ORIGINAL PAPER

A dual-model approach to blind quality assessment of noisy images

GUANGTAO ZHAI[1], ANDRE KAUP[2], JIA WANG[1] AND XIAOKANG YANG[1]

We propose a new paradigm for no-reference image quality assessment (IQA) exploiting neurological and psychophysical properties of the human visual system (HVS)[1]. Physiological and psychological evidences exist that HVS has different behavioral patterns under low and high noise/artifact levels. In this paper, we propose a dual-model approach for blind IQA under near-threshold and suprathreshold noise conditions. The underlying assumption for the proposed dual-model approach is that for images with low-level near-threshold noise, HVS tries to gauge the strength of the noise, so image quality can be well approximated via measuring strength of the noise. On the other hand, for images with structures overwhelmed by high-level suprathreshold noise, perceptual quality assessment relies on a cognitive model and the HVS tries to recover meaningful contents from the noisy pixels using past experiences and prior knowledge encoded into an internal generative model of the brain. And image quality is therefore closely related to the agreement between the noisy observation and the internal generative model explainable part of the image. Specifically, under near-threshold noise condition, a noise level estimation algorithm using natural image statistics is used, while under suprathreshold condition, an active inference model based on the free-energy principle is adopted. The near- and suprathreshold models can be seamlessly integrated through a mathematical transformation between estimates from both models. The proposed dual-model algorithm has been tested on additive Gaussian noise contaminated images. Experimental results and comparative studies suggest that although being a no-reference approach, the proposed algorithm has prediction accuracy comparable with some of the best full-reference IQA methods. The dual-model perspective of IQA presented in this paper is expected to facilitate future research in the field of visual perceptual modeling.

I. INTRODUCTION

The psychovisual process of image quality assessment (IQA) is an interaction between the human visual system (HVS) and the visual stimuli [2]. For visual communication systems, visual stimuli is generally a combination of the meaningful image contents and the inevitably introduced noises or artifacts during acquisition, compression, and transmission. According to the psychophysical study of human sensory-threshold [3], for visual stimuli, the inherent artifacts can be classified into three categories [4, 5]: *Imperceptible noise* or *subthreshold noise* is considered invisible to the HVS and is therefore often out of the question for visual quality assessment. *Near-threshold noise* is observable

to the HVS, and it may obscure some fine image structures but does not obstacle the process of visual recognition. *Suprathreshold noise* is so intensive that it overwhelms some key image structures and features that are essential for visual perception. The threshold-related behavior of visual perception has a physiological explanation that the HVS has a set of sensory neurons being tuned to specific orientations and frequencies [6], and the excitation of some neurons may reduce the activity of others. This effect is also known as lateral inhibition [7, 8].

It has long been realized that near- and suprathreshold visual stimuli should be discriminated since the psychovisual processes involved are quite different [10–12]. The psychological studies of the "priming" effect of human vision confirm that viewing context can have substantial impacts on visual perception [13]. In the premise of threshold-related visual perception, it is easy to imagine that the process of visual quality assessment will also be affected by the visibility level of the noise or artifacts. For images containing near-threshold noise, the IQA process of the HVS is a task of detecting artifacts and distortions from the observation. The image quality is better if less artifact can be detected. On the other hand, for images contaminated by

[1]Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China. Phone: +81-21-34204511.
[2]Chair of Multimedia Communication and Signal Processing, Friedrich-Alexander-University, Erlangen, 91058, Germany

Corresponding author:
Guangtao Zhai
Email: zhaiguangtao@sjtu.edu.cn

[1]The paper is an extension of the conference version that has been presented at Picture Coding Symposium 2013 [1].
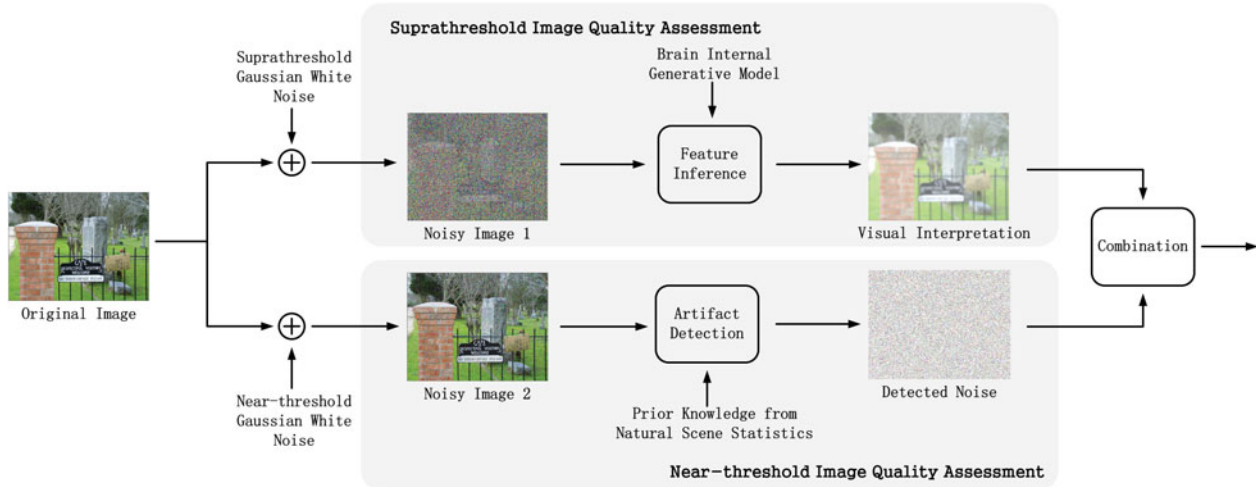
Fig. 1. IQA in the near- and suprathreshold conditions (noisy images taken from LIVE database [9]).

suprathreshold noise, subjective IQA becomes a work of inferring meaningful image structures using past experience and prior knowledge. The image quality is better if more significant structures can be discovered.

Although apparent gap exists between near- and suprathreshold IQA, currently there lacks the effective IQA method to account for the difference between near- and suprathreshold noise conditions. And most of existing quality metrics try to handle all noise and artifact levels irrespective of the physiological and psychological differences between near- and suprathreshold noise conditions mentioned above. For the problem of blind (no-reference) IQA, since no information about the original image is available, both the artifacts and meaningful image features have to be detected and discovered solely from the noisy observations. Therefore, the discrimination between near- and suprathreshold noise conditions becomes even more imperative. In this work, we propose a dual-model approach to quality assessment of noisy images under near- and suprathreshold noise conditions. For theoretical simplicity and operational amenability, our study in this work is focused on additive Gaussian white noise (AGWN), which is considered as one of the most simple type of distortions because the noisy signals are independent among pixels. Although being straightforward by itself, AGWN is often considered as a building component for other more complicated distortion types [14–16]. For example, impacts of quantization noises on perceptual quality can be modeled using Gaussian blurring plus AGWN as introduced in [14]. This is part of the reason why AGWN has been widely studied in the literature of IQA [9, 17, 18]. The study of AGWN can therefore serve as a pilot for the research of other more complicated types of distortions.

