientist of the Chinese National Engineering Research Centre of Digital Television (NERC-DTV), an industry/government consortium in DTV technology research and standardization and the Chair of Future of Broadcast Television Initiative (FOBTV) Technical Committee.