Reduced-Reference Image Quality Assessment in Free-Energy Principle and Sparse Representation

Yutao Liu, Guangtao Zhai, *Member, IEEE*, Ke Gu, Xianming Liu, Debin Zhao, and Wen Gao, *Fellow, IEEE*

***Abstract*—The free-energy principle in recent studies of brain theory and neuroscience models the perception and understanding of the outside scene as an active inference process, in which the brain tries to account for the visual scene with an internal generative model. Specifically, with the internal generative model, the brain yields corresponding predictions for its encountered visual scenes. Then, the discrepancy between the visual input and its brain prediction should be closely related to the quality of perceptions. On the other hand, sparse representation has been evidenced to resemble the strategy of the primary visual cortex in the brain for representing natural images. With the strong neurobiological support for sparse representation, in this paper, we approximate the internal generative model with sparse representation and propose an image quality metric accordingly, which is named FSI (free-energy principle and sparse representation-based index for image quality assessment). In FSI, the reference and distorted images are, respectively, predicted by the sparse representation at first. Then, the difference between the entropies of the prediction discrepancies is defined to measure the image quality. Experimental results on four large-scale image databases confirm the effectiveness of the FSI and its superiority over representative image quality assessment methods. The FSI belongs to reduced-reference methods, and it only needs a single number from the reference image for quality estimation.**

***Index Terms*—Free-energy principle, image quality assessment (IQA), reduced-reference (RR), sparse representation, visual saliency.**

Manuscript received September 15, 2016; revised December 26, 2016; accepted July 6, 2017. Date of publication July 21, 2017; date of current version January 17, 2018. This work was supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2015CB351804 and in part by the National Science Foundation of China under Grant 61672193. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ivan Bajic. *(Corresponding author: Xianming Liu.)*

Y. Liu, X. Liu, and D. Zhao are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: yt.liu@hit.edu.cn; xmliu.hit@gmail.com; dbzhao@hit.edu.cn).

G. Zhai is with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@gmail.com).

K. Gu is with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke@bjut.edu.cn).

W. Gao is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and also with the National Engineering Laboratory for Video Technology and Key Laboratory of Machine Perception, School of Electrical Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2017.2729020

I. Introduction

IMAGE quality assessment (IQA) plays a mass of roles in image processing applications. First, IQA can be used to evaluate or monitor the image quality, which is the original intention of IQA. Besides, it can also be used to control the execution of the image processing systems. The third common application of IQA is that it often serves as the performance measure for various image processing algorithms, such as image compression [1], restoration [2] and enhancement [3], [4] algorithms etc. Hence, IQA becomes an important topic in both scientific research and practical applications.

Since most digital images are ultimately consumed by humans, the most reliable way for evaluating the image quality is the subjective assessment by humans, which refers to asking viewers to directly rate the image quality according to their perception of the images. However, this quality assessment manner is always expensive, cumbersome and also improper for the real-time image processing systems. To tackle this, researchers devote to developing objective IQA methods which can automatically evaluate the image quality. In this paper, we focus our attention on objective IQA methods.

Last decades have seen a lot of sophisticated objective IQA approaches, which in general can be divided into three categories according to the access to the reference or original image, which are full-reference (FR), reduced-reference (RR) and no-reference (NR) methods respectively. Based on the definition, FR methods assume that the reference image is completely available when assessing the image quality. The traditional FR method, mean-squared error (MSE) or its deduced peak signal-to-noise ratio (PSNR) is most widely used owing to its simplicity and clear physical meaning. However, it is also found to correlate poorly with subjective ratings under some conditions [5]. Toward this end, Wang *et al*. proposed a structural similarity index (SSIM) [5] to measure the image quality, which can be regarded as a milestone in IQA studies. Specifically, the authors hypothesized that the human visual system (HVS) is highly adapted to extract structural information from the visual scene. Therefore, measuring the structure distortion between the reference image and distorted image can well infer the image quality. Then a series of FR methods were proposed successively. Sheikh *et al*. proposed an information fidelity criterion (IFC) by introducing information theory into IQA through quantifying the shared information in reference and distorted images [6]. Afterwards IFC was improved to visual information fidelity (VIF) in [7]. Visual signal-to-noise ratio (VSNR) was presented based on

near-threshold and suprathreshold properties of the HVS [8]. A feature similarity index (FSIM) which compares two low-level features, namely, gradient magnitude and phase congruency, of the reference and distorted images was proposed in [9]. The perceptual similarity (PSIM) which fuses micro- and macrostructures was recently developed in [10].

Compared with FR methods, NR methods are more desirable to provide the image quality without referencing the original image. In literature, some representative NR methods have also been constructed, e.g., distortion identification-based image verity and integrity evaluation (DIIVINE) [11], NR free energy-based robust metric (NFERM) [12], blind image integrity notator using DCT statistics (BLIINDS-II) [13], blind/referenceless image spatial quality evaluator (BRISQUE) [14], etc. All these methods are designed in the similar way, which is feature extraction followed by training a prediction module with those extracted features. The main difference in them concentrates on the features they employed that characterize the image quality. For example, BLIINDS-II exploits features in the DCT domain, while BRISQUE operates in the spatial domain. Besides the aforementioned general-purpose NR methods, there are still some methods dedicated to specific distortions, e.g., JPEG compression [15], [16], blur [17], [18], noise [19], contrast change [20], [21].

Although NR IQA doesn't require the reference image for quality assessment, it is still at an immature stage and keeps a challenging task to blindly predict the image quality. A tradeoff solution to FR IQA and NR IQA is RR IQA, which employs partial information, or some features from the reference image in quality computation. The essence of RR IQA lies in what features should be extracted for precise quality estimation. In [22], the marginal distribution of the wavelet coefficients was firstly modeled through the generalized Gaussian distribution and the model parameters were employed as the features to represent the image quality. Then, this idea was further improved by introducing an divisive normalization transform (DNT) step after wavelet decomposition and the quality prediction performance was thus enhanced [23]. In [24], the normalized histogram of the decomposition coefficients after CSF masking and JND thresholding was utilized as the RR feature. The authors in [25] borrowed the design philosophy of SSIM and distinguished structural and nonstructural changes of the statistical features extracted from DNT domain to design a RR algorithm. Reduced reference entropic differencing (RRED) [26] was proposed to measure the differences between the entropies of wavelet coefficients of reference and distorted images. Clearly, the RR features applied in RRED are the entropies of the image's wavelet coefficients. Xu *et al*. put forward a RR approach, which compares the difference of fractal dimension between the reference and distorted images [27]. In [28], Zhai *et al*. proposed free-energy-induced distortion metric (FEDM), in which a perceptual distance between the reference and distorted images in free energy was defined to predict the psychovisual quality of the distorted image.

Different from most existing RR IQA methods, in this paper, we propose a novel RR IQA metric FSI (Free-energy principle and Sparse representation-based Index for image quality assessment), which is directly derived from the perception mechanism of the brain. On one hand, the free-energy principle models the perception of the visual scene as an active inference process governed by an internal generative model. Using the generative model, the brain can actively infer predictions for the scenes. Since nobody can have knowledge of everything in the world, the internal generative model can't be universal for each one. Therefore, it's reasonable to assume that there exists a discrepancy between the visual input and its brain prediction, which is believed to be closely related to the quality of perceptions [12], [28], [29]. On the other hand, sparse representation has been proven to resemble the strategy for representing natural images in the primary visual cortex of the brain, which is mainly expressed in several aspects [30]–[32]: Firstly, natural images can generally be described in terms of a small number of structural primitives, such as edges, lines, or other elementary features, which can be well captured by sparse representation; Secondly, the receptive fields of simple cells in mammalian primary visual cortex can be characterized as being spatially localized, oriented and bandpass, which is similar to that developed by sparse representation. Thirdly, several theoretical and computational studies have also suggested that neurons in the brain encode visual sensory information with a small number of active neurons at any given point in time, which coincides with the mechanism of sparse representation. Under these neurobiological causes, in this paper, we approximate the internal generative model with sparse representation and propose FSI accordingly. Specifically, in FSI, the reference and distorted images are firstly predicted by sparse representation. Then the difference between the entropies of the prediction discrepancies is defined as the quality index to indicate the image quality. Extensive experiments conducted on four large-scale image databases confirm the effectiveness of our proposed FSI and its superior performance over the competing RR methods, e.g., RRED and FEDM. It should be noted that the needed information of FSI is just one number from the reference image.