For near-threshold AGWN, it will be shown that high correlation exists between subjective quality score and strength (variance) of the noise. This coincides with our conjecture above that IQA is a noise detection/estimation problem under near-threshold conditions. So the near-threshold model of the proposed approach is based on a noise estimation algorithm using natural scene statistics. As aforementioned, for suprathreshold AGWN, IQA transfers to the problem of recovering original image structures from noisy observations. This echoes the general belief that vision is a process of inference, making assumptions and drawing conclusions based on experiences [19, 20], or using an internal generative model [21–23]. Therefore a Bayesian brain theory [21] inspired the inference method based on the free-energy principle [22, 23] will be used as the suprathreshold model of the proposed blind IQA scheme. A mathematical transformation will be used to calibrate the estimates from the near- and suprathreshold models and drive the dual-model distortion metric (DMDM) for blind quality assessment. Plot of the proposed dual-model approach to quality assessment of noisy images is illustrated in Fig. 1. We emphasize that the scope of this paper is beyond the mere introduction of a new blind image quality metric. The dual-model approach combining an artifact detection stage and an active inference stage presented in this paper should be applicable to the general problem of perceptual modeling and perceptual quality assessment and is expected to facilitate future research in the field.

The rest of the paper is organized as follows: Section II introduces the near- and suprathreshold models of the proposed dual-model approach in detail. Section III deals with the unification of the two models. Section IV provides the experimental results and comparative studies. Finally, Section V concludes the paper.

II. DMDM FOR BLIND QUALITY ASSESSMENT

System diagram of the proposed DMDM for blind quality assessment is illustrated in Fig. 2, wherein the near-threshold model based on noise estimation using natural image statistics and the suprathreshold model based on image feature inference using the free-energy principle
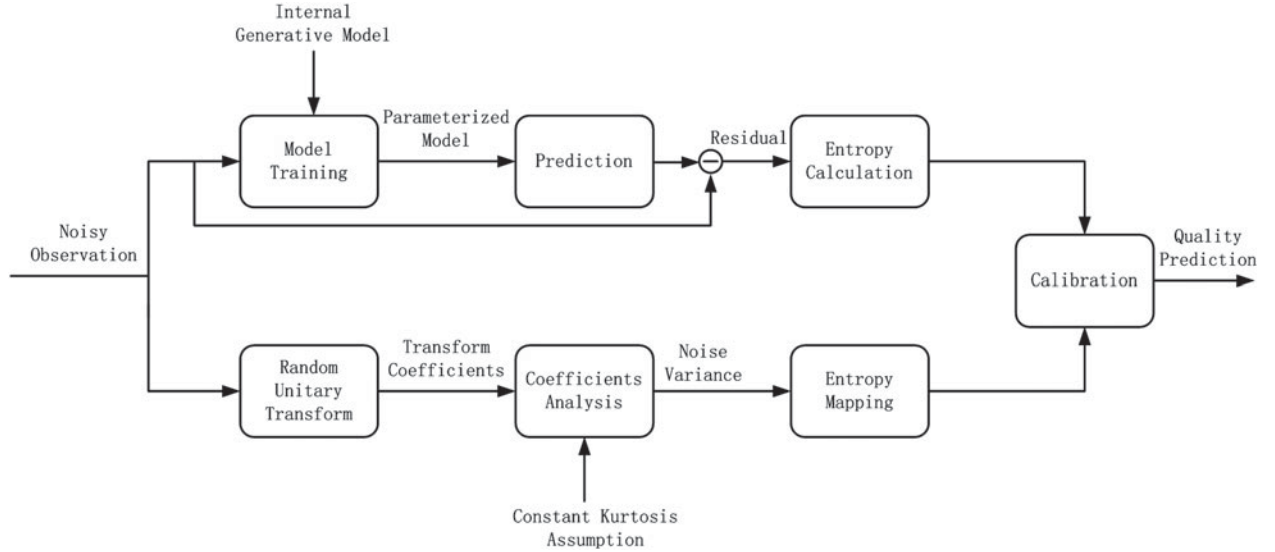
Fig. 2. Block diagram of the dual-model approach for blind quality assessment of noisy images.

will be introduced in detail in Sections II-A and II-B, respectively.

A) Near-threshold model: noise estimation using natural image statistics

Assuming an additive noise model $Y = X + N$ in the pixel domain, where X is the original image, N is the noise and Y is the noisy observation of X. If X and N are independent, then in the transform domain we also have

$$y = x + n, \qquad (1)$$

where x and y are the transformation of X and Y, and n is the noise in transform domain. Note that N and n will have the same variance, as long as the transform is unitary. It has been discovered that the high-order statistics (e.g. Kurtosis) of transform coefficients of natural images are stable across different frequency indexes [24–27], which can be used for noise variance estimation.

Kurtosis of x is defined as the fourth cumulant divided by the square of the second cumulant, which equals the normalized fourth central moments minus 3

$$K(x) = \frac{\kappa_4(x)}{\kappa_2^2(x)} = \frac{\mu_4(x)}{\sigma^4(x)} - 3. \qquad (2)$$

From the relationship between cumulant and the central moments $k_4(\cdot) = \mu_4(\cdot) - 3\sigma^4(\cdot)$ we have

$$\begin{aligned}
\mu_4(y) &= k_4(y) + 3\sigma^4(y) \\
&= k_4(x) + k_4(n) + 3\sigma^4(y) \\
&= \mu_4(x) - 3\sigma^4(x) + \mu_4(n) - 3\sigma^4(n) + 3\sigma^4(y) \\
&= K(x)\sigma^4(x) + K(n)\sigma^4(n) + 3\sigma^4(y), \qquad (3)
\end{aligned}$$

where we used the fact that x and n are independent. From (2) we have $\mu_4(\cdot) = (K(\cdot) + 3)\sigma^4(\cdot)$, and letting this into (3) we further have

$$K(y) = \frac{\sigma^4(x)}{\sigma^4(y)} K(x) + \frac{\sigma^4(n)}{\sigma^4(y)} K(n), \qquad (4)$$

which reveals that kurtosis of y are essentially weighted averages of that of x and n and the weights are determined by the ratios between the standard deviation of signals. Again, from the independence between x and y, we have $\sigma^2(y) = \sigma^2(x) + \sigma^2(n)$, letting which into (4), we arrive at

$$K(y) = \left(\frac{\sigma^2(y) - \sigma^2(n)}{\sigma^2(y)}\right)^2 K(x) + \left(\frac{\sigma^2(n)}{\sigma^2(y)}\right)^2 K(n), \qquad (5)$$

which enables us to quantify the relationships between the kurtosis of the original and the noisy transformed signals using the second moments (variances) of the noise n and the noisy observation y.

