NMF-Based Image Quality Assessment Using Extreme Learning Machine

Shuigen Wang, *Student Member, IEEE*, Chenwei Deng, *Senior Member, IEEE*, Weisi Lin, *Fellow, IEEE*, Guang-Bin Huang, *Senior Member, IEEE*, and Baojun Zhao

***Abstract*—Numerous state-of-the-art perceptual image quality assessment (IQA) algorithms share a common two-stage process: distortion description followed by distortion effects pooling. As for the first stage, the distortion descriptors or measurements are expected to be effective representatives of human visual variations, while the second stage should well express the relationship among quality descriptors and the perceptual visual quality. However, most of the existing quality descriptors (e.g., luminance, contrast, and gradient) do not seem to be consistent with human perception, and the effects pooling is often done in *ad-hoc* ways. In this paper, we propose a novel full-reference IQA metric. It applies non-negative matrix factorization (NMF) to measure image degradations by making use of the parts-based representation of NMF. On the other hand, a new machine learning technique [extreme learning machine (ELM)] is employed to address the limitations of the existing pooling techniques. Compared with neural networks and support vector regression, ELM can achieve higher learning accuracy with faster learning speed. Extensive experimental results demonstrate that the proposed metric has better performance and lower computational complexity in comparison with the relevant state-of-the-art approaches.**

***Index Terms*—Extreme learning machine (ELM), human visual system (HVS), image quality assessment (IQA), non-negative matrix factorization (NMF).**

I. Introduction

DURING the past years, with the rapid proliferation of digital imaging and communication technologies, image quality assessment (IQA) has been playing an important role in a wide variety of applications, such as image acquisition, transmission, watermarking, compression, restoration, enhancement, and reproduction. The goal of IQA is to measure image quality degradation, and the IQA metrics are generally employed to evaluate and compare the performance of various processing systems and/or to optimize the parameters settings in processing. For example, the most popular structural similarity (SSIM) index by Wang *et al.* [1] has been used in image and video coding [2].

Since human visual system (HVS) is the ultimate receiver of sensory information in most cases, IQA based on subjective viewing (as defined in ITU-R Recommendation BT. 500 [3]) is the most reliable way. However, subjective evaluation is too cumbersome, time consuming, and expensive to be used in real-time and automated systems. Therefore, it is necessary to develop objective metrics to automatically measure the image quality. To this end, the research community has developed numerous IQA methods in the past decades.

According to the availability of a reference image, objective IQA metrics can be classified into full reference (FR), reduced-reference (RR), and no reference (NR) methods. In this paper, the discussion is confined to FR methods, where the original "distortion-free" image is known as the reference image.

The recent developed FR IQA schemes include SSIM [1], multiscale SSIM (MS-SSIM) [4], information content weighted SSIM (IW-SSIM) [5], gradient similarity (GSIM) [6], visual gradient similarity (VGS) [7], feature similarity (FSIM) [8], most apparent distortion (MAD) [9], visual information fidelity (VIF) [10], internal generative mechanism (IGM) [11], and additive impairment and detail loss measure (ADM) [12].

The SSIM [1] extracts luminance, contrast, and structural information assuming that HVS is highly sensitive to these distortions. It combines these three factors with a nonlinear weighted multiplier, which is the same as the MS-SSIM [4] and IW-SSIM [5]. Since gradient conveys important visual information for scene understanding, Liu *et al.* [6] applied the GSIM to measure the changes in contrast and structure, and combined luminance together for final quality prediction. The VGS model [7] fuses contrast, gradient directions, and amplitudes by intrascale pooling. Apart form these metrics, the phase congruency is used together with gradient in FSIM [8]. Larson and Chandler [9] advocated that multiple strategies should be employed in evaluating image quality by the HVS, and the distortions of near-threshold and supra-threshold (clearly visible) should be measured separately. In [10], the VIF views IQA as an information fidelity problem, and quantifies the loss of image information using Gaussian scale mixtures. In [11] and [12], IGM and ADM also use contrast and gradient as image distortion descriptors. Although these distortion measurements (e.g., luminance, contrast, and gradient) used in the aforementioned metrics are simple and

Manuscript received February 24, 2015; revised December 4, 2015; accepted December 19, 2015. Date of publication February 3, 2016; date of current version December 14, 2016. This work was supported by the National Natural Science Foundation of China under Grant 61301090. This paper was recommended by Associate Editor P. S. Sastry. *(Corresponding author: Chenwei Deng.)*

S. Wang, C. Deng, and B. Zhao are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: cwdeng@bit.edu.cn).

W. Lin is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798.

G.-B. Huang is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

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/TCYB.2015.2512852

easy to implement, they are not consistent with human perception. Therefore, more researches on effective distortion description are needed toward consistency with subjective perception.

On the other hand, the influence of each distortion factor (in terms of luminance, contrast, and gradient) should be weighted or pooled together for the overall image quality gauging. Generally speaking, there are two different strategies (i.e., nonlearning-based methods and learning-based ones) for distortion pooling. The above-mentioned metrics fuse the effects of different distortion factors using simple summations [used in peak signal-to-noise ratio (PSNR)] or weighted multiplications (used in SSIM). Such pooling techniques, however, seem to be *ad-hoc* with limited theoretical grounds. For example, a simple summation or averaging operation implicitly constraints the relationship among distortion effects and image quality score to be linear; a weighted summation requires the determination of appropriate weighting coefficients, and there is no general method available for this. Fortunately, learning-based pooling strategies have the abilities to overcome those pooling/fusion limitations. They aim to use machine learning techniques to deduce a mathematical function to model the relationship among different distortion effects and image quality. Since the required weights/parameters of the resulting function are optimized by training with subjective quality scores [e.g., mean opinion score (MOS)] provided by human observers, the human vision knowledge toward IQA is inexplicitly incorporated into the trained model. Therefore, the quality score obtained by learning-based model is expected to be more consistent with human perception.

The rest of this paper is organized as follows. Section II presents an overview of the related works on learning-based IQA metrics. In Section III, we give detailed analysis and discussion of the proposed perceptual quality metric using non-negative matrix factorization (NMF) and extreme learning machine (ELM). In Section IV, substantial experimental results and related analysis are demonstrated. Finally, the conclusion is presented in Section V.

II. Related Works

Learning-based IQA is one of the new trends for perceptual IQA. In recent few years, more and more IQA works have been developed by employing machine learning techniques for visual quality evaluation, which can be categorized into distortion-fused and model-fused metrics.

Neural networks (NNs) and support vector regression (SVR) are applied to build distortion-fused models for visual quality assessment [13]–[21]. In RR/NR metrics [13]–[15], to quantify the losses of image quality, the scene statistics of luminance or wavelet/DCT coefficients are modeled, and SVR is adopted for fusing the distortion effects. In [16] and [17], NNs are employed for FR IQA with mean values/standard deviations and luminance/contrast exploited as distortion descriptors, respectively. In [18], the image distortions are represented by 2-D mel-cepstrum, while the perceptual quality measurements in [19] are the singular vectors and values of singular value decomposition (SVD). In [20] and [21], NNs are adopted for objective video quality assessment, and energy, gradient, and frame errors are used for distortion representations. However, as analyzed in Section I, the distortion descriptors (e.g., mean values/standard deviations and luminance/contrast) used in these metrics do not agree well with final human perception. Furthermore, the two learning algorithms, NNs and SVR, may suffer from some problems, including overfitting, trial human intervention, time consuming, and local optimization [22].

With respect to the model-fused metrics, SVR is employed to develop a content-dependent multimetric fusion (CD-MMF) for measuring the image degradations in [23]. CD-MMF combines multiple objective IQA metrics (e.g., SSIM, VIF, PSNR, and FSIM) using SVR. Thus, a large number of image samples need to be collected, which are associated with subjective scores and predicted objective quality scores by different IQA metrics. However, it presents several different models for each database, and it is difficult to select one model performing well on all databases. In addition, the evaluation performances are not robust across different databases and the computational complexity is high.

In this paper, we propose a novel leaning-based FR objective IQA metric, and the contributions are in twofold.

1) NMF [24] is exploited to measure the image distortions. The property of NMF parts-based representation creates the possibilities to reflect high-level parts-based visual processing in human perception. It is believed that the high-level features of the image content would be closer to the final perception of human beings [25], [26], and thus modeling and evaluating the high-level visual contents and their corresponding distortions is expected to be more effective for IQA.
2) An emergent machine learning scheme (i.e., ELM [27]) is adopted for the distortion effects pooling. With the aids of ELM, human vision knowledge toward IQA is incorporated into the process of distortion pooling. Moreover, ELM has been demonstrated to have higher learning accuracy with much faster learning speed than NNs and SVR in numerous applications, such as face recognition [28], image segmentation [29], human action recognition [30], and NR IQA [31]. Suresh *et al.* [31] used a k-fold selection ELM (KS-ELM) for the evaluation of JPEG distortion. However, actually the KS-ELM is n random k-fold training with the original ELM algorithm [22], [32], which results in roughly n times computational complexity but similar training accuracy compared with that of ELM. In our method, to balance the performance and computational complexity, the original regularized ELM [27], [33], [34] is employed.

III. Proposed Algorithm

The block diagram of proposed method is shown in Fig. 1. It can be seen that NMF is first applied to represent the original and distorted images, respectively. And then, the resultant NMF bases are utilized to form quality similarity vectors. By using the given subjective scores, ELM is employed for fusing the similarity vectors and subjective results to predict the final perceptual image quality.