The remainder of this paper is organized as follows. In Section II, we review related works of IQA approaches derived from free-energy principle and sparse representation respectively. In Section III, we present the proposed metric for IQA in detail. Experimental results and analysis are given in Section IV. Finally, we conclude this paper in Section V.

II. Related Works

In this section, we deliver a brief review of IQA works based on free-energy principle and sparse representation respectively.

A. IQA Approaches Based on the Free-Energy Principle

As aforementioned, the free-energy principle mainly models the brain activities when perceiving and understanding the visual scenes, which can help us further understand the HVS and inspire the study of new IQA algorithms.

The first IQA model based on free-energy principle is FEDM [28] as mentioned in previous section, which can be considered as the beginning work that introduced free-energy principle into IQA. In FEDM, the authors firstly modeled

the cognitive process through mathematical formulation. Then the linear auto-regressive (AR) model was chosen to simulate the internal generative model in the brain. At last, the differences of free energy between the reference and distorted images was defined to measure the image quality. Considering the brain works with an internal generative mechanism for visual perception, in [33], Wu *et al*. simulated the generative mechanism with AR prediction and decomposed an image into two portions, which are the predicted portion and the residual portion. Then they proposed a FR IQA approach by measuring the degradations of these two portions. As the free-energy principle indicates the discrepancy between the image itself and its brain prediction is correlated with the perceptual quality, Gu *et al*. presented a NR IQA model [12], in which features that can measure the prediction discrepancy, e.g., the gradient magnitude similarity, phase congruency similarity and absolute difference between the image and its brain prediction, were calculated for the construction of the IQA model. By analysing the characteristics of the prediction coefficients, Gu *et al*. proposed a NR IQA approach dedicated to assessing the image's visual sharpness [34]. In addition, free energy was applied to measure the joint effect of multiply distortions in assessing the quality of images degraded by multiply distortions [35].

In summary, the above works from the free-energy principle validate that the free-energy principle can definitely prompt IQA studies.

B. Sparse Representation-Induced IQA Models

Sparse representation refers to representing a signal with a linear superposition of a small number of primitives. This representation strategy has been proved to resemble the neural behaviors in the visual cortex which is responsible for most of our conscious perception of the visual world. Under the neurobiological implications, researchers have also developed sparse representation-based IQA approaches and achieved promising results. Here we review some of them to illustrate.

In [36], He *et al*. proposed a NR IQA method, in which sparse representation is applied in two steps. First, the natural scene statistics (NSS) features extracted from the wavelet domain were represented by sparse representation. Second, the differential mean opinion scores were weighted by the sparse representation coefficients to get the final quality estimation. In [37], a NR sparse representation-based sharpness index was proposed, in which an overcomplete dictionary was firstly trained on natural images, then the blurred image was represented over the dictionary and the normalized energy of the representation coefficients was defined to measure the image sharpness. A sparse feature fidelity (SFF) metric was proposed in [38]. In SFF, sparse features were acquired by a feature detector, which is pre-trained on natural image samples through independent component analysis (ICA). Then the image quality was measured by comparing the sparse features of the reference and distorted images. In [39], Qi *et al*. proposed a RR stereoscopic IQA method based on monocular and binocular perceptual information, in which monocular and binocular features extracted from sparse representation were fed into support vector machine (SVM) to train a model which was used to predict the stereoscopic image quality later.

III. The Proposed FSI for IQA

A. Modelling Visual Perception Mechanism

As the free-energy principle conjectures, the cognitive process is governed by an internal generative model in the brain. With this internal model, the brain is able to generate the corresponding predictions for its encountered visual scenes. For operational amenability, the brain internal model for visual perception, denoted by $\mathcal{G}$, is often supposed to be parametric, which predicts the visual scenes by adjusting its parameters. For clearness, we denote $\mathbf{g}$ as the parameter vector which contains all the parameters of the model $\mathcal{G}$. Therefore, given an image I, its 'surprise' can be computed by integrating the joint distribution $P(I, \mathbf{g})$ over the space of the parameter vector $\mathbf{g}$ as

$$-\log P(I) = -\log \int P(I, \mathbf{g}) d\mathbf{g}. \tag{1}$$

Here we bring an auxiliary term $Q(\mathbf{g}|I)$ into both the denominator and numerator of the right part in (1) which doesn't change its equality as follows:

$$-\log P(I) = -\log \int Q(\mathbf{g}|I) \frac{P(I, \mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g} \tag{2}$$

where $Q(\mathbf{g}|I)$ can be regarded as the posterior distribution of the model parameters given image I, which can be thought of as an approximate posterior to the true posterior of the model parameters $P(\mathbf{g}|I)$ calculated by the brain. When perceiving the input image I, the brain intends to minimize the discrepancy between the approximate posterior $Q(\mathbf{g}|I)$ and the true posterior $P(\mathbf{g}|I)$. According to Jensen's inequality, (2) can be translated into

$$-\log P(I) \leq -\int Q(\mathbf{g}|I) \log \frac{P(I, \mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g}. \tag{3}$$

According to statistical physics and thermodynamics [40], the right side of equation (3) is defined as "free energy" as follows:

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

Clearly, $F(\mathbf{g})$ defines an upper bound of 'surprise' for image I. For intuitive understanding, with $P(I, \mathbf{g}) = P(\mathbf{g}|I)P(I)$, we further derive (4) as

$$\begin{aligned} F(\mathbf{g}) &= \int Q(\mathbf{g}|I) \log \frac{Q(\mathbf{g}|I)}{P(\mathbf{g}|I)P(I)} d\mathbf{g} \\ &= -\log P(I) + \int Q(\mathbf{g}|I) \log \frac{Q(\mathbf{g}|I)}{P(\mathbf{g}|I)} d\mathbf{g} \\ &= -\log P(I) + \mathbf{KL}(Q(\mathbf{g}|I) \| P(\mathbf{g}|I)) \end{aligned} \tag{5}$$

where $\mathbf{KL}(\cdot)$ refers to the Kullback-Leibler divergence between the approximate posterior and the true posterior distributions and it's nonnegative. It is clearly seen that the free energy $F(\mathbf{g})$ is

greater than or equal to the image 'surprise'-$\log P(I)$.[1] In visual perception, the brain tries to minimize $\mathbf{KL}(Q(\mathbf{g}|I)\|P(\mathbf{g}|I))$ of the divergence between the approximate posterior and its true posterior distributions when perceiving image I.

B. The Approximation of the Internal Generative Model

The modelling of visual perception mechanism in free-energy principle assumes that the internal model in the brain is parametric and it explains the encountered scenes by minimizing the divergence between the approximate posterior and the true posterior distributions of the model parameters. This formulation models the brain perceiving an outside image in a probabilistic manner. For practical and computational employment of the free-energy principle into IQA, figuring out the brain internal model should come to the first. However, the true form of the internal generative model is still unknown till now. To tackle this problem, researchers proposed to approximate the brain model with existing image models. In the free-energy-induced IQA works mentioned in Section II-A, the internal model $\mathcal{G}$ was often approximated with the linear auto-regressive (AR) model for its simplicity and ability to represent a wide range of natural images, which is described as

$$y_n = \mathcal{X}^k(y_n)\mathbf{a} + e_n \tag{6}$$

where y_n is a pixel to be predicted, $\mathcal{X}^k(y_n)$ is the vector consist of k nearest neighbors of y_n, $\mathbf{a} = (a_1, a_2, ..., a_k)^T$ is the AR coefficient vector, the superscript "T" is transpose operation. e_n represents the additive Gaussian noise. To acquire the AR coefficient vector $\mathbf{a}$, the following optimization problem is presented:

$$\mathbf{a}^* = \underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2 \tag{7}$$

where $\mathbf{y} = (y_1, y_2, , , y_k)^T$ and $\mathbf{X}(i,:) = \mathcal{X}^k(y_i)$. This equation can be conveniently solved with the least square method and the solution $\mathbf{a}^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. Here $\mathbf{a}^*$ acts as the internal model parameter vector $\mathbf{g}$ actually. Then y_n can be predicted with $\mathcal{X}^k(y_n)\mathbf{a}^*$. By predicting each pixel in this manner, the whole reconstructed image can be finally obtained, which stands for the brain prediction for the input image. Compared to AR prediction, sparse representation denotes a new linear strategy to represent the observed image. Suppose that $\mathbf{x} \in \mathbb{R}^n$ is a vectorized patch to be represented, then $\mathbf{x}$ can be denoted by a linear combination of primitives from a predefined or trained dictionary as