Recently, Zoran and Weiss [28] propose to estimate noise variance by applying the equation (5) upon multiple sub-bands of discrete cosine transform (DCT) coefficients of the noisy images. However, Zhai and Wu's work [29] suggests that the dominant horizontal and vertical edges in images may lead to a violation of the scale invariance assumption for DCT subbands [25, 26]. On the other hand, Huang [30, 31] first noticed that random transform coefficients of natural images have high kurtosis. In addition, the research in [29] validate that for random transform coefficients the kurtosis is not only high but also remains approximately a constant, which is a very useful property in estimating noise variance, and to preserve the noise energy in the transform domain, a unitary transform has to be used. For a $k \times k$ data matrix A, its two-dimensional (2D) random unitary transform (RUT) [29] can be computed as follows: we first compute the QR decomposition of a $k \times k$ random matrix C as $QR = C$. We then extract the diagonal elements of the upper triangular matrix $V = \text{diag}(R)$ and normalize it as $V = V/|V|$. A new upper triangular matrix

is formed as $R' = \text{diag}(\mathcal{V})$ and the RUT matrix is computed as $T = QR'$. The unitary of T is obvious since $T^T T = (QR')^T QR' = R'^T R' = I_k$, where we use the fact that R' is a diagonal matrix with normalized entries. The 2D RUT for A can then be computed as $B = TAT^T$.

By employing (5) on the subbands of the RUT coefficients, we can estimate the noise variance by solving the following constrained nonlinear programming problems:

$$\widehat{K}(x), \widehat{K}(n), \widehat{\sigma}^2(n) = \underset{K(x),K(n),\sigma^2(n)}{\arg\min} \sum_{i \in \mathcal{I}} \quad (6)$$

$$\left\| \widehat{K}(y_i) - \left(\frac{\widehat{\sigma}^2(y_i) - \sigma^2(n)}{\widehat{\sigma}^2(y_i)} \right)^2 K(x) + \left(\frac{\sigma^2(n)}{\widehat{\sigma}^2(y_i)} \right)^2 K(n) \right\|_1$$

subject to $K(x), K(n) \geq -2$,

where $\mathcal{I}$ denotes the set of frequency indexes of RUT and the constraints are from the definition of Kurtosis. We emphasize that the analysis in this section is only based on the independent additive noise model as given in (1), and there is no underlying assumption about the distributions of transform coefficients x and n.

B) Suprathreshold model: structure inference with generative model

Recently, Friston [22, 23] proposed the free-energy principle to explain and unify several brain theories in biological and physical sciences about human action, perception, and learning. Similar to the Bayesian brain hypothesis, a basic premise of the free energy-based brain theory is that the cognitive process is governed by an internal generative model. Using the generative model, human brain renders predictions of those encountered scenes in a constructive manner, and the constructive model is essentially a probabilistic model that decomposes into a likelihood term and a prior term. Visual perception is then the process of inverting this likelihood term so as to infer the posterior possibilities of the given scene. Not surprisingly, since the internal generative model cannot be universal (implying the limitation of one's knowledge and experience), there always exists a gap between the encountered scene and brain's prediction. As mentioned in the Introduction section, for images with suprathreshold noise, features, and cues that are essential for visual perception are overwhelmed, so visual perception reasonably follows the manner of inference as outline by the free-energy principle, and under those conditions, the gap between the external input and its generative-model-explainable part is intimately related to the quality of perceptions and therefore determines the perceptual quality of the image.

For operational amenability, we assume that the internal generative model $\mathcal{G}$ for visual perception is parametric, which explains perceived scenes by adjusting the vector $\boldsymbol{\theta}$ of parameters. Given a visual stimuli, or an image I, its "surprise" (measured by entropy) can be computed by integrating the joint distribution $P(I, \boldsymbol{\theta}|\mathcal{G})$ over the space of model parameters $\boldsymbol{\theta}$

$$-\log P(I|\mathcal{G}) = -\log \int P(I, \boldsymbol{\theta}|\mathcal{G}) d\boldsymbol{\theta}. \quad (7)$$

We further introduce an auxiliary term $Q(\boldsymbol{\theta}|I)$ into both the denominator and numerator in (7) and have

$$-\log P(I|\mathcal{G}) = -\log \int Q(\boldsymbol{\theta}|I) \frac{P(I, \boldsymbol{\theta}|\mathcal{G})}{Q(\boldsymbol{\theta}|I)} d\boldsymbol{\theta}. \quad (8)$$

Here $Q(\boldsymbol{\theta}|I)$ is a posterior distribution of the model parameters given the image. It can be thought of as an approximate posterior to the true posterior of the model parameters $P(\boldsymbol{\theta}|I, \mathcal{G})$ that can be calculated by the brain. The brain minimizes the discrepancy between the approximate posterior $Q(\boldsymbol{\theta}|I, \mathcal{G})$ and the true posterior $P(\boldsymbol{\theta}|I, \mathcal{G})$ when perceiving I, or adjusting the parameters $\boldsymbol{\theta}$ of $Q(\boldsymbol{\theta}|I, \mathcal{G})$ to best explain I. Therefore, it is also called the *recognition* density in Bayesian brain theory [21]. Note that in (8), the negative "surprise" $\log P(I|\mathcal{G})$ is also known as the log-evidence of the image data I given the model. Clearly, the minimization of surprise equals the maximization of model evidence. In our analysis below, the dependency on the generative model $\mathcal{G}$ is dropped for simplicity. Using Jensen's inequality, from (8) we have

$$-\log P(I) \leq -\int Q(\boldsymbol{\theta}|I) \log \frac{P(I, \boldsymbol{\theta})}{Q(\boldsymbol{\theta}|I)} d\boldsymbol{\theta}, \quad (9)$$

and the right-hand side which is defined as the free energy

$$F(\boldsymbol{\theta}) = -\int Q(\boldsymbol{\theta}|I) \log \frac{P(I, \boldsymbol{\theta})}{Q(\boldsymbol{\theta}|I)} d\boldsymbol{\theta}. \quad (10)$$