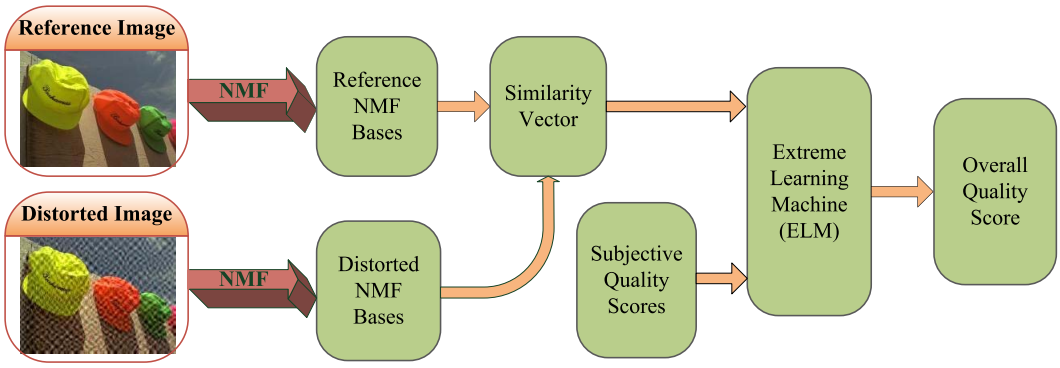

Fig. 1. Block diagram of the proposed IQA metric.

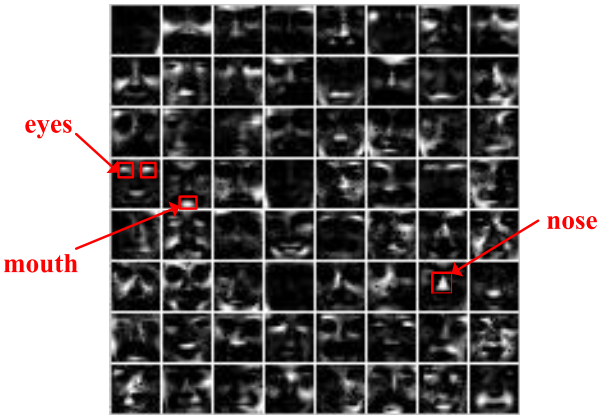

Fig. 2. Parts-based representations of CBCL facial dataset [35]. Here 64 basis images of matrix $\mathbf{W}$ in (1) are presented to show different parts of human faces, such as eyes, noses, and mouths.

A. NMF-Based Distortion Description

According to the work in [24], NMF is with parts-based representation property (as shown in Fig. 2) by using non-negativity constraints. This property has been exploited for numerous applications, such as face recognition [36]–[38] and document clustering [39], [40]. In this paper, the parts-based representation of NMF is utilized for image quality gauging.

Suppose that we have n data points $\{\mathrm{x}_i\}_{i=1}^{n}$. Each data point $\mathrm{x}_i \in \mathbb{R}^m$ is an m-dimensional vector. The vectors are placed in the columns and the whole data set is represented by a matrix $\mathbf{X} = [\mathrm{x}_i, \ldots, \mathrm{x}_n] \in \mathbb{R}^{m \times n}$. NMF aims to find two non-negative matrix factors $\mathbf{W}$ and $\mathbf{H}$ where the product of the two factors is an approximation of the original matrix [36], represented as

$$\mathbf{X} \approx \mathbf{W}\mathbf{H}^{\mathbf{T}} \tag{1}$$

where $\mathbf{W}$ is an $m \times k$ matrix and $\mathbf{H}^T$ is a $k \times n$ matrix such that $\mathbf{H}$ is an $n \times k$ matrix. The approximation is quantified by a cost function which can be constructed by some distance measurements, such as the square of the Euclidean distance (also known as the Frobenius norm) between the two matrices. The goal of NMF can then be restated as follows: to factor $\mathbf{X}$ into matrices $\mathbf{W}$ and $\mathbf{H}^T$ so that the following objective function is minimized:

$$\mathcal{O}_F = \|\mathbf{X} - \mathbf{W}\mathbf{H}^T\|^2 \tag{2}$$

where $\mathcal{O}_F$ is the error between the original matrix $\mathbf{X}$ and the factorization result $\mathbf{W}\mathbf{H}^T$. The matrix $\mathbf{W}$ denotes as bases matrix and $\mathbf{H}$ is the corresponding coefficient matrix. Since (2) is not convex in both $\mathbf{W}$ and $\mathbf{H}$, multiplicative update rules [24] are practically adopted for NMF optimization.

When NMF is used for image decomposition, each column vector of $\mathbf{W}$ (i.e., w_i) can be regarded as an image basis, which tends to represent the fundamental parts of image contents. Each data point x_i is approximated by a linear combination of these k bases w_i, weighted by the components of $\mathbf{H}$. In other words, NMF maps each m-dimensional data x_i to k-dimensional h_i, and the new representation space is spanned by the k bases w_i.

Generally speaking, k is related to the image size, image content, NMF approximation error $\mathcal{O}_F$, and computational complexity. Ideally, k should be adaptively determined for each specific image. However, finding optimal k for different image contents would be very complicated and time-consuming. Fortunately, we found that for various images, $\mathcal{O}_F$ can converge fast with a relatively small k, and the resulting bases matrix $\mathbf{W}$ is able to represent image content well. Therefore, we practically set $k \ll m$ and $k \ll n$.

An experiment was performed to demonstrate the relationship between k and $\mathcal{O}_F$. We randomly chose 100 images from each of the TID [41] and LIVE [42] databases with the image sizes of 512×384 and 768×512, respectively. Fig. 3 shows the averaging NMF convergence curves. One can see that as k increases, $\mathcal{O}_F$ converges fast. When $k \geq [15\% \times \min(m, n)]$, $\mathcal{O}_F$ changes slightly. Moreover, we also found that in this case, the IQA performance changes little. However, the computational complexity becomes higher with the increasing of k. Therefore, for quality assessment of an image with the size of $m \times n$, due to the fast convergence property of NMF, k need not be adaptively configured, and can be empirically set as a value around $15\% \times \min(m, n)$. In our experiments in Section IV, for computational convenience, we set $k = 64$ for all the images in the six databases.

Considering an $m \times n$ original image $\mathbf{X}^{(r)}$ and its corresponding distorted image $\mathbf{X}^{(d)}$, the factorization results are represented as follows:

$$\begin{cases} \mathbf{X}^{(r)} \approx \mathbf{W}^{(r)} \left(\mathbf{H}^{(r)}\right)^T \\ \mathbf{X}^{(d)} \approx \mathbf{W}^{(d)} \left(\mathbf{H}^{(d)}\right)^T \end{cases} \tag{3}$$

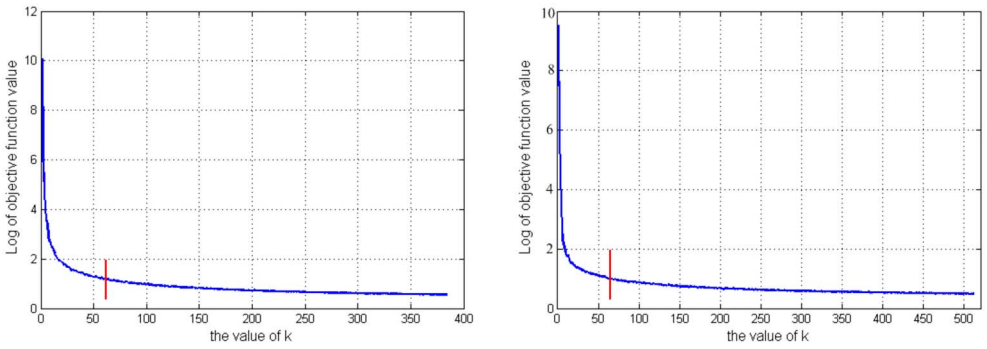

Fig. 3. Illustration of the relationship between k and the objective function value $\mathcal{O}_F$. The curves are the average values of 100 distorted images randomly chosen from (a) TID [41] and (b) LIVE [42] databases. The red solid line in each database denotes $k = 64$, which is around the value of $15\% \times \min(m, n)$.

where $\mathbf{W}^{(r)}$ and $\mathbf{W}^{(d)}$ are the $m \times k$ NMF bases matrices of the original and distorted images, respectively. Each column vector of $\mathbf{W}^{(r)}$ (or $\mathbf{W}^{(d)}$), i.e., $w_i^{(r)}$ (or $w_i^{(d)}$), can be regarded as a basis. $\mathbf{H}^{(r)}$ and $\mathbf{H}^{(d)}$ are the corresponding $n \times k$ weighting matrices. In the proposed metric, $\mathbf{W}^{(r)}$ and $\mathbf{W}^{(d)}$ are used as distortion descriptors to be further processed. One should note that, to ensure the order of $w_i^{(r)}$ corresponds to that of $w_i^{(d)}$, the initialized matrices $\mathbf{W}$ and $\mathbf{H}$, and the number of iterations should be the same for $\mathbf{X}^{(r)}$ and $\mathbf{X}^{(d)}$ in the optimization of NMF.

B. Distortion Effects Pooling Using ELM

In this section, we will present how to fuse the distortion effects of bases matrix $\mathbf{W}$ to predict the final quality score using ELM [22]. Note that the ELM theories [27], [32]–[34] have proved that random feature mapping (with almost any nonlinear activation function) can provide universal approximation capability.