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{E} \tag{8}$$

where $\mathbf{D}$ is the dictionary denoted by $[\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3 ... \mathbf{d}_K]$ that contains K primitives. $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, ...\alpha_K\}^T$ denotes the representation coefficient vector and $\mathbf{E}$ points out the representation error. Here, the coefficient vector $\boldsymbol{\alpha}$ can be regarded as the brain model parameter vector $\mathbf{g}$. Different from AR prediction, sparse representation is performed on the basis of image patch, not pixel, which means sparse representation can be more efficient than the AR representation. In addition, the dictionary used in sparse representation can be predefined or trained during the representation process, which indicates that sparse representation is more flexible for predicting the visual scenes. The most important point, as we stated before, sparse representation has been verified to resemble the strategy for representing natural images in the primary visual cortex of the brain and already achieved success in IQA tasks. Based on these careful analysis, in this paper, we approximate the internal generative model in the brain with sparse representation. Specific implementation of sparse representation for natural images and the solution to representation coefficients will be given in the following.

C. Patch-Based Sparse Representation

According to the above analysis, we approximate the internal generative model with sparse representation. Usually, the basic unit in sparse representation for an image is image patch [41]–[44]. Mathematically, given an image I, we extract a patch $\mathbf{x}_k \in \mathbb{R}^{B_s}$ of size $\sqrt{B_s} \times \sqrt{B_s}$ from I by

$$\mathbf{x}_k = \mathbf{R}_k(I) \tag{9}$$

where $\mathbf{R}_k(\cdot)$ is the extraction operator that extracts the image patch $\mathbf{x}_k$ from image I at location k, $k = 1, 2, 3...n$, n gives the total number of image patches. The transpose operation of $\mathbf{R}_k(\cdot)$, denoted by $\mathbf{R}_k^T(\cdot)$ is to put the image patch $\mathbf{x}_k$ back to the position k in the reconstructed image. With all the extracted patches, image I can be reconstructed by

$$I = \sum_{k=1}^{n} \mathbf{R}_k^T(\mathbf{x}_k) . \Big/ \sum_{k=1}^{n} \mathbf{R}_k^T(\mathbf{1}_{B_s}) \tag{10}$$

where the notation "./" represents the operation of element-wise division of two vectors and $\mathbf{1}_{B_s}$ refers to the vector of size B_s whose elements are all 1. This equation indicates an abstraction strategy of averaging all the patches for recovering image I.

For the specific extracted patch $\mathbf{x}_k$, its sparse representation over a dictionary $\mathbf{D} \in \mathbb{R}^{B_s \times M}$ refers to finding a sparse vector $\boldsymbol{\alpha}_k \in \mathbb{R}^M$ (i.e., most of the elements in $\boldsymbol{\alpha}_k$ are zero or close to zero) to satisfy

$$\mathbf{x}_k = \mathbf{D}\boldsymbol{\alpha}_k \tag{11}$$

or approximate as

$$\mathbf{x}_k \approx \mathbf{D}\boldsymbol{\alpha}_k \quad s.t. \quad \|\mathbf{x}_k - \mathbf{D}\boldsymbol{\alpha}_k\|_p \leq \xi \tag{12}$$

where $\|\cdot\|_p$ refers to the l^p norm. ξ is a positive threshold. What we request for the representation coefficient vector satisfies

$$\boldsymbol{\alpha}_k^* = \underset{\boldsymbol{\alpha}_k}{\operatorname{argmin}} \|\boldsymbol{\alpha}_k\|_p \ s.t. \ \mathbf{x}_k = \mathbf{D}\boldsymbol{\alpha}_k . \tag{13}$$

This equation can be transformed into an unconstrained optimization problem for solving as

$$\boldsymbol{\alpha}_k^* = \underset{\boldsymbol{\alpha}_k}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{x}_k - \mathbf{D}\boldsymbol{\alpha}_k\|_2 + \lambda\|\boldsymbol{\alpha}_k\|_p \tag{14}$$

where the first term is the reconstruction fidelity constraint and the second term is to punish the sparsity of the representation

[1] "surprise" is also known as surprisal or negative log-evidence, which is a measure of self-information. Here, "surprise" is defined as -$\log P(I)$, indicating that the higher the probability that an image or a visual stimulus I is observed, the lower its "surprise" is.

coefficient vector. λ is a positive constant for balancing the importance of these two terms. p takes 0 or 1. If $p = 0$, the sparsity of $\boldsymbol{\alpha}_k$ is strictly measured by l^0-norm, which refers to calculating the number of nonzero coefficients in $\boldsymbol{\alpha}_k$. That's what we desire for the representation coefficients. However, the l^0-minimization is non-convex and NP-hard. It is usually solved by greedy algorithms, such as the orthogonal matching pursuit (OMP) algorithm [45]. An alternative way for solving the l^0-minimization problem is to replace l^0-norm with l^1-norm, namely, $p = 1$. Then (14) becomes l^1-minimization, which is convex and can be solved by some large-scale methods [46]–[48].

By solving (14), we can get the corresponding representation coefficient vector $\boldsymbol{\alpha}_k^*$ for representing the image patch $\mathbf{x}_k$. Then we bring $\mathbf{D}\boldsymbol{\alpha}_k^*$, the sparse representation of $\mathbf{x}_k$, into (10) leading to

$$I' = \sum_{k=1}^{n} \mathbf{R}_k^T (\mathbf{D}\boldsymbol{\alpha}_k^*). \Big/ \sum_{k=1}^{n} \mathbf{R}_k^T (\mathbf{1}_{B_s}) \tag{15}$$

where I' refers to the sparse representation for the entire image I, which serves as the brain prediction for image I as we supposed before.

D. The Perceptual Quality Index

With the internal generative model, the brain yields the corresponding prediction for the input image. However, the internal model can't be universal resulting in a discrepancy between the image and its brain prediction. The prediction discrepancy is believed to be closely related to the quality of human perceptions. More precisely, the quality of perceptions can be quantified mathematically by the uncertainty of the prediction discrepancy [28]. Then, the perceptual quality degradation can be measured by the uncertainty variation of the prediction discrepancy. At first, we define the prediction residual as the discrepancy between the image and its brain prediction as

$$R = |I - I'| \tag{16}$$

where R refers to the prediction residual image, I represents the input image and I' represents the brain prediction of image I, which is obtained through sparse representation in (15). "$|\cdot|$" is the magnitude operation. The uncertainty of the prediction discrepancy R can be measured by its entropy. Additionally, as visual saliency studies demonstrate the HVS would selectively pay attention to the 'salient' regions of the image, while pay little attention to the visual insignificant regions [49]–[51], then immediately calculating the entropy of R will underestimate this characteristic as entropy is unable to discriminate visual importance. Considering this, we define a salient prediction residual image R_s by percentile strategy which is widely adopted in IQA algorithms [34]. Specifically, we constitute R_s by the pixels in R that are corresponding to the $l\%$ most salient pixels of image I. Suppose S_I is the saliency map of image I, each pixel value in S_I indicates the visual importance degree of the corresponding pixel in I. We firstly create a 0–1 value mask by

$$M = \mathbb{F}(S_I) \tag{17}$$

where M refers to the 0-1 mask matrix, $\mathbb{F}(\cdot)$ refers to the assignment function that assigns 1 to $M(i,j)$ if $S_I(i,j)$ belongs to the $l\%$ largest values in S_I, otherwise $M(i,j)$ is assigned to 0. Then R_s can be calculated by

$$R_s = R. * M \tag{18}$$

where ".*" represents the operation of element-wise multiplication of two matrixes. In implementation, S_I can be obtained in advance by some mature saliency prediction methods, such as GBVS [49], IS [52] etc. Next, we calculate the entropy of the salient prediction residual image as

$$E = -\sum_{i=0}^{255} p_i \log_2 p_i \tag{19}$$

where E gives the entropy of R_s, p_i is the probability density of ith gray scale in R_s.