By letting $P(I, \boldsymbol{\theta}) = P(\boldsymbol{\theta}|I)P(I)$ into (10) we have

$$\begin{aligned} F(\boldsymbol{\theta}) &= \int Q(\boldsymbol{\theta}|I) \log \frac{Q(\boldsymbol{\theta}|I)}{P(\boldsymbol{\theta}|I)P(I)} d\boldsymbol{\theta} \\ &= -\log P(I) + \int Q(\boldsymbol{\theta}|I) \log \frac{Q(\boldsymbol{\theta}|I)}{P(\boldsymbol{\theta}|I)} d\boldsymbol{\theta} \\ &= -\log P(I) + KL\left(Q(\boldsymbol{\theta}|I)\|P(\boldsymbol{\theta}|I)\right). \quad (11) \end{aligned}$$

According to Gibbs' inequality, the Kullback–Leibler divergence $KL(\cdot\|\cdot)$ between the recognition (approximate) posterior and true posterior parameter distribution in (11) is non-negative, i.e. $KL\left(Q(\boldsymbol{\theta}|I)\|P(\boldsymbol{\theta}|I)\right) \geq 0$, with equality if and only if $Q(\boldsymbol{\theta}|I) = P(\boldsymbol{\theta}|I)$. Here the free energy $F(\boldsymbol{\theta})$ defines a strict upper bound of the "surprise" or negative model evidence. As indicated in (11), for fixed image data I, the free energy is suppressed by minimizing the divergence term. In other words, the brain tries to lower the divergence $KL\left(Q(\boldsymbol{\theta}|I)\|P(\boldsymbol{\theta}|I)\right)$ between the approximate recognition density of model parameters and its true posterior density when perceiving a given scene. And following the analysis we made at the beginning of this subsection, the computed free-energy term can be used directly as a no reference measure of image quality. For a clean image with little noise contamination, an ideal generative model is expected to recognize the essential image features well and the free

energy will be low, indicating a higher image quality. While for noisy image with perceptually important features being besmirched, the free energy induced the generative model will be high, indicating a lower perceptual quality.

In practice, the internal generative model can be further simplified as a linear model such as the noncausal autoregressive (AR) model and the linear AR model can be locally optimized from the observation. Some implementation details will be briefly introduced in the next section and more theoretical treatments regarding the free-energy principle and the approximation using an AR model can be found in [32]. The calibration and combination of the near- and suprathreshold models toward a final dual-model quality metric will be discussed in the next section.

III. MODEL CALIBRATION AND INTEGRATION

The near- and suprathreshold quality metrics based on natural image statistics and the free-energy principle have been introduced in Sections II-A and II-B, respectively. For near-threshold quality assessment, the noise estimation is used as a distortion measure, whereas for suprathreshold quality assessment, the free energy of the noisy observation is used as a distortion measure. Obviously, the noise strength measure by variance and free-energy measure by entropy are of different dimension so the two-quality model cannot be combined directly, and this section discusses the unification and calibration of the two distortion measures and defines the so-called DMDM, for blind quality assessment.

Assuming an additive noise model $Y = X + N$, where $N \sim \mathcal{N}(0, \sigma)$ is the Gaussian noise term. For $N \sim \psi(n) = (1/\sqrt{2\pi\sigma^2}) \times e^{-n^2/2\sigma^2}$, its differential entropy can be computed as

$$
\begin{aligned}
H(\psi) &= -\int \psi(n) \ln \psi(n) dn \\
&= -\int \psi(n) \left(-\frac{n^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2} \right) dn \\
&= \frac{1}{2} + \frac{1}{2} \ln 2\pi\sigma^2 = \frac{1}{2} \ln 2\pi e\sigma^2.
\end{aligned}
\tag{12}
$$

Note that the entropy in (12) is in *nats*, and by changing the logarithm base, we get the entropy in *bits* as

$$
H(\psi) = \frac{1}{2} \log 2\pi e\sigma^2.
\tag{13}
$$

Equation (13) provides a means to convert the estimate from the near-threshold noise estimation model to the entropy-based quality measure as given by the near-threshold quality model

$$
H^{near}(Y) = \frac{1}{2} \log 2\pi e\hat{\sigma}^2
\tag{14}
$$

with $\hat{\sigma}$ being the variance estimate from (6) and Y is the noisy observation.

One should excise caution when combining the transformed entropy measure from near-threshold model and the entropy measure from the suprathreshold model. This is because the free-energy value, when being approximated using a locally learnt linear AR model as introduced in Section II-B, contains contributions from the meaningful image structures. Ideally, for a perfect internal generative model, the approximation error of the free-energy model contains only the AGWN. However, in practice, since a linear AR model is used as an delegate of the idealistic internal model and is optimized with the noisy samples, the estimation error inevitably carries some of the image structures and has to be calibrated before model integration.

The linear estimation using a linear AR model from noisy observation Y can be written as

$$
\hat{Y} = AY_0,
\tag{15}
$$

where Y_0 is a vector consists of samples from the neighborhood and A is the coefficient matrix. To minimize the estimation error $E = Y - \hat{Y}$, from the orthogonal principle, the error must be perpendicular to the observation Y_0. Therefore we have

$$
E[(Y - \hat{Y})Y_0^T] = E[(Y - AY_0)Y_0^T] = 0.
\tag{16}
$$

From (16) we can get $\Sigma_{YY_0^T} = A\Sigma_{Y_0}$, so the optimal linear coefficient matrix is

$$
A = \Sigma_{YY_0^T} \Sigma_{Y_0}^{-1}.
\tag{17}
$$

The covariance matrix of error vector E can be written as

$$
\begin{aligned}
\Sigma_E &= E[(Y - \hat{Y})(Y - \hat{Y})^T] \\
&= E[(Y - AY_0)(Y - AY_0)^T] \\
&= E[(Y - AY_0)(Y^T - Y_0^T A^T)] \\
&= \Sigma_Y - \Sigma_{YY_0^T} A^T - A\Sigma_{Y_0 Y^T} + A\Sigma_{Y_0} A^T.
\end{aligned}
\tag{18}
$$