1) Distortion Effects Pooling: As shown in Section III-A, NMF bases matrices $\mathbf{W}^{(r)}$ and $\mathbf{W}^{(d)}$ of the original and distorted images have been separately obtained. The changes between $\mathbf{W}^{(r)}$ and $\mathbf{W}^{(d)}$ are then measured using cosine similarity as

$$c_j = \cos(\theta_j) = \frac{\left\langle w_j^{(r)}, w_j^{(d)} \right\rangle}{\|w_j^{(r)}\| \|w_j^{(d)}\|}, \quad j = 1, 2, \ldots, k \tag{4}$$

where $w_j^{(r)}$ (and $w_j^{(d)}$) is the jth column vector of $\mathbf{W}^{(r)}$ (and $\mathbf{W}^{(d)}$) of the original image (and the corresponding distorted image). Here k is the same with that in (3), and $\langle \cdot \rangle$ denotes the dot product, $\|\cdot\|$ is the 2-norm of the vector, and θ_j is the angle between $w_j^{(r)}$ and $w_j^{(d)}$.

Mathematically, the range of c_j belongs to $[-1\ 1]$. However, since $w_j^{(r)}$ and $w_j^{(d)}$ are obtained by NMF, the entities of $w_j^{(r)}$ and $w_j^{(d)}$, i.e., $w_{ji}^{(r)}$ and $w_{ji}^{(d)}$, are non-negative values. Therefore, in this paper, c_j belongs to $[0\ 1]$, where $c_j = 1$ means that the two vectors are exactly the same, $c_j = 0$ indicates that the two vectors are independent, and in-between values represent the intermediate similarities.

The obtained similarity values c_j can form a vector $C = [c_1, c_2, \ldots, c_k]$, which is used as the final quality measurement for ELM pooling. It is worth noting that images with different levels of distortions exhibit different distributions of the vector C. Fig. 4 demonstrates this property using six different distortion types (i.e., awgn, blur, contrast change, fnoise, JPEG2000, and JPEG) and two distortion levels (denoted as "high-quality" and "low-quality") in the CSIQ database [43] (similar results are obtained on other databases). The red solid lines indicate that the distributions of C in the distorted images with higher quality (i.e., lower distortion level), while green dashed lines represent those of lower quality ones (i.e., higher distortion level). From Fig. 4(a)–(f), we can see that the distributions of C of the images with lower distortion are generally above those with higher distortion, and the distinctions are obvious. Therefore, the quality measurement C can be considered as a discriminative feature to accurately measure image degradations with different distortion levels.

Given the image quality feature C, our goal is to find a function of C to represent the quality score Q as

$$Q = f(C) \tag{5}$$

where f is a function mapping C to the final score Q, and Q can be further normalized to $[0\ 1]$. However, f is difficult to be solved in practice due to the limited knowledge and complexity of the HVS. To estimate f, in this paper, ELM [27] is adopted to learn the underlying complex relationship between C and subjective quality scores. Here subjective quality scores are used for training the parameters of the function f, and therefore, the resultant objective quality score is consistent with human perception.

2) ELM Regression: ELM is a simple and efficient learning algorithm for single-hidden layer feedforward neural networks (SLFNs) [22], [32], and it was demonstrated to have higher learning accuracy than other machine learning techniques (e.g., NNs and SVR) with much faster learning speed [22], [27], [32].

Given N arbitrary training samples (C_i, y_i), where C_i is the feature vector of the ith pair of original/distorted images and y_i is the subjective quality score of the corresponding distorted image, the goal of ELM is to find a function $f(C_i)$ which has the smallest deviation from the subjective quality score y_i for all the training data. The function $f(C_i)$ can be represented as

$$f(C_i) = \sum_{j=1}^{L} \beta_j g_j(C_i), \quad i = 1, \ldots, N \tag{6}$$

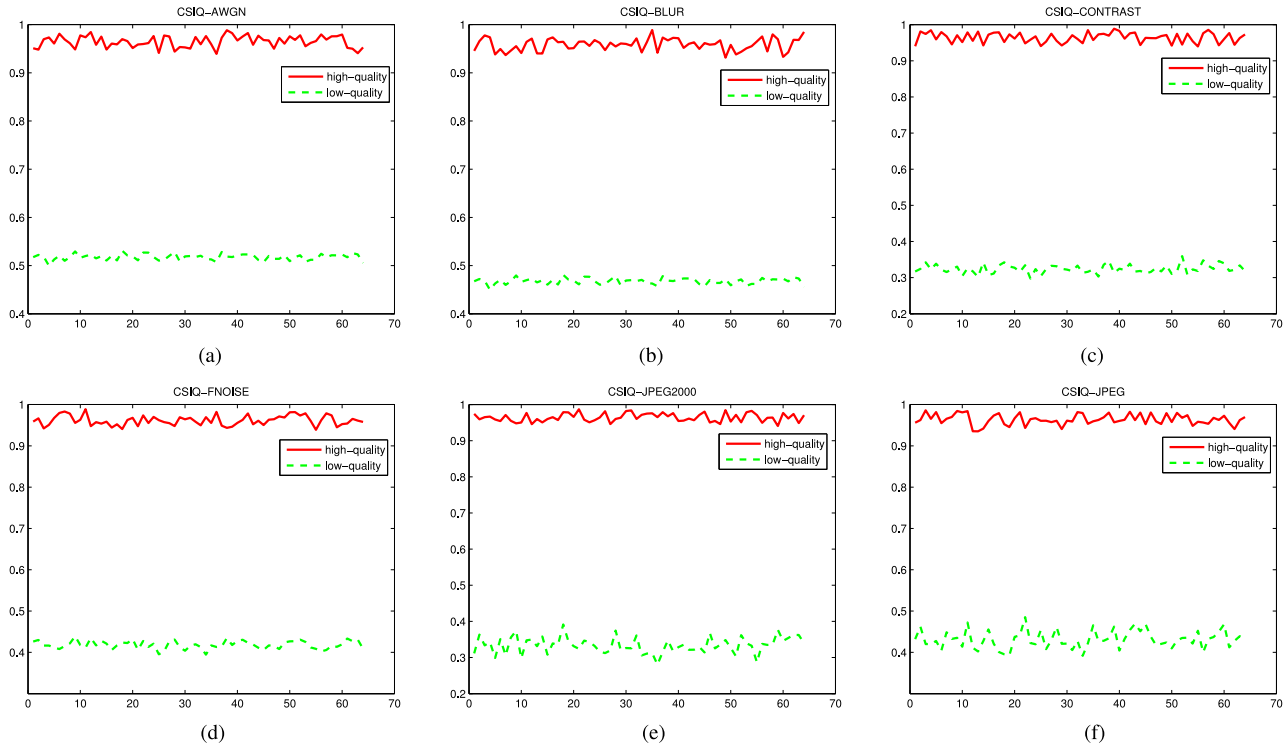

Fig. 4. Averaging feature vectors C of distorted images with six distortion types and two distortion levels from CSIQ database, and each distortion level has 30 images. The red solid lines indicate the feature vectors of high-quality images with slight distortion, while the green dashed lines represent the feature vectors of low-quality images with severe distortion. The x-axis denotes the index (i.e., j) of element c_j in C, and the y-axis denotes the value of c_j. (a) awgn. (b) Blur. (c) Contrast change. (d) fnoise. (e) JPEG2000. (f) JPEG.

where $\beta = [\beta_1, \ldots, \beta_L]^T$ is the output weighting vector, $g_j(C_i)$ is the activation function which can approximate N training samples with zero error means that $\sum_{i=1}^{N} \|f(C_i) - y_i\| = 0$, and it can be formulated as

$$g_j(C_i) = g(w_j \cdot C_i + b_j) \tag{7}$$

where w_j is the input weighting vector connecting the jth hidden node and the input nodes, b_j is the threshold of the jth hidden node. $w_j \cdot C_i$ denotes the inner product of w_j and C_i. In [22], it has been shown that the input weighting vector w_j and the bias term b_j can be randomly generated based on a continuous probability distribution, and all the hidden nodes are randomly generated as well and independent of each other. Therefore, β in (6) is the only parameter to be estimated. This is one of the reasons why ELM has fast learning speed.

For the N training samples (C_i, y_i), (6) can be written compactly as

$$\mathbf{Y_H}\beta = \mathbf{Y} \tag{8}$$

where $\mathbf{Y_H}$ is called the hidden layer output matrix of the NN, and can be shown as

$$\mathbf{Y_H} = \begin{pmatrix} g(w_1 \cdot C_1 + b_1) & \ldots & g(w_L \cdot C_1 + b_L) \\ \vdots & \vdots & \vdots \\ g(w_1 \cdot C_N + b_1) & \ldots & g(w_L \cdot C_N + b_L) \end{pmatrix}_{N \times L} \tag{9}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_L \end{pmatrix}_{L \times 1}, \mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}_{N \times 1} \tag{10}$$

where L is the number of hidden nodes in the SLFNs. In order to minimize the norm of the output weights, minimal norm least square method is employed in ELM instead of the standard optimization methods [22], [32]. Thus, the output weights vector β is estimated analytically as

$$\beta = \mathbf{Y}\mathbf{Y}_\mathbf{H}^\dagger \tag{11}$$

where $\mathbf{Y}_\mathbf{H}^\dagger$ is the Moore–Penrose generalized pseudo-inverse of the hidden layer output matrix $\mathbf{Y_H}$ whose ith column is the ith hidden node output with respect to inputs $C_1, \ldots, C_N$. Bartlett [44] has claimed that the smaller the norm weights are, the better generalization performance the networks tend to have for feedforward NNs. By adopting the Moore–Penrose generalized pseudo-inverse, $\beta = \mathbf{Y}\mathbf{Y}_\mathbf{H}^\dagger$ has the smallest norm among all the optimization solutions, and this is the reason why ELM has better generalization performance/higher learning accuracy than those of NNs and SVR [22].