As the free-energy principle indicates, the entropy E of the prediction discrepancy can characterize image quality degradations. To illustrate this intuitively, we chose three standard reference images with their corresponding distorted images from TID2013 database [53]. Three common distortion types were investigated which are Addictive Gaussian noise, Gaussian blur and JPEG compression respectively. The distortion levels are 0, 1, 2, 3, 4, 5, where 0 means no distortion and 5 refers to the worst distortion. Then we calculated Es of all the images and plot the results in Fig. 1. As can be observed, for the three images under each distortion type, E changes monotonously as the distortion level increases, which indicates E can effectively capture the image quality degradations.Therefore, we can measure the quality degradation by inspecting the variation of E between the reference image and its corresponding distorted image as

$$FSI = |E_r - E_d| \tag{20}$$

where E_r represents E of the reference image and E_d represents E of the corresponding distorted image. Here E can be regarded as the RR feature which is on behalf of the image quality. It's clear to see that the smaller FSI is, the higher the perceptual quality of the distorted image is. FSI equalling 0 means the highest perceptual quality. Since FSI requires information from the reference image for predicting the image quality, it belongs to RR IQA methods eventually. As the needed information of the reference image (the entropy of the salient prediction residual image) is just a single scalar, FSI can maximally reduce the data rate for RR quality evaluation.

IV. Experimental Results

In this section, we will validate the effectiveness of the proposed FSI through extensive experiments. Some important issues about FSI will also be discussed.

A. Experimental Protocol

To test the proposed FSI, we conducted experiments on four widely-adopted image databases, which are LIVE [56],

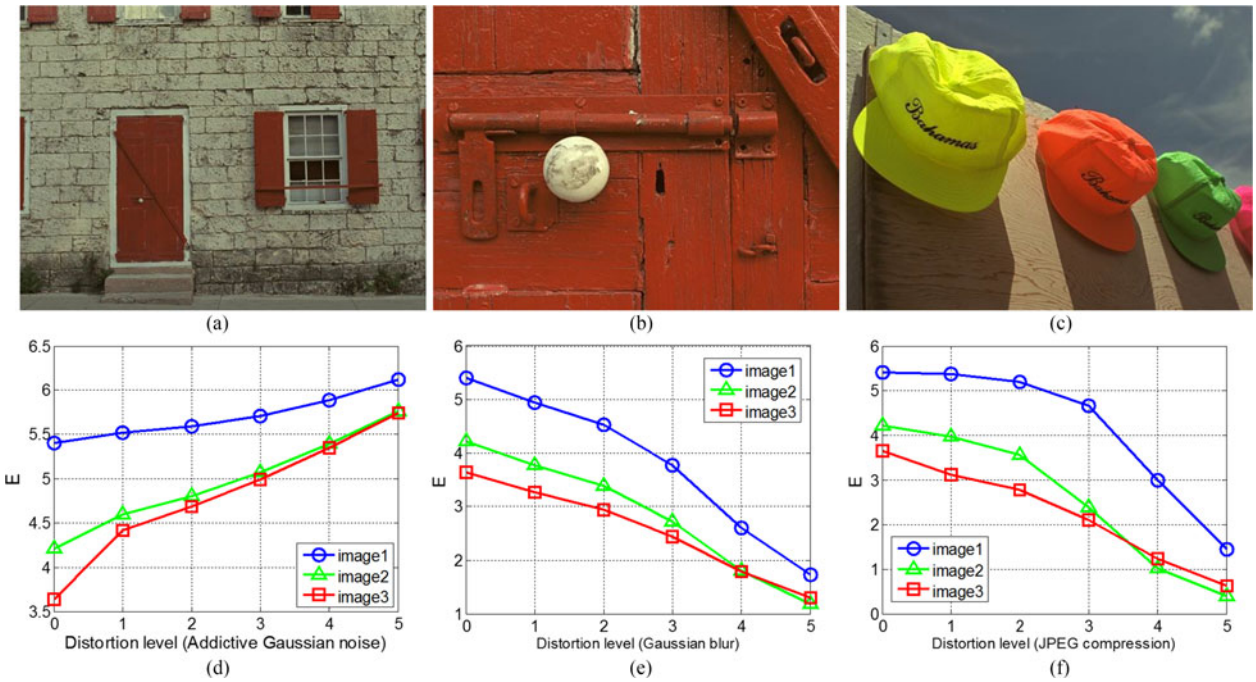


Fig. 1. Scatter plots of the entropies of the salient prediction residual images (E) against different distortion levels. (a)–(c) show the three standard images from TID2013 database which are denoted as image1, image2, and image3, respectively. (d)–(f) refer to the distortion types of Addictive Gaussian noise, Gaussian blur, and JPEG compression, respectively. The higher the distortion level is, the worse the quality of the image is.

TID2013 [53], CSIQ [57] and Toyama [58]. A brief introduction to these four databases is listed below:

1) The LIVE database contains a total of 779 distorted images from 29 pristine images. There are five distortion types involved, which are JPEG2000 compression, JPEG compression, white noise, gaussian blur and fast fading.
2) The TID2013 database is composed of 3000 distorted images which are generated from 25 reference images with 24 distortion types at 5 distortion levels. The distortion types are addictive gaussian noise, addictive noise in color components, spatially correlated noise, masked noise, high frequency noise, impulse noise, quantization noise, gaussian blur, image denoising, JPEG compression, JPEG2000 compression, JPEG transmission errors, JPEG2000 transmission errors, non eccentricity pattern noise, local block-wise distortions of different intensity, mean shift, contrast change, change of color saturation, multiplicative gaussian noise, comfort noise, lossy compression of noisy images, image color quantization with dither, chromic aberrations, sparse sampling and reconstruction.
3) The CSIQ database consists of 866 images distorted from 30 original images with six types of distortions, which are JPEG compression, JPEG2000 compression, addictive noise, global contrast decrements, additive pink gaussian noise and gaussian blur respectively.
4) The Toyama database includes 168 distorted images generated by JPEG compression and JPEG2000 compression respectively.

For simplifying the computational complexity and quick test, in this paper we have only performed FSI on the luminance channel of the image. Therefore, distortions focused on the image chrominance were not involved in our test, which include additive noise in color components, change of color saturation, image color quantization with dither, chromic aberrations in TID2013 database and additive pink gaussian noise in CSIQ database respectively.

For evaluating the prediction performance of the objective IQA models, we employed four statistical indexes, which are Kendall's rank correlation coefficient (KROCC), Spearman Rank order Correlation coefficient (SROCC), Pearsons linear correlation coefficient (PLCC) and root mean square error (RMSE) respectively. The SROCC and KROCC values can indicate the prediction monotonicity of the quality metric, PLCC reflects the prediction accuracy and RMSE points out the prediction consistency. Therefore, these four indexes demonstrate the prediction performance from different aspects. A superior IQA metric is expected to achieve values close to 1 in SROCC, KROCC and PLCC, while close to 0 in RMSE.

As suggested by VQEG [59], before computing PLCC and RMSE, the objective results are needed to be mapped to subjective ratings through nonlinear regression. Toward this end, we apply a five-parameter logistic function as

$$q(z) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2 \cdot (z - \beta_3))} \right) + \beta_4 \cdot z + \beta_5 \tag{21}$$

with z and $q(z)$ being the input objective score and the mapped score. β_j (j=1, 2, 3, 4, 5) are free parameters to be determined during the curve fitting process.