From (17) we have

$$
A\Sigma_{Y_0} A^T = \Sigma_{YY_0^T} \Sigma_{Y_0}^{-1} \Sigma_{Y_0} A^T = \Sigma_{YY_0^T} A^T.
\tag{19}
$$

And letting (19) into (18) we arrive at

$$
\Sigma_E = \Sigma_Y - A\Sigma_{Y_0 Y^T} = \Sigma_Y - \Sigma_{YY_0^T} \Sigma_{Y_0}^{-1} \Sigma_{Y_0 Y^T}.
\tag{20}
$$

On the other hand, from the independence between Y and N, we have $\Sigma_{YY_0^T} = \Sigma_{Y_0 Y^T} = \Sigma_X$ and $\Sigma_Y = \Sigma_X + \sigma_n^2 I$. Therefore (20) reduces to

$$
\Sigma_E = (\Sigma_X + \sigma_n^2 I) - \Sigma_X(\Sigma_X + \sigma_n^2 I)^{-1} \Sigma_X.
\tag{21}
$$

Using the equation of the inverse of sum of matrices[2] (21) can be further written as

$$
\Sigma_E = \sigma_n^2 I + (\Sigma_X^{-1} + (\sigma_n^2 I)^{-1})^{-1},
\tag{22}
$$

which indicates that the covariance of the residual signal from the linear prediction is larger than that of the noise

---

[2] $(B + C)^{-1} = B^{-1} - B^{-1}(B^{-1} + C^{-1})^{-1} B^{-1}$, and letting $B^{-1} = \Sigma_X$ and $C^{-1} = \sigma_n^2 I$.
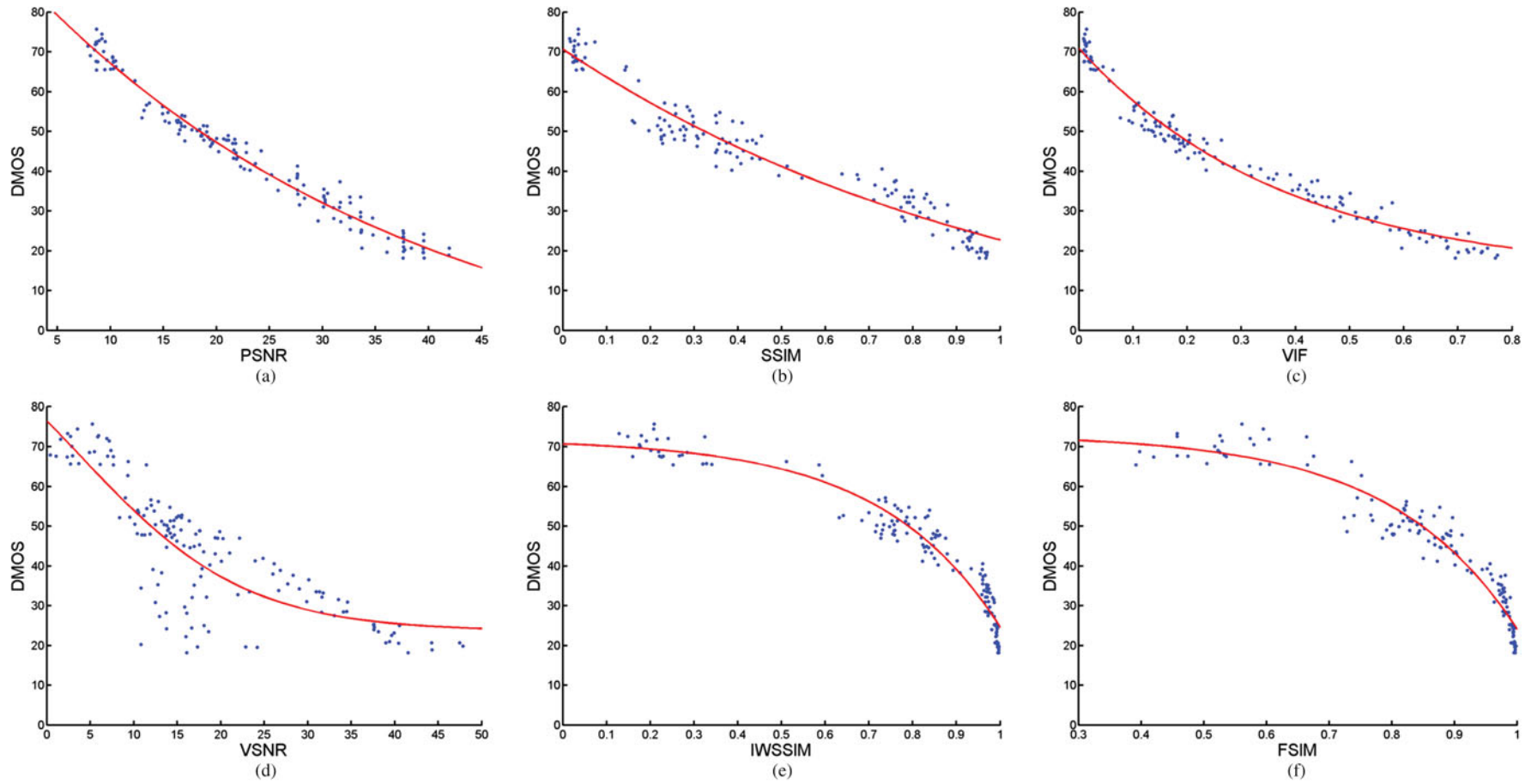
Fig. 3. Scatter plots of DMOS/MOS versus full-reference quality metrics on test databases LIVE [9]

signal alone, and under the simplest condition of pixel-wise estimation, (22) becomes

$$\sigma_e^2 = \sigma_n^2 + ((\sigma_x^2)^{-1} + (\sigma_n^2)^{-1})^{-1}, \qquad (23)$$

where σ_e^2 and σ_x^2 are respectively the variance for the error term and the image signal. Equation (23) indicates that variance of the estimate error of the free-energy model, as computed from a linear AR model, is larger than variance of the inherent AGWN in the observation. Therefore, care should be taken to properly shrink the free-energy term before combining with the estimate from the near-threshold model.