Practically, the orthogonal projection method [45] can be efficiently used to calculate the Moore–Penrose inverse: $\mathbf{Y}_\mathbf{H}^\dagger = (\mathbf{Y}_\mathbf{H}^T\mathbf{Y_H})^{-1}\mathbf{Y}_\mathbf{H}^T$, if $\mathbf{Y}_\mathbf{H}^T\mathbf{Y_H}$ is nonsingular; or $\mathbf{Y}_\mathbf{H}^\dagger = \mathbf{Y}_\mathbf{H}^T(\mathbf{Y}_\mathbf{H}^T\mathbf{Y_H})^{-1}$, if $\mathbf{Y_H}\mathbf{Y}_\mathbf{H}^T$ is nonsingular. According to the ridge regression theory [46], it is suggested to add a positive value $(1/\lambda)$ to the diagonal of $\mathbf{Y}_\mathbf{H}^T\mathbf{Y_H}$ or $\mathbf{Y_H}\mathbf{Y}_\mathbf{H}^T$. By doing so, the resultant solution is more stable and tends to have better generalization performance than basic ELM. With the positive value $(1/\lambda)$, we can have

$$\beta = \mathbf{Y}_\mathbf{H}^T\left(\frac{\mathbf{I}}{\lambda} + \mathbf{Y_H}\mathbf{Y}_\mathbf{H}^T\right)^{-1}\mathbf{Y}, \text{ if } N \le L \tag{12}$$

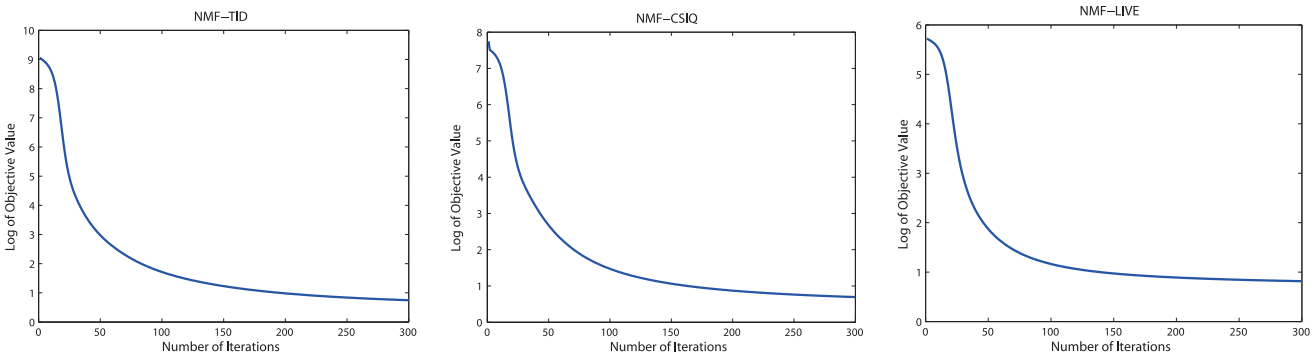

Fig. 5. Demonstration of NMF Convergence. The tested reference images are *I*1.bmp, 1600.png, and buildings.bmp from (a) TID, (b) CSIQ, and (c) LIVE databases, respectively.

where N is the number of training samples and L is the number of hidden nodes. When the number of training samples N is huge and larger than that of nodes L, then we can have

$$\beta = \left(\frac{\mathbf{I}}{\lambda} + \mathbf{Y}_{\mathbf{H}}^T \mathbf{Y}_{\mathbf{H}}\right)^{-1} \mathbf{Y}_{\mathbf{H}}^T \mathbf{Y}, \text{ if } N > L. \tag{13}$$

C. Computational Complexity Analysis

For the proposed metric, the computational complexity is mainly dominated by the factorization of images and the distortion pooling process. In this section, we will discuss the complexity from these two aspects.

The goal of NMF is to find two non-negative matrices $\mathbf{W}$ and $\mathbf{H}$ to minimize the objective function in (2). However, the objective function $\mathcal{O}_F$ is not convex in both variables $\mathbf{W}$ and $\mathbf{H}$. It is thus difficult to find the global minimum for $\mathcal{O}_F$. Lee and Seung [24, Sec. III-A, Algorithm 1] proposed an iterative update to find the locally optimal solution for the minimization problem. Therefore, the affecting factor is the convergence rate, which is determined by the matrix size (m and n) and the variable k in (1). It has been found that the larger size of the original matrix is, the slower NMF converges; the larger the variable k is, the slower NMF converges. From Fig. 3, we have known that $\mathcal{O}_F$ can converge fast with a small k, and for computational convenience, k can be empirically set as a value around $15\% \times \min(m, n)$. As for the relationship between NMF convergence and iteration times, three experiments on TID [41], CSIQ [43], and LIVE [42] databases were conducted to figure out the relationship between the iterating times and the approximation error [i.e., $\mathcal{O}_F$ in (2)]. Fig. 5 shows the convergence rate of NMF on the three image databases. It can be noted that NMF converges very fast. For all databases, NMF converges within 100 iterations. We can also see that the objective function value $\mathcal{O}_F$ of 50 iterations is similar with that of 100 iterations. Furthermore, the learning accuracy is almost the same, while the time cost for 50 iterations is much less than that for 100 iterations. Therefore, we set the number of iterations as 50 in our experiments.

As for ELM, from Section III-B2, one can note that the learning time of ELM is mainly due to the calculation of the output weighting vector β. The hidden nodes, input weights w_j and hidden layer biases b_j of ELM are all randomly generated based on a continuous probability distribution without fine tuning which is a necessary process for SVR/NNs. This is an important difference from other learning algorithms. For SVR, some additional parameters (e.g., kernel parameters ρ, C, and ϵ in [19]) need to be tuned, and this induces heavy computation. Moreover, it has been demonstrated that the generalization performance of ELM is stable on a wide range of hidden nodes [22], [27], and only a small number of hidden nodes are needed for training a large number of samples, which can save much training time with better generalization performance.

Based on the analysis above, we can see that the computational complexity of the proposed scheme would be low. In the next section, we will give the execution time of our algorithm comparing with other schemes on the public image databases.

IV. Experimental Results and Analysis

In this section, extensive experimental results are presented to evaluate the overall accuracy of the proposed IQA metric, in comparison with the existing relevant state-of-the-art metrics, including PSNR, MS-SSIM [4], VIF [10], IW-SSIM [5], MAD [9], FSIM [8], GSIM [6], VGS [7], IGM [11], ADM [12], SVDR [19], and CD-MMF [23]. In addition, the performances on the individual distortion types are then demonstrated. A cross database validation is conducted for further testing the robustness of the proposed metric compared with the other two learning-based metrics (i.e., SVDR and CD-MMF). Two additional experiments are conducted to further confirm the effectiveness of NMF-based distortion measurements and ELM-based distortion pooling technique, respectively. Finally, average execution time comparison results are provided to verify the efficiency of proposed model.

A. Databases and Evaluation Criteria

In this paper, six publicly available and subject-rated benchmark databases are used, including TID [41], CSIQ [43], LIVE [42], IVC [47], MICT [48], and A57 [49]. The number of distorted images is 1700, 866, 779, 185,

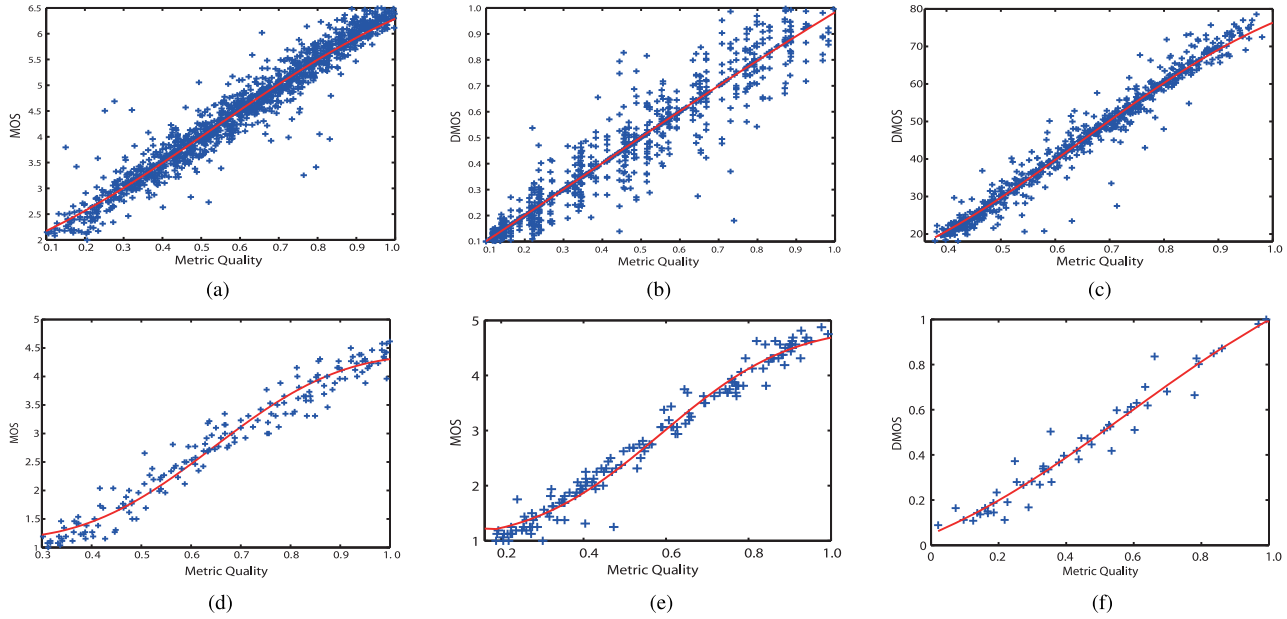

Fig. 6. Scatter plots of subjective scores versus the proposed metric scores on IQA databases. (a) TID. (b) CSIQ. (c) LIVE. (d) IVC. (e) MICT. (f) A57.