TABLE I
OVERALL PERFORMANCE COMPARISON ON LIVE, TID2013, CSIQ, AND TOYAMA DATABASES

Database	Index	PSNR FR	SSIM [5] FR	SFF [38] FR	IL-NIQE [54] NR	NFERM [12] NR	C-DIIVINE [55] NR	FEDM [28] RR	RRED [26] RR	FSI (pro.) RR
LIVE (779 images)	SROCC	0.8756	**0.9479**	**0.9649**	**0.8978**	training	training	0.7947	0.7653	0.8826
	KROCC	0.6865	**0.7963**	**0.8365**	**0.7128**	training	training	0.5964	0.5833	0.6957
	PLCC	0.8723	**0.9449**	**0.9632**	**0.9025**	training	training	0.7976	0.6880	0.8821
	RMSE	13.3597	**8.9455**	**7.3461**	**11.7702**	training	training	16.4786	19.8278	12.8720
TID2013 (2500 images)	SROCC	**0.6675**	**0.8018**	**0.8637**	0.5349	0.3509	0.3810	0.1221	0.5926	0.5798
	KROCC	**0.4881**	**0.6056**	**0.6736**	0.3811	0.2470	0.2663	0.0827	0.4313	0.4101
	PLCC	**0.6750**	**0.8105**	**0.8778**	0.6069	0.4882	0.5411	0.1842	0.6641	0.6111
	RMSE	**0.9270**	**0.7359**	**0.6020**	0.9986	1.0965	1.0566	1.2349	0.9394	0.9945
CSIQ (716 images)	SROCC	0.8206	**0.8829**	**0.9656**	0.8099	0.8047	0.8187	0.8308	0.6877	**0.9175**
	KROCC	0.6229	**0.6993**	**0.8360**	0.6209	0.6330	0.6418	0.6236	0.5015	**0.7479**
	PLCC	0.8018	0.8627	**0.9670**	0.8671	0.8798	**0.8807**	0.8113	0.6477	**0.9265**
	RMSE	0.1606	0.1359	**0.0685**	0.1338	0.1277	**0.1273**	0.1571	0.2047	**0.1011**
Toyama (168 images)	SROCC	0.6132	**0.8794**	**0.8992**	0.7114	0.8498	**0.8773**	0.7779	0.4532	0.8014
	KROCC	0.4443	**0.6939**	**0.7217**	0.5105	0.6587	**0.7095**	0.5864	0.3220	0.6033
	PLCC	0.6428	**0.8887**	**0.9030**	0.7247	0.8517	**0.8758**	0.7804	0.4972	0.8070
	RMSE	0.9587	**0.5738**	**0.5378**	0.8625	0.6558	**0.6041**	0.7826	1.0859	0.7391
Dir. AVG	SROCC	0.7442	**0.8780**	**0.9234**	0.7385	0.6685	0.6923	0.6314	0.6247	**0.7953**
	KROCC	0.5605	**0.6988**	**0.7670**	0.5563	0.5129	0.5392	0.4723	0.4595	**0.6142**
	PLCC	0.7480	**0.8767**	**0.9277**	0.7753	0.7399	0.7659	0.6434	0.6242	**0.8067**
Wei. AVG	SROCC	**0.7306**	**0.8462**	**0.9016**	0.6572	0.4717	0.4982	0.3963	0.6356	0.7035
	KROCC	**0.5466**	**0.6610**	**0.7340**	0.4896	0.3491	0.3678	0.2922	0.4674	0.5294
	PLCC	**0.7324**	**0.8478**	**0.9101**	0.7117	0.5891	0.6296	0.4309	0.6590	0.7240

B. Implementation Issues

In FSI, the reference and distorted images were firstly predicted through patch-based sparse representation as described in Section III-C. In this process, we divided the image into 8×8 patches, namely B_s equals 64. The overcomplete DCT dictionary was employed as the predefined dictionary $\mathbf{D}$ for sparse representation because of its wide use in image processing algorithms. The dimension of $\mathbf{D}$ was 64×144 with totally 144 atoms available for representing each patch. To be specific, $\mathbf{D}$ was created in line with [60], namely, forming a 1D-DCT $\mathbf{A}_{1D}$ of size 8×12 firstly, where the k-th atom (k=1, 2,..., 12) is given by $a_k = \cos((i-1)(k-1)\pi/12)$, i=1, 2,... 8. Then all the atoms except the first one were disposed by removing their mean. At last, the dictionary $\mathbf{D}$ was calculated by a Kronecker-product $\mathbf{D} = \mathbf{A}_{1D} \otimes \mathbf{A}_{1D}$. In implementation, we embedded $\mathbf{D}$ into the algorithm so that additional storage and transmission for $\mathbf{D}$ can be saved. In (14), the parameter p was set to 0 and this equation was solved by OMP algorithm [45]. The sparsity (the number of nonzero coefficients for representing each patch) was set to 6. The popular GBVS model was employed in our proposed method for saliency prediction and the threshold l was set to 30 experimentally. Experiments about saliency prediction can be referred to in Section IV-G.

C. Overall Prediction Performance Evaluation

In this section, the prediction performance in terms of SROCC, KROCC, PLCC and RMSE of the proposed FSI with competing methods is given. The overall results are tabulated in Table I, in which the best three results in each indice are highlighted in boldface. As can be seen, we compare FSI with eight representative IQA metrics, which are PSNR, SSIM [5], SFF [38], IL-NIQE [54], NFERM [12], C-DIIVINE [55], FEDM [28], RRED [26]. Among them, PSNR, SSIM and SFF are the classical FR methods, IL-NIQE, NFERM and C-DIIVINE represent state-of-the-art NR methods, FEDM and RRED belong to RR methods. It should be noted that RRED has different modes according to the amount of needed information from the original image for quality estimation. Here, we selected the mode of referencing one single scalar from the original image, which is the same as FSI. Certainly, there are still some other RR models as we introduced in Section I, while their number of features extracted for IQA are different from FSI, e.g., [24] needs 24 features, [22] needs 18 features, etc. FEDM and RRED are both in need of one feature, therefore we include them for fair comparison. In Table I, "Dir. AVG" refers to directly averaging indexes (SROCC, KROCC and PLCC) over the four databases. "Wei. AVG" refers to the weighted average with the weight determined by the number of images in each database, which is calculated as

$$\bar{\delta} = \frac{\sum_i \delta_i \cdot \pi_i}{\sum_i \pi_i} \tag{22}$$

where $\bar{\delta}$ is the weighted average, δ_i refers to one of SROCC, KROCC and PLCC for the ith database, π_i gives the number of images in the ith database, i.e., 779 for LIVE, 2500 for TID2013, 716 for CSIQ and 168 for Toyama. The LIVE database is used for training a prediction model for NFERM and C-DIIVINE. Therefore, the performance results of NFERM and C-DIIVINE on the LIVE database are not included and their average results are calculated over the other three databases.

TABLE II
STATISTICAL SIGNIFICANCE COMPARISON WITH T-TEST

t-test	PSNR FR	SSIM [5] FR	SFF [38] FR	IL-NIQE [54] NR	NFERM [12] NR	C-DIIVINE [55] NR	FEDM [28] RR	RRED [26] RR
LIVE	0	−1	−1	−1	–	–	1	1
TID2013	−1	−1	−1	0	1	1	1	−1
CSIQ	1	1	−1	1	1	1	1	1
Toyama	1	−1	−1	1	0	−1	0	1

The symbol 1, 0, or −1 indicates that the proposed metric (FSI) is statistically (with 95% confidence) better, indistinguishable, or worse than the compared metric.