However, a precise correction using (23) requires the variance of the original signal x, which is out of question for blind quality assessment algorithm. In fact, the difference term $((\sigma_x^2)^{-1} + (\sigma_n^2)^{-1})^{-1}$ in (23) defines a harmonic mean between σ_x^2 and σ_n^2. On the other hand, σ_x^2 can change drastically from one location to another due to different local features of the image signal, making the relationship between σ_e^2 and σ_x^2 even more complicated. Fortunately, in practice, on average for the whole image, it is possible to assume a constant shrinkage

$$H^{supra}(Y) = \xi \cdot F(\theta), \qquad (24)$$

where $F(\theta)$ is the free-energy value computed with a linear AR model using the noisy observation Y. We empirically find $\xi = 0.89$ which gives a reasonably well-calibration result.

To combine the estimates from the near-threshold model in (14) and suprathreshold model in (24) and to derive the final DMDM value H^{DM}, we can simply use a binary integration

$$H^{DM}(Y) = \begin{cases} H^{near}(Y) & \text{if } H^{near}(Y) \leq \zeta, \\ H^{supra}(Y) & \text{otherwise,} \end{cases} \qquad (25)$$

where ζ is a predefined switching threshold and we empirically set $\zeta = 6.2$ in this research. Note that besides the binary combination scheme used above, other linear/nonlinear combinations are also possible.

Matlab code of the proposed DMDM algorithm can be found at multimedia.sjtu.edu.cn.

IV. EXPERIMENTAL RESULTS AND COMPARATIVE STUDY

The test images used in this work are taken from the widely used LIVE database [9]. Note some of other emerging test databases, such as TID2008 [18] and CSIQ [17] also have subsets of AGWN contaminated images, but the noise levels (variances) used in both TID2008 and CSIQ databases are very limited[3], making themselves not suitable for a fair and trustable test of image quality metric based on noise level estimation.

[3]Only four and five noise variance levels were used respectively in the sub-databases TID2008 and CSIQ.

Table 1. Comparisons of full-reference image quality metrics on AGWN images in LIVE database.

	PSNR	SSIM [33]	VIF [15]	VSNR [34]	IWSSIM [16]	FSIM [35]
CC	0.986	0.969	0.981	0.841	0.969	0.964
SROCC	0.985	0.963	0.985	0.817	0.966	0.965
MAE	2.164	3.256	3.115	6.643	3.346	3.401
RMSE	2.680	3.916	2.536	8.620	3.943	4.196

We first report the performances of some classic and contemporary general-purpose full-reference quality metrics on AGWN contaminated images. The full-reference quality metrics tested in this paper include the ubiquitous peak signal-to-noise ratio (PSNR), state-of-the-art SSIM [33], VIF [15], and VSNR [34], and two of the newest quality metrics IWSSIM [16] and FSIM [35]. Figure 3 shows the scatter plots of scores of the quality metrics versus difference mean opinion score (DMOS) of the test images from the LIVE database. And solid line in those plots is the nonlinear regression curve using the 4 parameter logistic function suggested by the video quality expert group (VQEG) [39]

$$M(s) = \frac{\beta_1 - \beta_2}{1 + \exp(-(s - \beta_3)/\beta_4)} + \beta_2, \qquad (26)$$

where s is the input score, while $M(s)$ is the mapped score and $\beta_1, \ldots, \beta_4$ are free parameters to be determined during the curve-fitting process. The numerical performance analysis suggested by VQEG [39] is provided in Table 1, including four commonly used testing criteria: the correlation coefficient (CC), Spearman rank-order correlation coefficient (SROCC), mean absolute error (MAE) and root-mean-squared error (RMSE). Note that higher correlations and lower mean errors in the table indicate better performance of the quality metric. It can be concluded from Fig. 3 is that the five general-purpose full-reference quality metrics PSNR, SSIM, VIF, IWSSIM, and FSIM all perform fairly well for the IQA of AGWN contaminated images. And the numerical performances in Table 1 indicate that the linear/nonlinear correlations between the IQA scores and the DMOS are as good as $0.96 \sim 0.98$, much higher than the metrics' performances on other subset of the LIVE database[4]. However, those full-reference quality metrics all require the original pristine images as inputs, which greatly restrains the application scenario, and we will focus on how to design a blind/no-reference quality metric in this work.

The scatter plots of DMOS versus some no-reference image quality metrics on the LIVE database are illustrated in Fig. 4. The results of the scale-invariant noise estimation (SINE) i.e. near-threshold model and the no-reference free energy-based distortion metric (NFEDM), i.e. the suprathreshold model are shown respectively in Figs. 4(a) and 4(b). Not surprisingly, the near-threshold model and the suprathreshold model have dissimilar yet

[4]The performances of PSNR, SSIM, and VIF on other subsets of the LIVE database can be found in [32].
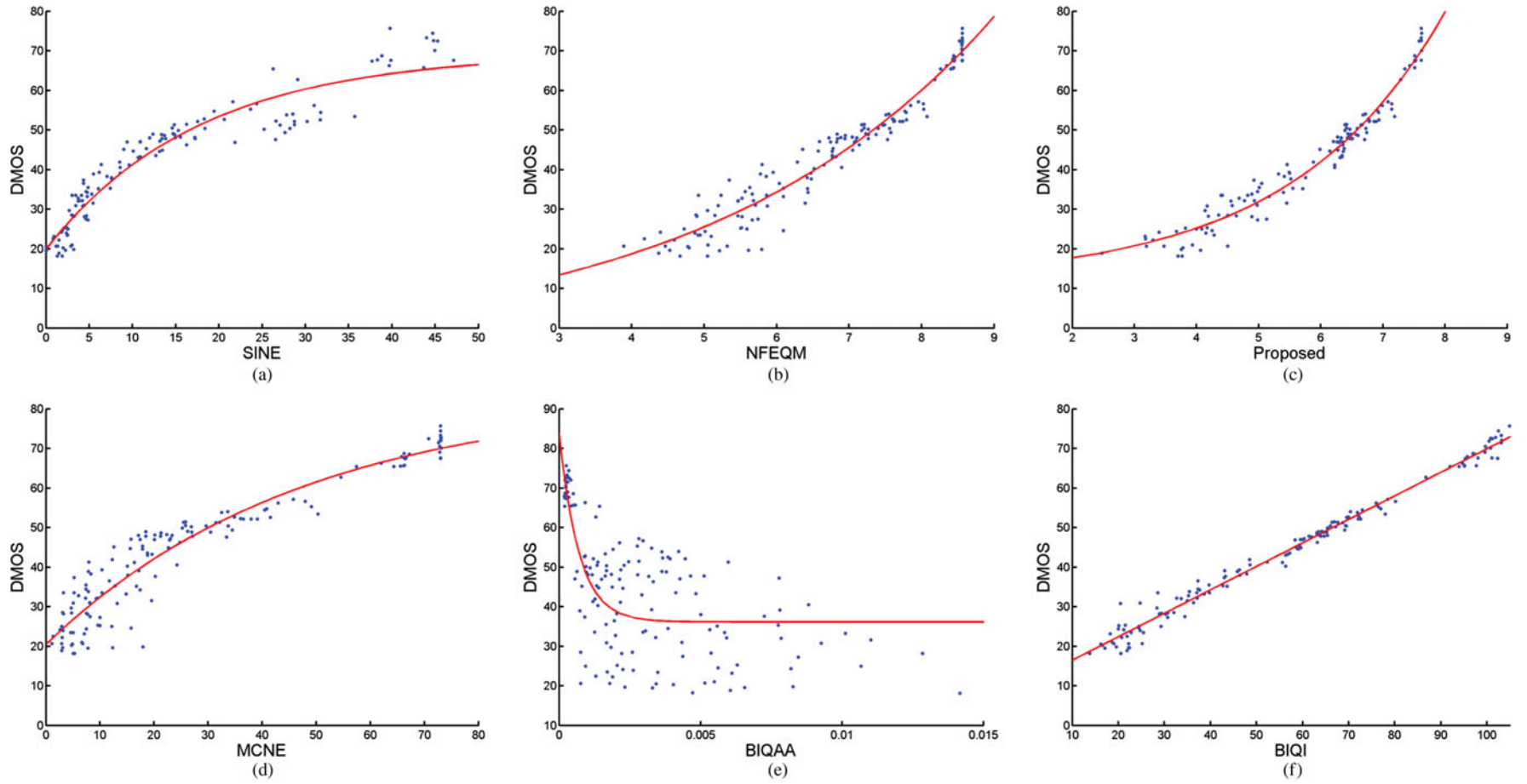
Fig. 4. Scatter plots of DMOS/MOS versus no-reference quality metrics on test databases LIVE [9]