168, and 54, respectively. The number of distortion types for each database is 17, 6, 5, 5, 2, and 6.

To evaluate the performance of the IQA metrics on a common analysis space, a five-parameter logistic mapping between the objective outputs $\mathbf{s}_o$ and the subjective scores is adopted to nonlinearly regress the objective scores $\mathbf{s}_o$

$$\mathbf{s}_r = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(\mathbf{s}_o - \beta_3))} \right) + \beta_4 \, \mathbf{s}_o + \beta_5 \qquad (14)$$

where $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ are the parameters to be fitted by minimizing the sum of squared errors between the mapped values $\mathbf{s}_r$ and subjective scores.

The mapped values $\mathbf{s}_r$ are then compared with the ground truth, i.e., MOS or differential MOS (DMOS), based on five evaluation criteria: 1) Spearman rank-order correlation coefficient (SRCC); 2) Kendall rank-order correlation coefficient (KRCC); 3) Pearson linear correlation coefficient (PLCC); 4) root mean squared error (RMSE); and 5) outlier ratio (OR). The SRCC and KRCC can measure the prediction monotonicity and the other three can assess the prediction accuracy [50]. A better objective IQA measurement has higher SRCC, KRCC, and PLCC, while lower RMSE and OR values. As for a perfect match between the mapped objective scores and the subjective scores, SRCC = KRCC = PLCC = 1 and RMSE = OR = 0.

B. Test Procedure

The k-fold cross validation strategy [51] is widely used in the verification of learning-based IQA metrics [13]–[19]. Since the proposed metric is a learning-based model, in this paper, we also employ k-fold cross validation for the performance and robustness testing.

As for the k-fold cross validation, the whole set of data need to be split into k equal or roughly equal chunks. Each chunk is tested separately, and the averaging value of the k testing chunks is regarded as the final result of the whole dataset. Generally, the k-fold validation contains two embedded loops.

1) For the testing of one chunk, the remaining $(k-1)$ chunks are used for training, and the testing result can be obtained by the trained model. Moreover, to achieve more stable and convincing performance, the training/testing processes should be repeated for several times (1000 times in our simulation), and the averaging value of the 1000 testing scores is used as the testing result of the said chunk.
2) Each of the k chunks is used as the testing chunk successively, and then goes back to step 1).

The splitting of the data into k chunks is done carefully so that the image contents (one image content is defined as all the distorted versions of an original image) presented in one chunk do not appear in any other chunks, and the learning results are affected by the chunk size k. If k is too large, the number of training data would be too large and it may cause under-fitting; and similarly, over-fitting occurs when k is too small. Generally, k is often set as 5–10 [19]. In our simulations, as an example, the TID database has 25 original images, and in this case, the images are empirically split into eight chunks, with each of the first seven chunks containing three image contents, while the last chunk includes four image contents. In the same way, the CSIQ database with 30 original images is split into ten chunks with each chunk consisting of three image contents. The LIVE database with 29 original images is also split into ten chunks with each of the first nine chunks consisting of three image contents while two image contents for the last chunk. Similar splitting procedure is followed for other databases.

Furthermore, in the experiments, unipolar sigmoidal function $(1/(1+e^{-\lambda u}))$ is used as the ELM activation function with $\lambda = 0.1$. We use the same number of hidden nodes as 200 for six databases.

TABLE I
PERFORMANCE COMPARISON OF IQA METRICS ON SIX BENCHMARK DATABASES

DB	Criteria	Proposed	CD-MMF	SVDR	VGS	IGM	FSIM	ADM	GSIM	MAD	IW-SSIM	VIF	MS-SSIM	PSNR
TID (1700)	SRCC	**0.9466**	**0.9422**	0.7771	0.9022	0.8902	0.8805	0.8617	0.8554	0.8340	0.8559	0.7496	0.8528	0.5245
	KRCC	**0.8915**	**0.8864**	0.5841	0.8467	0.7104	0.6946	0.6842	0.6651	0.6445	0.6636	0.5863	0.6543	0.3696
	PLCC	**0.9513**	**0.9476**	0.7889	0.9095	0.8858	0.8738	0.8690	0.8462	0.8306	0.8579	0.8090	0.8425	0.5309
	RMSE	**0.4185**	**0.4289**	0.8246	0.5577	0.6228	0.6525	0.6620	0.7151	0.7474	0.6895	0.7888	0.7299	1.1372
CSIQ (866)	SRCC	**0.9727**	**0.9668**	0.8618	0.9662	0.9401	0.9242	0.9334	0.9126	0.9467	0.9213	0.9193	0.9138	0.8057
	KRCC	**0.8342**	**0.8266**	0.6870	0.8258	0.7872	0.7567	0.7716	0.7403	0.7970	0.7529	0.7534	0.7397	0.6080
	PLCC	**0.9763**	0.9675	0.8875	**0.9692**	0.9280	0.9120	0.9280	0.8979	0.9502	0.9144	0.9277	0.8998	0.8001
	RMSE	**0.0571**	0.0674	0.1210	**0.0647**	0.0978	0.1077	0.0980	0.1156	0.0818	0.1063	0.0980	0.1145	0.1575
	OR	**0.1121**	0.1638	0.2794	**0.1619**	0.2217	0.2252	0.2180	0.2298	0.1801	0.2460	0.2263	0.2448	0.3430
LIVE (799)	SRCC	**0.9760**	**0.9805**	0.8791	0.9696	0.9580	0.9634	0.9542	0.9554	0.9669	0.9567	0.9631	0.9445	0.8755
	KRCC	**0.8526**	**0.8574**	0.7178	0.8463	0.8319	0.8337	0.8228	0.8131	0.8421	0.8175	0.8270	0.7922	0.6864
	PLCC	**0.9758**	**0.9802**	0.8788	0.9686	0.9578	0.9597	0.9360	0.9437	0.9674	0.9522	0.9598	0.9430	0.8721
	RMSE	**5.8659**	**5.4134**	7.6861	6.7910	7.9248	7.6780	9.6270	9.0376	6.9235	8.3470	7.6734	9.0956	13.3680
	OR	**0.3227**	**0.3015**	0.6902	0.3498	0.3954	0.3933	0.5390	0.4069	0.4146	0.5310	0.5456	0.6187	0.6826
IVC (185)	SRCC	**0.9468**	**0.9382**	0.8796	0.9221	0.9027	0.9262	0.9030	0.9294	0.9146	0.9125	0.8966	0.8847	0.6885
	KRCC	**0.7956**	**0.7876**	0.6925	0.7548	0.7288	0.7564	0.7255	0.7626	0.7406	0.7339	0.7165	0.7012	0.5220
	PLCC	**0.9531**	**0.9453**	0.8829	0.9319	0.9129	0.9376	0.9130	0.9399	0.9210	0.9231	0.9028	0.8934	0.7199
	RMSE	**0.3870**	**0.3956**	0.5721	0.4420	0.4973	0.4236	0.4960	0.4160	0.4747	0.4686	0.5239	0.5474	0.8456
MICT (168)	SRCC	**0.9508**	**0.9411**	0.8620	0.8905	0.8910	0.9059	0.9370	0.9233	0.9362	0.9202	0.9086	0.8864	0.6130
	KRCC	**0.8442**	**0.8264**	0.6716	0.7011	0.7093	0.7302	0.7903	0.7541	0.7823	0.7537	0.7029	0.6413	0.4447
	PLCC	**0.9553**	**0.9456**	0.8668	0.8953	0.8990	0.9252	0.9420	0.9287	0.9405	0.9248	0.9144	0.8935	0.6426
	RMSE	**0.3709**	**0.4071**	0.6241	0.5575	0.5780	0.5248	0.4210	0.4640	0.4251	0.4761	0.5066	0.5621	0.9588
	OR	**0.0283**	**0.0318**	0.0774	0.0728	0.0714	0.0476	0.0710	0.0357	0.0714	0.0408	0.0536	0.0833	0.2202
A57 (54)	SRCC	**0.9554**	**0.9498**	0.8533	0.9118	0.8859	0.9181	0.8725	0.9002	–	0.8709	0.6223	0.8394	0.6189
	KRCC	**0.8492**	**0.8413**	0.6804	0.7628	0.7191	0.7639	0.6912	0.7205	–	0.6842	0.4589	0.6478	0.4309
	PLCC	**0.9501**	**0.9411**	0.8846	0.9299	0.9188	0.9078	0.8803	0.8976	–	0.9034	0.6158	0.8504	0.6587
	RMSE	**0.0713**	**0.0831**	0.1146	0.0904	0.0970	0.0933	0.1166	0.1084	–	0.1050	0.1936	0.1293	0.1849
Weighted Mean	SRCC	**0.9554**	**0.9372**	0.8278	0.9316	0.9165	0.9121	0.9011	0.8960	0.8974	0.8963	0.8369	0.8874	0.6759
	KRCC	**0.8557**	**0.8394**	0.6463	0.8295	0.7546	0.7445	0.7370	0.7218	0.7335	0.7218	0.6777	0.7054	0.5017
	PLCC	**0.9568**	**0.9379**	0.8398	0.9363	0.9156	0.9059	0.9003	0.8865	0.8973	0.8967	0.8652	0.8802	0.6805
SRCC	# best	**5**	1	0	0	0	0	0	0	0	0	0	0	0
KRCC	# best	**5**	1	0	0	0	0	0	0	0	0	0	0	0
PLCC	# best	**5**	1	0	0	0	0	0	0	0	0	0	0	0

C. Overall Performance Comparison

In order to make a comprehensive analysis on the proposed metric, a performance comparison is presented with the existing IQA metrics on the overall distortions of the six public image databases. Fig. 6 shows the scatter plots of the evaluation results of the proposed scheme, which demonstrates the high consistency between the proposed metric and the subjective evaluation results (i.e., MOS/DMOS). Moreover, the values of SRCC, KRCC, PLCC, RMSE, and OR of the proposed scheme and other 12 IQA schemes are listed in Table I, where the two best schemes of each database are highlighted in boldface. Note that since the standard deviations among the subjects are not released in TID, IVC, and A57, the OR values are not computed for these databases.