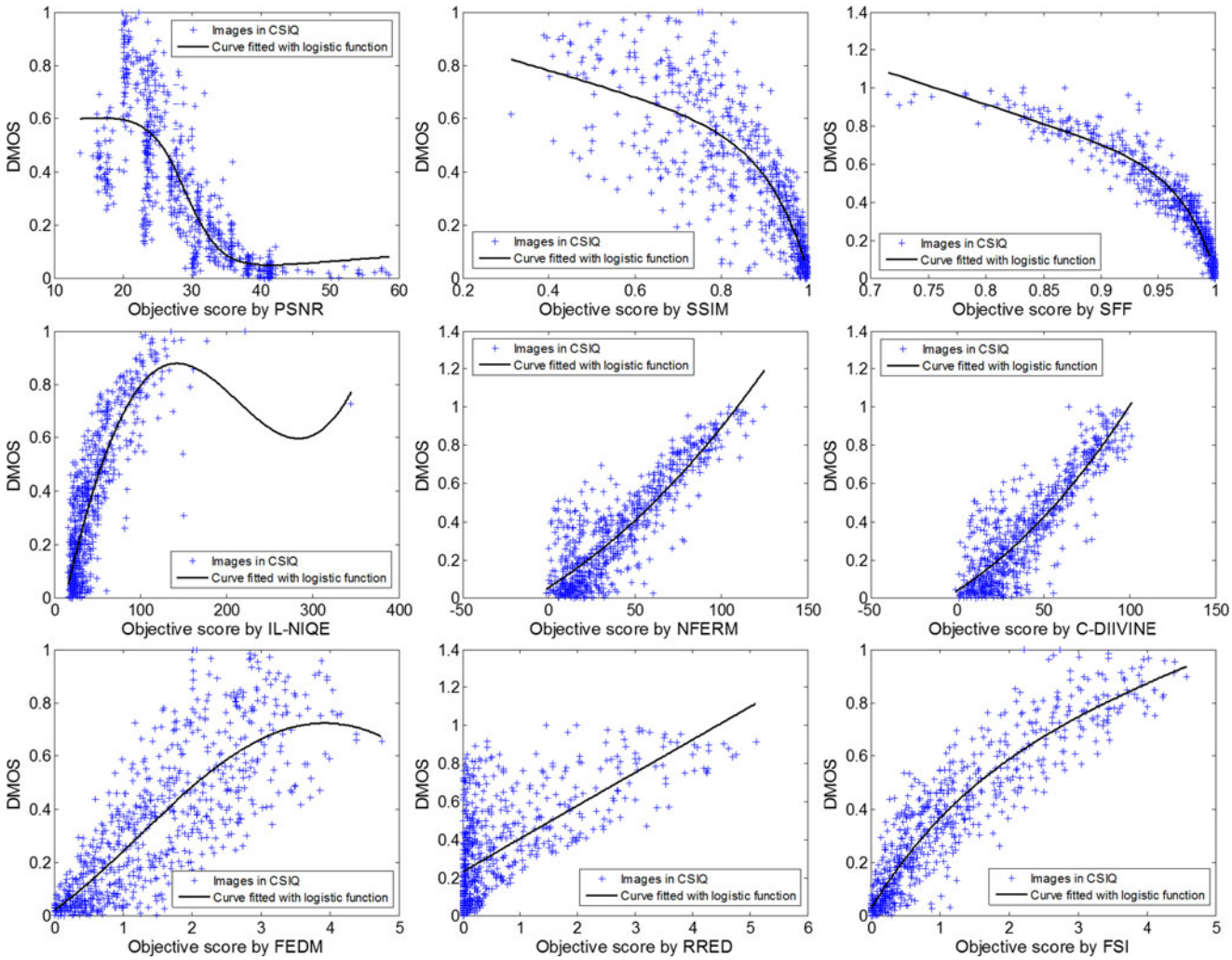


Fig. 2. Scatter plots of subjective scores (DMOS) against objective scores obtained by IQA models on the CSIQ database.

Compared with the FR methods in Table I, we find that the FR method SFF is the best performed method undoubtedly, while the proposed method FSI still achieves better performance than PSNR on LIVE, CSIQ and Toyama databases and outperforms SSIM on CSIQ database, given the fact that only one number of the original image is referenced for FSI. Compared with the NR methods, FSI is able to perform consistently well on the four databases, while the NR methods may work better on some database but can't provide consistent results on other databases. For example, C-DIIVINE achieves good results on Toyama database, it fails to deliver equivalent performance on TID2013 and CSIQ databases. It should be emphasized that FSI outperforms its same kind methods FEDM and RRED remarkably except RRED exceeds FSI on TID2013 database.

To evaluate the statistical significance of the proposed method with the compared approaches, we employ t-test to the obtained scores of the objective IQA models. The statistical significance results are reported in Table II, where the symbol "1", "0" and "−1" implies the proposed method is statistically (with 95% confidence) better, indistinguishable or worse than the compared

TABLE III
SROCC VALUES OF THE IQA METRICS ON INDIVIDUAL DISTORTION TYPES

Database	Dis. Type	PSNR FR	SSIM [5] FR	SFF [38] FR	IL-NIQE [54] NR	NFERM [12] NR	C-DIIVINE [55] NR	FEDM [28] RR	RRED [26] RR	FSI (pro.) RR
LIVE	FF	0.8907	**0.9556**	**0.9529**	0.8328	training	training	0.8229	**0.9155**	0.8861
	GB	0.7823	**0.9517**	**0.9752**	0.9158	training	training	0.7594	0.9517	**0.9642**
	JP2K	0.8954	**0.9614**	**0.9672**	0.8942	training	training	0.9200	**0.9234**	0.9023
	JPEG	0.8809	**0.9764**	**0.9786**	0.9419	training	training	0.9225	0.8358	**0.9623**
	AWGN	**0.9854**	0.9694	**0.9859**	**0.9807**	training	training	0.9152	0.9161	0.9231
	Dir. AVG	0.8869	**0.9629**	**0.9719**	0.9131	training	training	0.8680	0.9085	**0.9276**
TID2013	AGN	**0.9291**	0.8671	**0.9066**	**0.8760**	0.8582	0.8436	0.7485	0.7496	0.7086
	SCN	**0.9200**	0.8515	**0.8982**	**0.9231**	0.2180	0.6261	0.6920	0.7800	0.7034
	MN	**0.8323**	**0.7767**	**0.8185**	0.5121	0.2207	0.6620	0.7189	0.4007	0.7210
	HFN	**0.9140**	0.8634	**0.8977**	0.8683	0.8814	**0.8824**	0.7889	0.7772	0.7710
	IN	**0.8968**	0.7503	**0.7871**	**0.7554**	0.1728	0.7354	0.7383	0.5323	0.7040
	QN	**0.8808**	**0.8657**	0.8607	**0.8726**	0.7747	0.0963	0.0732	0.7308	0.2618
	GB	0.9149	**0.9668**	**0.9675**	0.8145	0.8498	0.8698	0.8896	**0.9672**	0.9501
	DEN	**0.9480**	**0.9254**	0.9091	0.7491	0.6389	0.8155	0.7998	**0.9159**	0.8312
	JPEG	**0.9189**	**0.9200**	**0.9273**	0.8355	0.8720	0.8841	0.7832	0.6974	0.8576
	JP2K	0.8840	**0.9468**	**0.9571**	0.8581	0.8097	0.9055	0.8396	0.8970	**0.9060**
	JGTE	**0.7685**	**0.8493**	**0.8831**	0.2821	0.1322	0.3246	0.7445	0.6304	0.3632
	J2TE	**0.8883**	**0.8828**	**0.8708**	0.5243	0.1681	0.4575	0.6094	0.7211	0.6358
	NEPN	**0.6863**	**0.7821**	**0.7668**	0.0803	0.0645	0.0675	0.5049	0.4173	0.4455
	Block	0.1552	**0.5720**	0.1786	0.1355	0.2023	0.0239	**0.5375**	0.1708	**0.5591**
	MS	**0.7671**	**0.7752**	**0.6654**	0.1845	0.0218	0.0320	0.5438	0.5611	0.6198
	CTC	0.4400	0.3775	0.4691	0.0133	0.2185	0.4162	**0.4958**	**0.5433**	**0.5683**
	MGN	**0.8905**	**0.7803**	**0.8434**	0.6924	0.7164	0.7363	0.7007	0.6905	0.6373
	CN	**0.8411**	**0.8566**	**0.9007**	0.3600	0.1433	0.0132	0.4890	0.7182	0.5287
	LCNI	**0.9145**	**0.9057**	**0.9262**	0.8287	0.6541	0.7001	0.6599	0.6272	0.3605
	SSR	0.9042	**0.9461**	**0.9522**	0.8650	0.7850	0.8844	0.8297	**0.9310**	0.8815
	Dir. AVG	**0.8147**	**0.8231**	**0.8193**	0.6015	0.4701	0.5488	0.6594	0.6730	0.6507
CSIQ	GCD	0.8621	0.7922	**0.9536**	0.4996	0.3774	0.3720	**0.9550**	0.9382	**0.9550**
	JP2K	0.9361	**0.9605**	**0.9762**	0.9059	0.9048	0.8931	0.8945	**0.9387**	0.9342
	JPEG	0.8879	**0.9543**	**0.9641**	0.8993	0.9222	0.9157	0.9166	0.8220	**0.9508**
	GB	0.9291	0.9609	**0.9751**	0.8576	0.8964	0.9076	0.8522	**0.9649**	**0.9634**
	AWGN	**0.9363**	0.8974	**0.9469**	0.8497	**0.9220**	0.8966	0.8246	0.8010	0.8490
	Dir. AVG	0.9103	**0.9131**	**0.9632**	0.8024	0.8046	0.7970	0.8886	0.8930	**0.9305**
Toyama	JPEG	0.2868	0.8590	**0.9018**	0.7091	0.8642	**0.8820**	0.7574	0.6352	**0.8922**
	JP2K	0.8605	**0.9399**	**0.9475**	0.7383	0.8741	0.8744	**0.8979**	0.4498	0.7988
	Dir. AVG	0.5737	**0.8995**	**0.9246**	0.7237	0.8691	**0.8782**	0.8277	0.5425	0.8455