Table 2. Comparisons of no-reference image quality metrics on AGWN images in LIVE database.

	DMDM	SINE [28]	NFEDM [32]	MCNE [36]	BIQAA [37]	BIQI [38]
CC	0.981	0.944	0.971	0.947	0.711	0.992
SROCC	0.978	0.957	0.968	0.942	0.615	0.990
MAE	2.867	5.26	3.823	3.900	9.297	1.394
RMSE	2.284	3.822	2.960	5.128	11.21	1.911

complementary performances across different noise levels. It is noticed in Fig. 4(a) that for images of relatively high quality (e.g. DMOS$\leq$ 45), SINE seems to correlate well the subjective scores, confirming the assumption that for images with near-threshold noise, perceptual quality assessment is process of gauging noise levels. Therefore, image quality can be well approximated through noise estimation. Note that at very low noise levels (e.g. $\sigma \leq 2.5$) the performance of SINE drops because of the possible interference between image signal and noises. Meanwhile, it is found in Fig. 4(b) that NFEDM performs better for images of relatively low quality (e.g. DMOS$>$ 45), indicating the plausibility of a free-energy-based inference model for blind quality assessment of images corrupted by suprathreshold noises[5]. The above observations not only demonstrate the existence of the dissimilar mechanisms of perceptual quality assessment for near- and suprathreshold noise conditions, but also justify the appropriateness of the proposed dual-model approach to IQA combining a noise estimation model and an image inference model. Inheriting merits from both near- and suprathreshold model, the proposed DMDM obviously has a more consistent performance over all noise levels, outperforming both near- and suprathreshold models, as illustrated in Fig. 4(c).

For a thorough comparison, some blind quality metrics, including the mean curvature-based noise estimation (MCNE) [36], blind IQA through anisotropy (BIQAA) [37], and the Blind Image Quality Index (BIQI) [38] are also included in our comparison in Figs. 4(d)–4(f). Note that BIQI is a training-based approach with parameters optimized on the LIVE database itself [38], so its performance is understandably far better than other metrics. And MCNE and BIQAA are clearly outperformed by the proposed DMDM on a wide quality range. A comparison between proposed DMDM and those full-reference metrics shown in Fig. 3 suggests that though being a no-reference method, DMDM has performance almost equivalent to some of the best full-reference metrics, such as SSIM and VIF.

Numerical performances of the no-reference quality assessment algorithms, using metrics suggested by the VQEG are listed in Table 2. The linear and nonlinear correlation coefficients between DMDM and DMOS are about 0.98, indicating a prediction accuracy comparable with those of PSNR and VIF and superior than SSIM, VSNR,

IWSSIM, and FSIM, as listed in Table 1. Moreover, the MAE and RMSE performances of DMDM are even better than those of VIF and most of other full-reference metrics. Again, BIQI results included in Table 2 is just for reference, since BIQI was trained on the LIVE database, and its direct comparison with other metrics is unfair.

V. CONCLUSION

Despite all the literatures and discussions of IQA, a fundamental question appear not to have been posed clearly, let alone answered: is there any difference between quality assessment under different noise/artifact levels. In this paper, we demonstrate that there does exist dissimilar psychovisual mechanism of the HVS under near- and suprathreshold noise levels. Taking the artifact of AGWN as an example, we approximate the behavior of HVS using a dual-model approach: a noise estimate model based on natural image statistic for near-threshold noise and an image inference model based on the free-energy principle for suprathreshold noise. We show that the proposed dual-model-based distortion metric outperforms some of state-of-the-art blind and full-reference quality metrics. The proposed dual-model approach to IQA sheds light on the necessity, possibility and effectuality of incorporating cognitive factors into the problem of IQA, and the dual-model framework is also extendable to other more complicated distortion types and is expected to improve the performances of existing image quality metrics that were designed to be universal, handling all noise/artifact levels.

ACKNOWLEDGEMENTS

REFERENCES

[1] Zhai, G.; Kaup, A.; Wang, J.; Yang, X.: A dual-model approach to blind quality assessment of noisy images, in *Picture Coding Symposium (PCS), 2013*, December 2013, 29–32.

[2] Wang, Z.; Bovik, A.C.: Modern Image Quality Assessment. Synthesis Lectures on Image, Video & Multimedia Processing, *Morgan & Claypool Publishers*, San Rafael, CA, 2006.

[3] Levine, M.: Fundamentals of Sensation and Perception, 3rd ed., *Oxford University Press*, London, 2000.

[4] Hay, G.A.; Chesters, M.S.: Limiting factors in threshold and suprathreshold vision. Nature, **228** (1970), 1216–1218.