From Table I, it can be seen that the proposed metric has good performances across all the databases. It is the most consistent metric on TID, CSIQ, IVC, MICT, and A57 databases in terms of all criteria like SRCC and PLCC. The proposed IQA method achieves exciting performances that the PLCC values are 0.9513, 0.9763, and 0.9758 for TID, CSIQ, and LIVE, respectively. Furthermore, to provide an overall indication of the comparative performance of different schemes, the weighted mean values of SRCC, KRCC, and PLCC are listed in Table I. Our proposed scheme also achieves the best performance in terms of the weighted mean values. In the last three rows of the table, we count the times of best performance for each metric on SRCC, KRCC, and PLCC. The proposed metric achieves 5 best, 5 best, and 5 best in terms of SRCC, KRCC, and PLCC, and 1 best, 1 best, and 1 best for CD-MMF out of six databases, respectively. For other metrics, there is no "best" performance yet.

D. Performance on Individual Distortion Type

To further analyze the performance of the proposed scheme, we provide a comparison between the proposed metric and other 12 metrics on different distortion types. The SRCC criterion is used since the other criteria lead to similar conclusion. The TID, CSIQ, and LIVE databases are selected because they are the three biggest databases among all six image databases. Table II shows the experimental results with the two best metrics highlighted in boldface for each distortion type. The definitions of the entries in the last three row of the table are listed as follows.

1) *"# best":* The number of distortion types that a metric performs best on SRCC.
2) *"# worst":* The number of distortion types that a metric performs worst on SRCC.
3) *"X Versus Proposed":* The score of winning times of a metric versus the proposed on SRCC.

From Table II, we can see that the proposed metric performs the best on **16** distortion types [the second best is 7 for VGS, and 0 for SVDR and CD-MMF (poor robustness)]. The proposed metric achieves the best results on awgn and blur

TABLE II
SRCC VALUES OF DIFFERENT IQA METRICS FOR EACH DISTORTION TYPE

DB	distortion types	Proposed	CD-MMF	SVDR	VGS	IGM	FSIM	ADM	GSIM	MAD	IW-SSIM	VIF	MS-SSIM	PSNR
TID	awgn	**0.9237**	0.9055	0.7600	0.8870	0.9069	0.8566	0.8630	0.8577	0.8388	0.8028	0.8799	0.8094	**0.9114**
	awgn-color	**0.9402**	0.8790	0.7203	0.8970	0.8947	0.8527	0.8390	0.8091	0.8258	0.8015	0.8785	0.8064	**0.9068**
	spatial corr-noise	0.8842	0.9138	0.7875	0.9010	**0.9152**	0.8483	0.8980	0.8907	0.8678	0.7909	0.8703	0.8195	**0.9229**
	masked noise	**0.9143**	0.8156	0.6363	0.7860	0.7968	0.8021	0.7360	0.7409	0.7336	0.8068	**0.8698**	0.8155	0.8487
	high-fre-noise	0.8967	**0.9341**	0.8638	**0.9410**	0.9223	0.9093	0.8970	0.8936	0.8864	0.8732	0.9075	0.8685	0.9323
	impulse noise	**0.9152**	0.8608	0.6630	0.7410	0.8160	0.7452	0.5120	0.7229	0.6499	0.6579	0.8331	0.6868	**0.9177**
	quantization noise	**0.8992**	**0.8980**	0.8130	0.8770	0.8788	0.8564	0.8500	0.8752	0.8160	0.8182	0.7956	0.8537	0.8699
	gblur	**0.9696**	0.9495	0.8120	0.904	**0.9682**	0.9472	0.9140	0.9589	0.9197	0.9580	0.9546	0.9607	0.8682
	denoising	0.9164	**0.9708**	0.8893	0.9660	0.9704	0.9603	0.9450	**0.9724**	0.9434	0.9463	0.9189	0.9571	0.9381
	jpg-comp	**0.9596**	0.9594	0.8855	**0.9710**	0.9484	0.9279	0.9410	0.9392	0.9275	0.9181	0.9170	0.9348	0.9011
	jpg2k-comp	0.9412	**0.9839**	0.9027	0.9790	**0.9845**	0.9773	0.9720	0.9759	0.9707	0.9749	0.9713	0.9736	0.8300
	jpg-trans-error	**0.9234**	0.8596	0.8347	0.8400	0.8635	0.8708	0.8510	**0.8835**	0.8661	0.8560	0.8582	0.8736	0.7665
	jpg2k-trans-error	**0.9148**	**0.9011**	0.7928	0.8440	0.8893	0.8544	0.8400	0.8925	0.8394	0.8313	0.8510	0.8525	0.7765
	pattern-noise	**0.9796**	0.7921	0.6600	0.8140	0.7295	0.7491	**0.8380**	0.7372	0.8287	0.7719	0.7608	0.7336	0.5931
	block-distortion	0.8436	0.7797	0.8013	0.7910	0.7902	**0.8492**	0.1610	**0.8862**	0.7970	0.7889	0.8320	0.7617	0.5852
	mean shift	**0.9056**	0.5229	0.5152	0.5560	0.4887	0.6720	0.5890	0.7170	0.5161	0.6757	0.5132	**0.7374**	0.6974
	contrast	0.8346	**0.8512**	0.4360	**0.8820**	0.6411	0.6481	0.4920	0.6737	0.2723	0.6273	0.8190	0.6400	0.6126
CSIQ	awgn	**0.9687**	**0.9628**	0.8298	0.9570	0.9638	0.9262	0.9583	0.9440	0.9600	0.9380	0.9571	0.9471	0.9363
	jpg-comp	**0.9755**	0.9532	0.9267	**0.9870**	0.9663	0.9654	0.9660	0.9632	0.9660	0.9662	0.9705	0.9622	0.8882
	jpg2k-comp	0.9641	0.9765	0.9381	**0.9850**	**0.9774**	0.9685	0.9748	0.9648	0.9770	0.9683	0.9672	0.9691	0.9363
	1/f noise	**0.9568**	0.9537	0.8666	0.9520	0.9427	0.9234	0.9488	0.9387	**0.9540**	0.9059	0.9509	0.9330	0.9338
	blur	**0.9839**	0.9668	0.9367	**0.9820**	0.9724	0.9729	0.9726	0.9589	0.9660	0.9782	0.9747	0.9720	0.9289
	contrast	**0.9804**	0.9473	0.3479	**0.9610**	0.9546	0.9420	0.9508	0.9508	0.9170	0.9539	0.9361	0.9521	0.8622
LIVE	jpg2k-comp	**0.9788**	0.9703	0.9425	**0.9790**	0.9675	0.9717	0.9711	0.9587	0.9380	0.9751	0.9654	0.9683	0.9551
	jpg-comp	0.9817	0.9809	0.9436	**0.9870**	0.9810	0.9834	0.9790	0.9098	0.9490	0.9645	0.9793	**0.9842**	0.9657
	awgn	**0.9924**	0.9866	0.8349	**0.9920**	0.9874	0.9652	0.9820	0.9774	0.9710	0.9667	0.9731	0.9845	0.9785
	blur	**0.9776**	0.9721	0.9123	0.9700	0.9538	0.9708	0.9650	0.9517	0.8990	0.9719	0.9584	**0.9722**	0.9413
	jpg2k-trans-error	**0.9700**	**0.9689**	0.9028	0.9600	0.9194	0.9499	0.9519	0.9399	0.8830	0.9442	0.9321	0.9652	0.9027
	# best	**16**	0	0	7	1	0	0	2	0	0	0	0	2
	# worst	**0**	**0**	12	**0**	1	**0**	2	1	4	**0**	1	**0**	7
	X vs. Proposed	**–**	3:25	0:28	10:18	5:23	6:22	5:23	5:23	3:25	3:25	4:24	4:24	4:24

TABLE III
PLCC VALUES OF THE CROSS DATABASE VALIDATION (FOR SAME IMAGES APPEARING IN THE TRAINING AND TESTING SETS, "–" IS REPRESENTED IN THE TABLE)

Test database/ Model		TID	CSIQ	LIVE	IVC	MICT	A57
TID	Proposed	–	**0.9265**	0.9593	**0.9350**	**0.9276**	**0.9485**
	SVDR [20]	–	0.8831	0.8862	0.8755	0.8558	0.8854
	CD-MMF [24]	–	0.9127	**0.9644**	0.9277	0.9132	0.9454
CSIQ	Proposed	**0.8608**	–	0.9107	**0.9324**	**0.9404**	**0.9029**
	SVDR [20]	0.7550	–	0.9086	0.8828	0.8327	0.8843
	CD-MMF [24]	0.8536	–	**0.9109**	0.9279	0.9353	0.8817
LIVE	Proposed	**0.8994**	**0.9777**	–	**0.9335**	**0.9404**	**0.9051**
	SVDR [20]	0.7428	0.8581	–	0.8877	0.8526	0.8807
	CD-MMF [24]	0.8984	0.9710	–	0.9300	0.9374	0.8839