method in each column. As can be observed in Table II, in most cases the proposed method is better than other metrics except SSIM and SFF over the four databases, which proves the superiority of FSI statistically.

Furthermore, we show the scatter plots of subjective scores against objective scores given by the IQA models on CSIQ in Fig. 2, where the blue "+" represents the test images and the black curves are fitted through (21). It can be observed that the points of FSI cluster close to the fitted curve, which means the objective scores predicted by FSI well correlate with subjective scores.

D. Performance Comparison on Individual Distortion Types

Besides evaluating the overall performance of the IQA methods on the whole image databases, we also want to know their prediction ability for specific distortion types. Therefore, in this experiment, we examine the prediction performance of the IQA metrics on each distortion type. We report the experimental results in terms of SROCC in Table III, where the top three methods for each distortion type are highlighted in boldface and "Dir. AVG" refers to averaging the SROCC values over all the distortion types in each database. There are totally 32 groups of distorted images in this test.

In Table III, it's obvious that the FR methods SFF, SSIM and PSNR are the top three methods, which are marked 32, 26 and 18 times respectively. Apart from the FR methods, FSI is marked 11 times and ranks first followed by RRED with 8 times highlighted. Therefore, we conclude that FSI also performs well on specific distortion types.

E. Performance Comparison between AR and Sparse Representation

In Section III-B, we analyzed sparse representation can be more effective and efficient than AR representation in simulating the internal generative model for quality prediction. In order to verify this, we conducted experiments by simulating the internal model with AR and sparse representation respectively. Specifically, under FSI framework, we altered the prediction manner from sparse representation to AR representation with other operations in FSI all fixed. The sparse representation

TABLE IV
SROCC VALUES COMPARISON BETWEEN AR AND SR ON LIVE, TID2013, CSIQ, AND TOYAMA DATABASES

Database	AR	SR
LIVE	0.7665	0.8826
TID2013	0.5084	0.5798
CSIQ	0.8220	0.9175
Toyama	0.7856	0.8014

configurations have been stated in Section IV-B. The AR representation configurations are set the same as that of FEDM. The overall performance on the four databases are summarized in Table IV, where the prediction performance is measured by SROCC. "SR" refers to sparse representation. As can be observed in Table IV, the overall SROCC values of SR are consistently higher than that of AR on the four databases, which indicates that simulating the internal model with sparse representation is more effective than with AR representation. Besides, we inspect the computational time of AR and SR simulations respectively.

Specifically, we chose a standard image from TID2013 database (i01_01_1.bmp) and record the computational time of AR and SR respectively. The hardware platform is a Thinkpad X220 computer with a 2.5GHz CPU and 4G RAM. The software platform is Matlab R2012a. The running time of AR is 87.70 seconds, while the running time of SR is only 5.03 seconds. Clearly, the computational time of SR is much shorter than AR, which verifies sparse representation is much more efficient than AR representation for quality prediction. Therefore, we verify that approximating the internal generative model with sparse representation is not only more effective but also much more efficient than with AR representation.

F. Impact of Sparsity on Prediction Performance

In sparse representation, sparsity refers to the number of atoms for representing each image patch, it also refers to the number of nonzero coefficients in the representation coefficient vector.

In this section, we investigate the impact of sparsity on the quality evaluation. To be specific, we conducted experiments on the CSIQ database by varying the sparsity in FSI. The experimental results are tabulated in Table V, where the overall prediction performance is measured by SROCC. It is noted that the prediction performance grows as the sparsity increases. When the sparsity is small, e.g., 1 or 2, the overall prediction performance is relatively low. While the sparsity becomes bigger than 5, the performance of FSI rises to a higher level. This is because sparse representation with small sparsity can't well approximate the internal generative model, which lowers the prediction performance. For visualization, we show an example of represented images with different sparsities in Fig. 3. It's clear to see that the reconstructed image with sparsity being 1 can't be represented accurately, which leads to bad visual quality, while the image reconstructed with sparsity equal to 6 reveals much better quality. As observed in Table V, when the sparsity becomes higher than 5, the performance changes slightly. In addition, a larger sparsity results in higher computational cost. For properly balancing the prediction performance and computational cost, we set the default value of sparsity to 6 in FSI.

TABLE V
IMPACT OF SPARSITY ON THE PREDICTION PERFORMANCE

Sparsity	SROCC
1	0.8220
2	0.8745
3	0.8966
4	0.9084
5	0.9151
6	0.9175
7	0.9186
8	0.9185

(a) (b)

Fig. 3. Example of reconstructed images with different sparsities. (a) Reconstructed image with sparsity = 1. (b) Reconstructed image with sparsity = 6.

G. Testing of Different Saliency Models at Different Proportions

Considering the important characteristic of HVS, i.e., visual saliency, we define the salient prediction residual image constituted by the $l\%$ most salient pixels of the prediction residual image. In this regard, saliency detection should be performed in advance. Without loss of generality, we tested seven representative saliency models for saliency detection, which are GBVS [49], IS [52], Covsal [50], SWD [61], LRK [62], FES [63] and RCSS [64]. In addition, we set l from 10 to 100 at a proper interval of 10 and conducted experiments on the CSIQ database. Table VI lists the prediction performance measured by SROCC and the best performance for each saliency model is marked in boldface. 100% means all pixels in the prediction residual image are involved in quality computation, which is also equivalent to no saliency detection for FSI. By observing Table VI, we find that the best results for all the saliency models are all higher than that on the proportion of 100%, which confirms taking visual saliency into consideration can further improve the predicting ability of FSI. As SROCC of GBVS at 30% achieves the highest value, we employ GBVS and set l to 30 in FSI experimentally.