[5] Lu, Z.; Lin, W.; Yang, X.; Ong, E.; Yao, S.: Modeling visual attentions modulatory aftereffects on visual sensitivity and quality evaluation. IEEE Trans. Image Process., **14** (11) (2005), 1928–1942.

[6] Valois, R.L.D.; Valois, K.K.D.: Spatial Vision, *Oxford University Press*, New York, 1990.

[7] Bekesy, G.V.: Sensory Inhibition, *Princeton University Press*, Princeton, NJ, 1967.

---

[5]A saturation effect of NFEQM is found under very large noise levels, this is because 8-bit images are used in this work and the approximated free-energy level will be bounded by a threshold lower than 9.

[8] Orchard, G.A.; Phillips, W.A.: Neural Computation: A Beginner's Guide. *Taylor and Francis*, New York, NY, 1991.

[9] Sheikh, H.; Wang, Z.; Cormack, L.; Bovik, A.: Live image quality assessment database release 2 [Online]. Available at: http://live.ece.utexas.edu/research/quality.

[10] Cornsweet, T.: Visual Perception. *Harcourt Publishing*, London, 1970.

[11] Sachs, M.B.; Nachmias, J.; Robson, J.G.: Spatial-frequency channels in human vision. J. Opt. Soc. Am. A, **61** (1971), 1176–1186.

[12] Sagi, D.; Hochstein, S.: Discriminability of suprathreshold compound spatial frequency gratings. Vis. Res., **23** (1983), 1595–1606.

[13] Kolb, B.; Whishaw, I.Q.: Fundamentals of Human Neuropsychology. *Worth Publishers Incorporated*, New York, NY, 2008.

[14] Sheikh, H.; Bovik, A.; de Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans. Image Process., **14** (12) (2005), 2117–2128.

[15] Sheikh, H.R.; Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process., **15** (2) (2006), 430–444.

[16] Wang, Z.; Li, Q.: Information content weighting for perceptual image quality assessment. IEEE Trans. Image Process., **20** (5) (2011), 1185–1198.

[17] Larson, E.C.; Chandler, D.M.: Most apparent distortion: full-reference image quality assessment and the role of strategy. J. Electron. Imaging, **19** (1) (2010), 011006.

[18] Ponomarenko, N.; Lukin, V.; Zelensky, A.; Egiazarian, K.; Carli, M.; Battisti, F.: TID2008 – a database for evaluation of full-reference visual quality assessment metrics. Adv. Mod. Radioelectron., **1** (2009), 30–45.

[19] von Helmholtz, H.: Treatise on Physiological Optics, 3rd ed. *Voss*, Hamburg, 1909.

[20] Sternberg, R.: Cognitive Psychology, 3rd ed. *Thomson Wadsworth*, Belmont, CA, 2003.

[21] Knill, D.C.; Pouget, A.: The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci., **27** (2004), 712–719.

[22] Friston, K.; Kilner, J.; Harrison, L.: A free energy principle for the brain. J. Physiol. Paris, **100** (2006), 70–87.

[23] Friston, K.: The free-energy principle: a unified brain theory? Nat. Rev. Neurosci., **11** (2010), 127–138.

[24] Simoncelli, E.P.; Olshausen, B.A.: Natural image statistics and neural representation. Annu. Rev. Neurosci., **24** (2001), 1193–1216.

[25] Srivastava, A.; Lee, A.B.; Simoncelli, E.P.; Zhu, S.-C.: On advances in statistical modeling of natural images. J. Math. Imaging Vis., **18** (1) (2003), 17–33.

[26] Mumford, D.B.: Empirical statistics and stochastic models for visual signals, in S. Haykin, J.C. Principe, T.J. Sejnowski & J. McWhirter (Eds.) New Directions in Statistical Signal Processing: from Systems to Brain. *Massachusetts Institute of Technology Press*, Cambridge, MA, 2006, pp. 1–34.

[27] Geisler, W.S.: Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol., **59** (2007), 167–172.

[28] Zoran, D.; and Weiss, Y.: Scale invariance and noise in natural images, in *Int. Conf. on Computer Vision*, 2009, 2209–2216.

[29] Zhai, G.; and Wu, X.: Noise estimation using statistics of natural images, *IEEE 2011 Int. Conf. on Image Processing*, 2010, 1857–1860.

[30] Huang, J.; Mumford, D.: Statistics of natural images and models, in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, 1999, 541–547.

[31] Huang, J.: Statistics of Natural Images and Models. Ph.D. Thesis, Division of Applied Mathematics, Brown University, 2000.

[32] Zhai, G.; Wu, X.; Yang, X.; Lin, W.; Zhang, W.: A psychovisual quality metric in free energy principle. IEEE Trans. Image Process., **21** (1) (2012), 41–52.

[33] Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process., **13** (4) (2004), 600–612.

[34] Chandler, D.M.; Hemami, S.S.: VSNR: a wavelet-based visual signal-to-noise ratio for natural images. IEEE Trans. Image Process., **16** (9) (2007), 2284–2298.

[35] Zhang, L.; Zhang, L.; Mou, X.; Zhang, D.: FSIM: a feature similarity index for image quality assessment. IEEE Trans. Image Process., **20** (8) (2011), 2378–2386.

[36] Li, X.: Blind image quality assessment, in *Int. Conf. on Image Processing*, vol. 1, 2002, 449–452.

[37] Gabarda, S.; Cristóbal, G.: Blind image quality assessment through anisotropy. J. Opt. Soc. Am. A, **24** (12) (2007), 42–51.

[38] Moorthy, A.; Bovik, A.: A two-step framework for constructing blind image quality indices. IEEE Signal Process. Lett., **17** (5) (2010), 513–516.

[39] VQEG: VQEG Final Report of FR-TV Phase II Validation Test, 2003 [Online]. Available at: www.vqeg.org

Guangtao Zhai received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China in 2009. He was a Humboldt research fellow at the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander-University, Germany. He is now a Research Professor at Shanghai Jiao Tong University, China. His research interests include multimedia signal processing and perceptual signal processing.

Andre Kaup is a Professor and Head at the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander-University, Germany. His research interests include image and video signal processing, coding of images, video, and multiview data, as well as multimedia communication.

Jia Wang is an Associate Professor at Shanghai Jiao Tong University, China. His research interests include image processing and information theory.

Xiaokang Yang is a Professor at Shanghai Jiao Tong University, China. His research interests include visual signal processing, machine learning, and pattern recognition.