TABLE IV
PLCC COMPARISONS OF NINE IQA METRICS USING THREE DIFFERENT DISTORTION DESCRIPTORS AND THREE DISTORTION POOLING METHODS ON SIX DATABASES

Test Database/ Metrics	TID	CSIQ	LIVE	IVC	MICT	A57
SVD-NNs	0.7835	0.8808	0.8716	0.8758	0.8600	0.8774
MMF-NNs	0.9396	0.9605	**0.9742**	0.9398	0.9386	0.9353
NMF-NNs	**0.9438**	**0.9656**	0.9658	**0.9439**	**0.9446**	**0.9415**
SVD-SVR	0.7889	0.8875	0.8788	0.8829	0.8668	0.8846
MMF-SVR	0.9476	0.9675	**0.9802**	0.9453	0.9456	0.9411
NMF-SVR	**0.9498**	**0.9726**	0.9715	**0.9496**	**0.9508**	**0.9473**
SVD-ELM	0.7928	0.8923	0.8811	0.8885	0.8704	0.8873
MMF-ELM	0.9490	0.9696	**0.9824**	0.9473	0.9476	0.9439
NMF-ELM	**0.9513**	**0.9764**	0.9756	**0.9528**	**0.9551**	**0.9505**

across the three big databases. On the TID database, the proposed approach achieves the best performance on nine types of distortions, such as awgn, awgn-color, masked noise, quantization noise, gblur, and pattern-noise. Moreover, it has good performance on pattern-noise, mean shift and contrast change, which are difficult to be evaluated for other metrics like SVDR, ADM, and MAD. Similar performances on CSIQ and LIVE are demonstrated in Table II.

E. Cross Database Validation

To test the generality and robustness of the proposed approach, a cross database evaluation is conducted. The three biggest image databases, namely, TID (1700 images), CSIQ (866 images), and LIVE (799 images), are selected as training data, since they have larger number of distorted images than the other three databases and cover most distortion types. The PLCC performance is shown in Table III, since the works in [19] and [23] only demonstrate PLCC results. SRCC, KRCC, and RMSE show similar results as PLCC.

As can be seen in Table III, the proposed metric performs well across all the databases. And we compared it with two other learning-based schemes (i.e., SVDR and CD-MMF). Our proposed metric outperforms them. It obtains the best performance on TID, CSIQ, IVC, MICT, and A57 databases for all training models. For the TID database, the proposed metric gives PLCC values of 0.8608 and 0.8994 with CSIQ and LIVE model, respectively, which are better than other metrics like CD-MMF, SVDR (learning-based), ADM, GSIM, and MAD (nonlearning-based). This is significant, since in this case, the training set (866 images) is only half of the size of the testing set (1700 images).

TABLE V
AVERAGE EXECUTION TIME (IN SECONDS PER IMAGE) FOR DIFFERENT METRICS

Metrics	Nonlearning-based							Learning-based		
	PSNR	GSIM [7]	MS-SSIM [5]	VGS [8]	IW-SSIM [6]	VIF [11]	MAD [10]	SVDR [20]	CD-MMF [24]	Proposed
Time(s)	0.0037	0.0873	0.1673	0.656	1.57	3.4829	4.1351	1.03	66.40	**0.632**

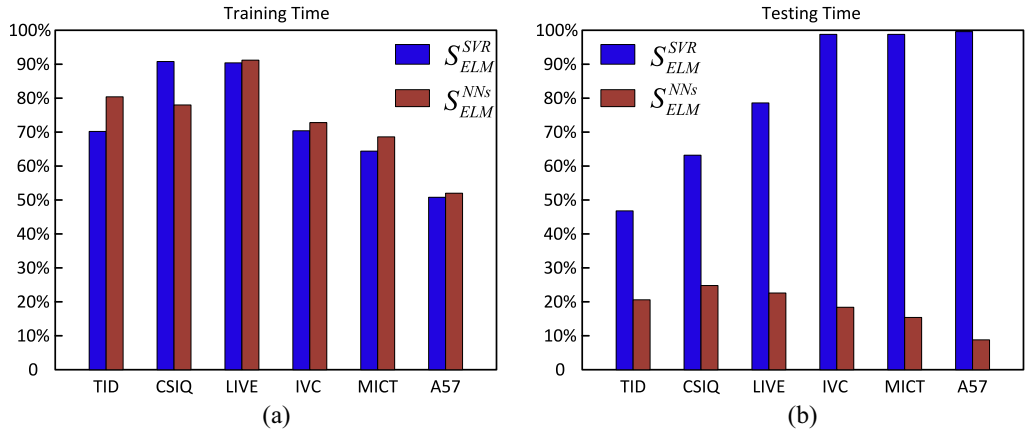

Fig. 7. Percentages of execution time saved by ELM compared with SVR and NNs. (a) Time saved in training process, i.e., $S_{\text{ELM}}^{\text{SVR}} = (t_{\text{SVR}_{\text{train}}} - t_{\text{ELM}_{\text{train}}})/t_{\text{SVR}_{\text{train}}}$ and $S_{\text{ELM}}^{\text{NNs}} = (t_{\text{NNs}_{\text{train}}} - t_{\text{ELM}_{\text{train}}})/t_{\text{NNs}_{\text{train}}}$. (b) Time saved in testing process, i.e., $S_{\text{ELM}}^{\text{SVR}} = (t_{\text{SVR}_{\text{test}}} - t_{\text{ELM}_{\text{test}}})/t_{\text{SVR}_{\text{test}}}$ and $S_{\text{ELM}}^{\text{NNs}} = (t_{\text{NNs}_{\text{test}}} - t_{\text{ELM}_{\text{test}}})/t_{\text{NNs}_{\text{test}}}$.

F. Comparison of Various Distortion Descriptors and Pooling Techniques

To further illustrate the effectiveness of the proposed NMF-based distortion descriptors (i.e., bases matrix **W**) and the advantages of ELM-based distortion pooling technique, we conducted the following additional comparisons. For distortion descriptors, we compared the proposed NMF-based one with other two commonly used ones, i.e., traditional SSIM-like (used in [23]) and SVD-based ones (used in [19]); for distortion pooling, the proposed ELM is verified with other two learning techniques, i.e., NNs and SVR. And therefore, nine IQA metrics are built and tested, including NMF-NNs, NMF-SVR, NMF-ELM, MMF-NNs, MMF-SVR, MMF-ELM, SVD-NNs, SVD-SVR [19], and SVD-ELM. Note that MMF-NNs, MMF-SVR, and MMF-ELM use the same features in CD-MMF [23], but fused by different learning techniques. For fair comparison, all the experimental environments and settings remain the same.

Table IV shows the comparison results of the aforementioned nine metrics. We first compared the performances of metrics using different distortion descriptors, but fused by the same pooling techniques. And therefore, the above-mentioned nine metrics are categorized into three groups: 1) SVD-NNs, MMF-NNs, and NMF-NNs; 2) SVD-SVR, MMF-SVR, and NMF-SVR; and 3) SVD-ELM, MMF-ELM, and NMF-ELM. One can see that the proposed feature (i.e., bases matrix **W**) achieves the best performances on five out of the six databases, i.e., TID, CSIQ, IVC, MICT, and A57.

Similarly, the performances of different distortion pooling methods are verified using the following groups of metrics: 1) SVD-NNs, SVD-SVR, and SVD-ELM; 2) MMF-NNs, MMF-SVR, and MMF-ELM; and 3) NMF-NNs, NMF-SVR, and NMF-ELM. From Table IV, one can also see that ELM-based methods perform better than NNs-based and SVR-based ones across all the six databases. It should be mentioned that, combining NMF-based feature with ELM-based pooling, the proposed metric achieves the best performance.

G. Comparison of Computational Complexity

In this section, we are to demonstrate the computational complexity of the proposed metric. First, the training/testing time of ELM, NNs, and SVR is compared, and Fig. 7 demonstrates the comparison results. In our experiment, we implemented the metrics using NMF-based features with three different learning techniques, ELM, SVR, and NNs. It shows that ELM has faster speed than SVR and NNs, in both training and testing. For example, in the CSIQ database, ELM-based methods can save 90.8% (88.2%) training time, and 63.3% (24.6%) testing time compared with SVR-based ones (NNs-based ones). The reasons are that fewer hidden nodes are needed for ELM, and the tuning for hidden parameters is not required.

Apart from the efficiency of distortion pooling methods, the average execution time of existing IQA models and proposed metric is further measured for each image in A57 [49] database (image resolution is 512×512) on a PC with 2.40-GHz Intel Core2 CPU and 2 GB of RAM. Table V shows the average execution time in seconds, with all codes being implemented in MATLAB. The metrics are categorized into nonlearning-based and learning-based ones. It is shown that the proposed scheme takes less time than VGS, IW-SSIM, VIF, and MAD. In particular, the execution time for MAD is more than 50 times of that for proposed metric. The PSNR, MS-SSIM, and GSIM (similar to MS-SSIM) take less time than proposed model, because they evaluate image degradations based on simple operations of pixels or gradient operators, however, as observed from

Tables I and III, these methods achieve much worse performance than the proposed method. As for the learning-based metrics, the proposed scheme is the fastest one comparing with SVDR and CD-MMF. The CD-MMF takes the longest time with 66.40 s, which is more than 100 times of the proposed metric.

V. Conclusion

IQA has been an important issue in various applications. In this paper, to address the existing drawbacks in distortion description/representation and distortion effects pooling of IQA, we have proposed an ELM-based IQA metric with distortions measured by NMF. Due to the non-negative property of the bases matrices, NMF can be considered as the parts-based representations of image scenes, which are more consistent with the human perception. The use of ELM exploits the advantages of machine learning to generate an effective mapping of distortion effects into overall quality scores. Experimental results on six publicly available image databases confirm that the proposed metric is highly consistent with the subjective testing scores. The performances on individual distortion types and cross-database validation demonstrate the effectiveness and robustness of the proposed scheme. Moreover, simulation results and comparisons of computational complexity with other methods show that the proposed method is computationally efficient.

References

[1] 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.

[2] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Rate bounds on SSIM index of quantized images," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1624–1639, Sep. 2008.

[3] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, IEEE Standard ITU-R BT.500-11, Int. Telecommun. Union, Geneva, Switzerland, Jun. 2002.

[4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2003, pp. 1398–1402.

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

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

[7] J. Zhu and N. Wang, "Image quality assessment by visual gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 919–933, Mar. 2012.

[8] 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.

[9] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, 2010, Art. ID 011006.

[10] 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.

[11] 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.

[12] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 935–949, Oct. 2011.

[13] 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.

[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 and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Proc. SPIE Conf. Human Vis. Electron. Imag.*, vol. 5666. San Jose, CA, USA, Jan. 2005, pp. 149–159.

[16] A. Bouzerdoum, A. Havstad, and A. Beghdadi, "Image quality assessment using a neural network approach," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol*, Rome, Italy, 2004, pp. 330–333.

[17] P. Carrai, I. Heynderickz, P. Gastaldo, R. Zunino, and P. Monza, "Image quality assessment by using neural networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 5. Scottsdale, AZ, USA, 2002, pp. V-253–V-256.

[18] M. Narwaria, W. Lin, and A. E. Cetin, "Scalable image quality assessment with 2D mel-cepstrum and machine learning approach," *Pattern Recognit.*, vol. 45, no. 1, pp. 299–313, 2012.

[19] M. Narwaria and W. Lin, "SVD-based quality metric for image and video using machine learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 347–364, Apr. 2012.

[20] P. Gastaldo, S. Rovetta, and R. Zunino, "Objective quality assessment of MPEG-2 video streams by using CBP neural networks," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 939–947, Jul. 2002.

[21] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1316–1327, Sep. 2006.

[22] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.

[23] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–1807, May 2013.

[24] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[25] E. A. Buffalo, P. Fries, R. Landman, H. Liang, and R. Desimone, "A backward progression of attentional effects in the ventral stream," *Proc. Nat. Acad. Sci.*, vol. 107, no. 1, pp. 361–365, 2010.

[26] B. C. Motter, "Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli," *J. Neurophysiol.*, vol. 70, no. 3, pp. 909–919, Sep. 1993.

[27] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[28] K. Choi, K.-A. Toh, and H. Byun, "Incremental face recognition for large-scale social network services," *Pattern Recognit.*, vol. 45, no. 8, pp. 2868–2883, 2012.

[29] C. Pan, D. S. Park, Y. Yang, and H. M. Yoo, "Leukocyte image segmentation by visual attention and extreme learning machine," *Neural Comput. Appl.*, vol. 21, no. 6, pp. 1217–1227, Sep. 2012.

[30] R. Minhas, A. A. Mohammed, and Q. M. J. Wu, "Incremental learning in human action recognition based on snippets," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 11, pp. 1529–1541, Nov. 2012.

[31] S. Suresh, R. V. Babu, and H. J. Kim, "No-reference image quality assessment using modified extreme learning machine classifier," *Appl. Soft Comput.*, vol. 9, no. 2, pp. 541–552, 2009.

[32] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.

[33] G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, pp. 3056–3062, Oct. 2007.

[34] G.-B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, pp. 3460–3468, Oct. 2008.

[35] MIT Center for Biological and Computation Learning. (2000). *CBCL Face Database #1*. [Online]. Available: http://www.ai.mit.edu/projects/cbcl

[36] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1299–1311, Jul. 2012.

[37] W.-S. Zheng, J. Lai, S. Liao, and R. He, "Extracting non-negative basis images using pixel dispersion penalty," *Pattern Recognit.*, vol. 45, no. 8, pp. 2912–2926, 2012.

[38] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Kauai, HI, USA, 2011, pp. 207–212.

[39] W. Xu, X. Liu, and Y. Gong, "Document clustering based on nonnegative matrix factorization," in *Proc. Annu. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Toronto, ON, Canada, 2003, pp. 267–273.

[40] F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons, "Document clustering using nonnegative matrix factorization," *Inf. Process. Manag.*, vol. 42, no. 2, pp. 373–386, 2006.
[41] N. Ponomarenko. (2008). *Tampere Image Database 2008 TID2008*. [Online]. Available: http://www.ponomarenko.info/tid2008.htm
[42] H. R. Sheikh *et al.* (2004). *MICT Image and Video Quality Assessment Research at Live*. [Online]. Available: http://live.ece.utexas.edu/research/quality/
[43] E. C. Larson and D. M. Chandler. (2010). *Categorical Image Quality (CSIQ) Database*. [Online]. Available: http://vision.okstate.edu/csiq
[44] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 525–536, Mar. 1998.
[45] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*, vol. 7. New York, NY, USA: Wiley, 1971.
[46] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
[47] A. Ninassi, P. L. Callet, and F. Autrusseau, "Pseudo non-reference image quality metric using perceptual data hiding," in *Proc. SPIE Conf. Human Vis. Electron. Imag.*, vol. 6057. San Jose, CA, USA, Jan. 2006, Art. ID 60570G.
[48] Y. Horita. (2010). *Image Quality Evaluation Database*. [Online]. Available: http://mict.eng.u-toyama.ac.jp/database_toyama/
[49] D. M. Chandler and S. S. Hemami. (2007). *VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images*. [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html
[50] Video Quality Expert Group (VQEG). (2003). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment II*. [Online]. Available: http://www.vqeg.org/
[51] P. Bartlett, S. Boucheron, and G. Lugosi, "Model selection and error estimation," *J. Mach. Learn.*, vol. 48, nos. 1–3, pp. 85–113, Jul. 2002.

Shuigen Wang (S'13) received the B.Sc. degree from the School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China, in 2012. He is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering, Beijing Institute of Technology, Beijing, China.

Since 2015, he has been with the LIVE Laboratory of Prof. Bovik, University of Texas at Austin, Austin, TX, USA, for visiting scholarship. His current research interests include image quality assessment, perceptual modeling, feature learning, and extraction.

Chenwei Deng (M'09–SM'15) received the Ph.D. degree in signal and information processing from the Beijing Institute of Technology, Beijing, China, in 2009.

He was a Post-Doctoral Research Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore. Since 2012, he has been an Associate Professor and then a Full Professor with the School of Information and Electronics, Beijing Institute of Technology. He has authored or co-authored over 50 technical papers in refereed international journals and conferences, and co-edited one book. His current research interests include video coding, quality assessment, perceptual modeling, feature representation, object recognition, and tracking.

Weisi Lin (M'92–SM'98–F'16) received the B.Sc. and M.Sc. degrees from Zhongshan University, Guangzhou, China, and the Ph.D. degree from King's College, London University, London, U.K.

He was the Laboratory Head of Visual Processing, and the Acting Department Manager of Media Processing with the Institute for Infocomm Research, Singapore. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He has authored over 200 refereed papers in international journals and conferences. His current research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.

Dr. Lin served as the Lead Guest Editor of a Special Issue on Perceptual Signal Processing of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING in 2012. He is the Chair of the IEEE multimedia communications technical committee (MMTC) in IEEE Communications Society Special Interest Group on Quality of Experience. He has been elected as a Distinguished Lecturer of Asia Pacific Signal and Information Processing Association 2012/13. He was the Lead Technical Program Chair of the Pacific-Rim Conference on Multimedia in 2012, and the Technical Program Chair of the IEEE International Conference on Multimedia and Expo in 2013. He is on the Editorial Board of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*. He is a Chartered Engineer, U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.

Guang-Bin Huang (M'98–SM'04) received the B.Sc. degree in applied mathematics and the M.Eng. degree in computer engineering from Northeastern University, Shenyang, China, in 1991 and 1994, respectively, and the Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 1999. During undergraduate period, he also concurrently studied in the Applied Mathematics Department and the Wireless Communication Department, Northeastern University, Boston, MA, USA.

From 1998 to 2001, he was a Research Fellow with the Singapore Institute of Manufacturing Technology (formerly known as the Gintic Institute of Manufacturing Technology), where he led/implemented several key industrial projects (e.g., a Chief Designer and the Technical Leader of Singapore Changi Airport Cargo Terminal Upgrading Project, etc.). Since 2001, he has been an Assistant Professor and an Associate Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University. His current research interests include machine learning, computational intelligence, and extreme learning machines.

Dr. Huang serves as an Associate Editor of *Neurocomputing*, *Neural Networks*, and *Cognitive Computation*. He serves as an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS.

Baojun Zhao received the Ph.D. degree in electromagnetic measurement technology and equipment from the Harbin Institute of Technology, Harbin, China, in 1996.

From 1996 to 1998, he was a Post-Doctoral Fellow with the Beijing Institute of Technology, Beijing, China, where he is currently a Full Professor, the Vice Director of Laboratory and Equipment Management, and the Director of the National Signal Acquisition and Processing Professional Laboratory. He has authored or co-authored over 100 publications. His current research interests include image/video coding, image recognition, infrared/laser signal processing, and parallel signal processing.

Prof. Zhao was a recipient of five provincial/ministerial-level scientific and technological progress awards in his research fields.