TABLE VI
OVERALL PREDICTION PERFORMANCE OF DIFFERENT SALIENCY MODELS AT DIFFERENT PROPORTIONS ON CSIQ

Saliency Model	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
GBVS [49]	0.9149	0.9163	**0.9175**	0.9147	0.9118	0.9101	0.9094	0.9090	0.9073	0.9069
IS [52]	0.9047	0.9076	**0.9096**	0.9074	0.9049	0.9031	0.9017	0.9027	0.9045	0.9069
Covsal [50]	0.9100	0.9135	**0.9159**	0.9144	0.9122	0.9118	0.9104	0.9091	0.9083	0.9069
SWD [61]	0.8967	0.9031	0.9082	**0.9087**	0.9078	0.9066	0.9053	0.9057	0.9065	0.9069
LRK [62]	0.9072	0.9083	0.9086	**0.9097**	0.9084	0.9065	0.9049	0.9040	0.9048	0.9069
FES [63]	0.9045	0.9088	**0.9119**	0.9106	0.9088	0.9072	0.9072	0.9070	0.9063	0.9069
RCSS [64]	0.8997	0.9064	**0.9106**	0.9081	0.9090	0.9080	0.9081	0.9074	0.9064	0.9069

V. CONCLUSION

In this paper, we have proposed a novel RR IQA metric FSI, which is based on free-energy principle and sparse representation. On one hand, the free-energy principle indicates the perception of the human brain is governed by an internal generative model, by which the brain generates predictions for its encountered scenes. The discrepancy between the visual input signal and its brain prediction is closely related to quality of perceptions. On the other hand, sparse representation resembles the strategy for representing natural images in the primary visual cortex of the brain. Conjunctively, we approximate the internal generative model with sparse representation and propose FSI accordingly. In FSI, the reference and distorted images are firstly predicted by sparse representation. Then the difference between the entropies of the prediction discrepancies is defined as the quality index. Experimental results on four large-scale image databases demonstrate FSI achieves comparative performance with PSNR and even outperforms SSIM on the CSIQ database, although it only needs one number of the original image. Moreover, for the same kind of methods, e.g., FEDM and RRED, FSI outperforms them by a large margin.

REFERENCES

[1] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

[2] X. Liu *et al.*, "Sparsity-based image error concealment via adaptive dual dictionary learning and regularization," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 782–796, Feb. 2017.

[3] X. Liu, G. Cheung, X. Wu, and D. Zhao, "Random walk graph laplacian-based smoothness prior for soft decoding of JPEG images," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 509–524, Feb. 2017.

[4] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[5] 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.

[6] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[7] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[8] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[9] 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.

[10] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.

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

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

[13] 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.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

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

[15] 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.*, Sep. 2002, vol. 1, pp. 477–480.

[16] X. Min *et al.*, "Blind quality assessment of compressed images via pseudo structural similarity," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2016, pp. 1–6.

[17] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2002, vol. 3, pp. 57–60.

[18] Y. Liu *et al.*, "Quality assessment for real out-of-focus blurred images," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 70–80, 2017.

[19] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep.-Oct., pp. 2209–2216.

[20] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480–1494, Sep. 2015.

[21] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

[22] Z. Wang *et al.*, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.

[23] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[24] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[25] 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.

[26] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

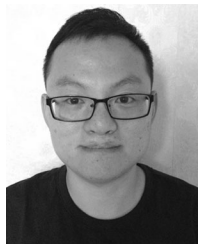
[27] Y. Xu, D. Liu, Y. Quan, and P. Le Callet, "Fractal analysis for reduced reference image quality assessment," *IEEE Trans. Image Process.*, vol. 24, no. 7, pp. 2098–2109, Jul. 2015.

[28] 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.

[29] K. Gu *et al.*, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.

[30] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[31] B. A. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
[32] B. A. Olshausen, "Principles of image representation in visual cortex," *Vis. Neurosc.*, vol. 2, pp. 1603–1615, 2003.
[33] 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.
[34] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.
[35] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcast.*, vol. 60, no. 3, pp. 555–567, Sep. 2014.
[36] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1146–1153.
[37] L. Li *et al.*, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.
[38] H.-W. Chang, H. Yang, Y. Gan, and M.-H. Wang, "Sparse feature fidelity for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 4007–4018, Oct. 2013.
[39] F. Qi, D. Zhao, and W. Gao, "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2338–2344, Dec. 2015.
[40] P. Richard, "Feynman. Statistical mechanics: A set of lectures," *Front. Phys.*, 1972.
[41] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[42] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[43] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
[44] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.
[45] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[46] J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.
[47] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
[48] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.
[49] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Int. Conf. Neural Inform. Process. Syst.*, 2006, pp. 545–552.
[50] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *J. Vis.*, vol. 13, no. 4, p. 11, 1–20, 2013.
[51] X. Min, G. Zhai, K. Gu, and X. Yang, "Fixation prediction through multimodal analysis," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 1, pp. 1–23, Oct. 2016.
[52] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, Jan. 2012.
[53] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th Eur. Workshop Vis. Inf. Process.*, 2013, pp. 106–111.
[54] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[55] Y. Zhang, A. K. Moorthy, D. M. Chandler, and A. C. Bovik, "C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes," *Signal Process., Image Commun.*, vol. 29, no. 7, pp. 725–747, 2014.
[56] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," 2016. [Online]. Available: http://live,ece.utexas.edu/research/qualit, Accessed on: July 17, 2007.
[57] E. C. Larson and D. Chandler, "Categorical image quality (CSIQ) database," 2010. [Online]. Available: http://vision.okstate.edu/csiq
[58] Y. Horita, K. Shibata, Y. Kawayoke, and Z. P. Sazzad, "MICT image quality evaluation database," 2011. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mictdb.html
[59] A. M. Rohaly *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment," ITU-T Standards Contribution COM, vol. 1, pp. 9–80, 2000.
[60] M. Elad, *Sparse and Redundant Representations*. New York, NY, USA: Springer, 2010.
[61] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 473–480.
[62] H. J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *J. Vis.*, vol. 9, no. 12, pp. 15–15, 2009.
[63] H. R. Tavakoli, E. Rahtu, and J. Heikkilä, "Fast and efficient saliency detection using sparse sampling and kernel density estimation," in *Proc. 17th Scandinavian Conf. Image Anal.*, 2011, pp. 666–675.
[64] T. N. Vikram, M. Tscherepanow, and B. Wrede, "A saliency map based on sampling an image into random rectangular regions of interest," *Pattern Recognit.*, vol. 45, no. 9, pp. 3114–3124, 2012.



Yutao Liu received the B.S. and M.S. degrees in computer science from Harbin Institute of Technology (HIT), Harbin, China, in 2011 and 2013, respectively, and is currently working toward the Ph.D. degree in the School of Computer Science and Technology, HIT.

From 2014 to 2016, he was with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China, as a Research Assistant. His current research interests include image processing, and image quality assessment.

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 Postdoctoral 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, Erlangen, Germany. His research interests include multimedia signal processing and perceptual signal processing. He was the recipient of the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012.

Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015. His research interests include quality assessment, contrast enhancement, visual saliency detection, and monitoring of air quality. He is currently the Associated Editor for the IEEE Access, and is the reviewer for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON BROADCASTING, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, IEEE ACCESS, *Information Sciences, Neurocomputing*, *Signal Processing: Image Communication*, the *Journal of Visual Communication and Image Representation*, *Digital Signal Processing*, *Multimedia Tools and Applications*, etc. He has reviewed more than 50 journal papers each year. He is the leading Special Session Organizer in VCIP2016 and ICIP2017. He was the recipient of the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016.

Xianming Liu received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology (HIT), Harbin, China, in 2006, 2008, and 2012, respectively. He is currently an Assistant Professor with the Department of Computer Science, HIT. From 2009 to 2012, he was with the National Engineering Laboratory for Video Technology, Peking University, Beijing, China, as a Research Assistant. In 2011, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Postdoctoral Fellow from 2012 to 2013. Since 2014, he has been a Postdoctoral Fellow with the National Institute of Informatics, Tokyo, Japan. His research interests include image/video coding, image/video processing, and machine learning.

Debin Zhao received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1985, 1988, and 1998, respectively. He is currently a Professor with the Department of Computer Science, Harbin Institute of Technology. He has authored or coauthored more than 200 technical articles in refereed journals and conference proceedings in the areas of image and video coding, video processing, video streaming and transmission, and pattern recognition.

Wen Gao (S'87–M'88–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He is currently a Professor of computer science with Peking University, Beijing, China. Before joining Peking University, he was a Professor of computer science with Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He has authored five books and more than 600 technical articles in refereed journals and conference proceedings in image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics. He serves the editorial board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, and the *Journal of Visual Communication and Image Representation*. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ICME and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations.