IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 20, 2019, accepted February 13, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901063

Blind Image Quality Assessment With Joint Entropy Degradation

WEIPING JI[1], JINJIAN WU[1], MAN ZHANG[1], ZUOZHI LIU[2], GUANGMING SHI[1], AND XUEMEI XIE[1]

[1]School of Electronic Engineering, Xidian University, Xi'an 710071, China
[2]School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China

Corresponding author: Jinjian Wu (jinjian.wu@mail.xidian.edu.cn)

This work was supported in part by the NSF of China under Grant 61772388, Grant 61632019, Grant 61621005, and Grant 61472301, and in part by the Natural Science Basic Research Plan in Shanxi Province under Grant 2017JM6055.

ABSTRACT Blind image quality assessment (BIQA) aims to evaluate the quality of an image without pristine image objectively, which is highly desired in many perception-oriented image processing systems. Distortions degrade the visual contents and cause image quality degradation. Moreover, the visual contents of an image suffer from individual degradations by different types and different levels of distortions, which makes us difficult to analyze the quality degradation. From the perspective of information theory, there is a decrease in the amount of the visual contents when images are distorted. Therefore, in this paper, the image quality is assessed through its visual entropy degradation. Researches on the neuroscience indicate that the simple cells in the local receptive field can be characterized as being spatially localized and oriented, then the local intensity, gradient, and orientation features are extracted to represent the visual contents of an image. By deducing the joint entropy equation, the joint entropy is related to the statistical distributions. Next, in order to measure the visual entropy, the joint statistical distributions of those features are calculated. Finally, by measuring the degradations on these distributions of distorted images, a novel BIQA method is proposed. The experimental results on the databases of LIVE, CSIQ, and TID2013 show that the proposed method has superior performance than other state-of-the-art BIQA methods.

INDEX TERMS Blind image quality assessment (BIQA), local receptive field, joint visual entropy, joint probability distribution.

I. INTRODUCTION

Digital images are ubiquitous in our daily life. However, images inevitably suffer from distortions during image processing [1], [2]. The image quality will decline which affects people's subjective perception. Therefore, it is essential to assess its perceived quality in image communication and processing. The quality of image assessed by human is usually time-consuming and expensive [3]. Hence, objective image quality assessment (IQA) which can automatically predict image quality consistently with human subjective perception has attracted a lot of attentions.

In the past two decades, a large amount of IQA methods have been introduced. Objective IQA methods are usually divided into three categories: full reference (FR) models, reduced reference (RR) models, and no reference (NR) [4] models. Most of these methods are FR (e.g., structure similarity [5]) and RR (e.g., orientation selectivity based visual pattern for RR IQA [6] and reduced-reference IQA with reference with visual information fidelity [7]) models. However, those methods require full reference images or certain descriptors of the reference images. In most practical situations, the reference images are not often available. Thus, no reference (NR) IQA that requires no information about the primary has became an active research topic in recent years [8]–[10].

Without the help and guidance of the reference, NR IQA is a more difficult problem. Early NR methods mainly used the prior knowledge of the distortion type for image quality prediction. Those methods are usually called distortion-specific NR IQA [3], [11]. The distortion-specific features are extracted for quality prediction in those NR IQA methods, such as the algorithm in [12] and [13] for blur, that in [14] for JPEG2000 compressed images, and that in [15] for

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Zhao.

JPEG compressed images. These methods based on distortion type only work for a certain type of distortion, and have a limitation in practice.

Recently, in order to assess the quality of images without the prior knowledge of distortion, the non-distortion-specific NR IQA methods have been studied gradually [16]–[18]. Most existing algorithms belong to knowledge-driven methods, such as that relying on human visual system (HVS) [19]or Natural Scene Statistical (NSS) [20]. HVS-based methods are also called bottom-up methods, which achieve the image quality by modeling some characteristics of HVS, or some physiology and psychophysics experiments. Although those methods have been found nearly universal acceptance, they have a lot of limitations. Actually, the HVS is a highly non-linear and complex system, but most existing HVS-based models of BIQA rely on linear or quasi-linear operators [5].

NSS-based models evaluate image quality based on the statistical distributions of certain filter responses in several different domains, such as spatial, wavelet, and DCT domains. Moorthy et al. studied the natural scene statistical of images, and obtained the image quality by measuring the changes on the generalized Gaussian distribution (GGD) coefficients in the wavelet domain, which is called DIIVINE [21]. Assuming that the statistics of the contrast and structure features extracted in the DCT domain can vary in a predictable way as the image quality changes, a BIQA model was proposed, which is called BLIINDS [22]. Moreover, Mittal et al. quantified possible losses of naturalness in the images by using scene statistic of locally normalized luminance coefficients, and introduced the BRISQUE for quality assessment [8]. Recently, Zhang *et al.* [23] proposed the IL-NIQE for BIQA which combined a series of NSS features in several domains. Though those methods have made great progress, there still exists a large gap between the objective method and the human subjective perception.

Distortions destroy the visual contents of images, and then degrade the subjective perception of human. Moreover, the visual contents of an image suffer from individual degradations by different types and different levels of distortions [24], which makes us difficult to analyze the quality degradation. However, distortions always degrade certain visual contents of images. From the perspective of information theory, there is a decrease in the amount of the visual contents when images are distorted. Therefore, the image quality can be measured by its visual information degradation. According to Shannon information theory, the amount of information that the source conveys to the outside can be defined by the information entropy. In other words, the quality degradation varies as the visual entropy changes.

In order to analyze and measure visual information, the representations of the visual contents are required. Inspired by the researches on the neuroscience, the local receptive field (LRF) in the primary visual cortex is highly adaptive to extract the local feature [25]. Moreover, the simple cells in the LRF can be characterized as being spatially localized and oriented [25]. In other words, the visual contents of an image can be characterized as the features delivering the spatial location and orientation of the image. Since these representations do not appear alone, the joint representations are used to describe the visual contents of the image. Thus, the quality degradation is analyzed as the joint entropy of degraded features representing visual information in images.

In this work, the image quality is assessed by measuring the joint visual entropy of images. Distortions decrease the visual information, and the amount of visual information in an image can be measured by information entropy. Moreover, the researches on the neuroscience indicate that the visual information of an image can be represented by a series of features. Thus, the degraded quality is analyzed as the changes of visual feature entropy.

The main contributions of our model are as follows: firstly, the image quality is assessed from a completely new perspective, the image quality degradation can be measured by the degraded visual information. Secondly, by analyzing the degradation of visual information, a novel BIQA method based on joint entropy degradation is proposed. Finally, the proposed method has a strong robustness, that is it achieves good performances in across-dataset evaluation.

The paper is organized as follows: In Section II, we give the details of joint visual entropy analysis. Section III describes the application of joint visual entropy in BIQA modeling. Experimental evaluations of the proposed methods on three databases compared with other state-of-the-art methods are presented in Section IV. We conclude the paper in Section V.

II. JOINT VISUAL ENTROPY ANALYSIS

Distortions degrade the visual information of an image, and cause the image quality degradation. Moreover, the visual contents of an image suffer from individual degradations by different types and different levels of distortions, which makes us difficult to analyze the image quality degradation. As can be seen in Fig. 1, the less distortions, the more visual contents contain. In other words, the distortion of an image can be evaluated by the visual content. From the perspective of information theory, the amount of visual information can be measured by the visual entropy. Recently, researches on the neuroscience indicate that the visual information can be represented by the features that describe the spatial location and orientation of the image [25]. Since the representations are expressed together, their joint visual

FIGURE 1. An example of a natural scene distorted by different distortions.

entropy can be used to represent the visual information of an image. Thus, the joint entropy of those visual features can efficiently represent the visual information degradations, and the image quality is predicted with its joint entropy degradation.

Based on the above analysis, the joint entropy degradation of visual information can be used to measure the image quality, and the visual information of an image is represented by a series of features. A hypothesis is that a set of features can be expressed by $X_1, X_2, \ldots, X_n$. According to the Shannon information theory, the joint entropy of these features is calculated as:

$$\begin{aligned} H(X_1, X_2, \ldots, H_n) &= \sum_{i=1}^{n} H(X_i|H_{i-1}, \ldots, H_1) \\ &= \frac{1}{n}\sum_{i=1}^{n} H(X_i) \\ &\quad + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} H(X_j|X_i) + \ldots \\ &\quad + \frac{1}{n!}\sum_{i=1}^{n}\ldots\sum_{k=1, k\neq i..q}^{n} H(X_k|X_q..X_i) \quad (1) \end{aligned}$$

As mentioned above, the joint visual entropy is related to single feature entropies and other conditional entropies (i.e. $H(X_i)$, $H(X_j|X_i)$,..., $H(X_k|X_q, \ldots, X_j, X_i)$). Moreover, according to the Shannon entropy equation, the entropy only relies on statistical distribution, which is shown in Eq. 2,

$$H(X) = -\sum_{i} p(x_i) log p(x_i). \quad (2)$$

Thus, $H(X_i) \propto p(X_i)$ (which also stands for the other parts, i.e., $H(X_j|X_i) \propto p(X_j|X_i)$,..., $H(X_k|X_q, \ldots, X_j, X_i) \propto p(X_k|X_q, \ldots, X_j, X_i)$). In other words, the statistical distributions of the image will change when an image is distorted, which results in entropy degradation. Thus, the image quality varies with the changes on each statistical distributions ($p(X_i)$, $p(X_j|X_i)$,...,$p(X_k|X_q, \ldots, X_j, X_i)$). Therefore, in order to assess the image quality with the visual information degradation, the statistical distributions of those features which represent the visual information of an image are required.

III. BIQA MODELING

In this section, the extraction of those features that represent the visual information of an image is presented firstly. Then, the statistical distributions of features are obtained. Finally, the degradations on statistical distributions of features are analyzed for BIQA modeling.

A. FEATURE EXTRACTION

Distortions degrade the visual information of an image. Moreover, different types and different levels of distortions generate individual degradations on visual contents. From the perspective of information theory, the amount of visual information that an image contains can be measured by the visual information entropy. In order to measure and analyze the visual information, the representations of the visual information of an image are required. Although there are many pixels in a discrete natural image, each pixel is highly correlated to its surrounding [24]. In other words, there a large amount of redundancies when the distribution of each pixel is calculated for visual entropy measurement directly. Inspired by the researches on the neuroscience, the LRF in the primary visual cortex is highly adaptive to extract the local feature for visual perception, and the simple cells in the LRF can be characterized as being spatially localized and oriented. In other words, the LRF is extremely sensitive to the changes of intensity and orientation [25], [26]. Thus, the visual information of an image can be represented by the local features such as local intensity, local orientation and local gradient of the image.

1) LOCAL INTENSITY EXTRACTION

As mentioned above, the simple cells in the LRF are extremely sensitive to the changes on intensity [25], [26]. The pixel values in an image change directly when an image is distorted. Thus, the intensity of images can partially represent the visual information of images. There are a large amount of redundancies when we calculate the distribution of each pixel, and each pixel in an image is highly correlated to its surrounding. Moreover, researches on natural scene statistical indicate that a local non-linear normalization to the luminance has efficient decorrelation function [8], [27]. Thus, a local non-linear normalization is adopted to represent the intensity of an image. For a given image I, the local luminance is normalized via local mean subtraction and divisive normalization. Such an operation is applied to a intensity image $I(i, j)$ to produce:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C}, \quad (3)$$

where, $i \in 1, 2, \ldots, M, j \in 1, 2, \ldots, N$ are spatial indices, and M, N are the height and width of an image, respectively. $C = 1$ is a constant that prevents instabilities from occurring when the $\sigma(i, j)$ tends to zero. And the $\mu(i, j)$ and $\sigma(i, j)$ are defined as:

$$\mu(i, j) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I_{k,l}(i, j), \quad (4)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}(I_{k,l}(i, j) - \mu(i, j))^2}, \quad (5)$$

where $w = \{w_{k,l}|k = -K, \ldots, K, l = -L, \ldots, L\}$ is a 2D circularly-symmetric Gaussian weighting function (i.e. $G(x, y|\sigma) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$) sampled out to 3 standard deviations and rescaled to unit volume. Since the smaller window size the better performance [8], we set the $K = L = 3$ in this implementation.

2) LOCAL GRADIENT EXTRACTION

According to the researches on neuroscience, the simple cells in the LRF are highly adapted to extract orientation features. Gradient is a better way to show the changes on the orientation. Thus, the gradient of an image is adopted to represent the visual information of an image partially. For a digital image, the gradient is usually computed by convolving an image with a linear filter such as the classic Roberts, Sobel, and Prewitt filters or some task-specific ones [28], [29]. Moreover, the gradient magnitude is defined as the root mean square of image directional gradients along two orthogonal directions. In this work, we adopt a Gaussian difference filter pair along the horizontal and vertical directions to calculate the gradient of an image in two orthogonal directions [30]. The horizontal direction filter operator $h_x(x, y|\sigma)$ is defined as:

$$\begin{aligned} h_x(x, y|\sigma) &= \frac{\partial}{\partial x} g(x, y|\sigma) \\ &= -\frac{1}{2\pi\sigma^2}\frac{x}{\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2}), \end{aligned} \tag{6}$$

and the vertical direction filter operator $h_y(x, y|\sigma)$ is defined as:

$$\begin{aligned} h_y(x, y|\sigma) &= \frac{\partial}{\partial y} g(x, y|\sigma) \\ &= -\frac{1}{2\pi\sigma^2}\frac{y}{\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2}), \end{aligned} \tag{7}$$

where, the $g(x, y|\sigma) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$ is isotropic Gaussian function, σ is the scale parameter. In this work, we define the scale parameter $\sigma = \frac{5}{6}$, and the size of windows is $5 * 5$. Thus, the gradient of an image is calculated as:

$$G = \sqrt{(I * h_x)^2 + (I * h_y)^2}, \tag{8}$$

3) LOCAL ORIENTATION SELECTIVITY EXTRACTION

Researches on the neuroscience indicate the simple cells in the LRF are highly sensitive to orientation and location. And orientation selectivity arises from the spatial arrangement of intracortical responses in a LRF of the primary visual cortex [31], [32]. When an image is perceived, the individual arrangements of excitatory/inhibitory interactions for different local receptive fields are excited. In other words, different kinds of orientation selectivity based visual patterns (OSVP) are generated for image understanding [33]–[35]. Therefore, the OSVP features are extracted for visual information presentation by imitating this orientation selectivity mechanism. In [34], the research indicates that the orientation selectivity is directly related to the arrangement of the interaction among cortical neurons in a LRF. Inspired by this mechanism and the arrangement of the correlations among neighbor pixels, the OSVP can be computed.

The OSVP of the pixel $x \in I$ in an image is defined as:

$$\mathcal{P}_v(x|\chi) = \mathcal{A}(\mathcal{I}(x|\chi)) = \mathcal{A}(\mathcal{I}(x|x_1, x_2, \ldots, x_n)), \tag{9}$$

where $\chi = \{x_1, x_2, \ldots, x_n\}$ is the arrangement of the spatial correlations with its circularly symmetric neighborhood, $\mathcal{A}(\cdot)$ is the arrangement of spatial correlations, and $\mathcal{I}(x|\chi)$ is the spatial correlations between x and x_i in $\mathcal{X}$. The OSVP feature of an pixel depends on the arrangement of intracortical responses (i.e.,excitatory and inhibitory interactions). According to these researches that neighbor neurons with similar preferred orientations always present excitatory interactions and these dissimilar ones present inhibitory interactions, the description of the pattern $\mathcal{P}_v$ can be simplified as the arrangement of interactions between the central pixel x and its local neighbors $\chi = \{x_1, x_2, \ldots, x_n\}$. Thus, the Eq. 9 can be recognized as:

$$\mathcal{P}_v(x|\chi) \approx \mathcal{A}(\mathcal{I}(x|x_1), \mathcal{I}(x|x_2), \ldots, \mathcal{I}(x|x_n)), \tag{10}$$

where $\mathcal{I}(x|x_i), i \in 1, 2, \ldots, n$, denotes the interaction between x and x_i.

There are two types of interactions in imitating the orientation selectivity mechanism, excitation and inhibition, which play distinct roles, one for excitatory neurons connecting to neurons that are well correlated in activity, and the other inhibitory neurons connecting to neurons that are anti-correlated. Moreover, the interaction type in the orientation selectivity mechanism is depended on the orientation similarity, which is defined as:

$$\mathcal{I}(x|x_i) = f(\Theta(x), \Theta(x_i)), \tag{11}$$

Since the interaction type depends on their preferred orientations, the orientation similarity between x and x_i is computed to represent $\mathcal{I}(x|x_i)$. In order to obtain the orientation similarity, the gradient direction θ of a pixel $x \in I$ is defined as its orientation:

$$\theta(x) = arctan\frac{G_v(x)}{G_h(x)}, \tag{12}$$

where G_h and G_v are the horizontal and vertical gradient magnitudes, respectively, which can be calculated as:

$$G_h = I * f_h, G_v = I * f_v, \tag{13}$$

where f_h is the horizontal Prewitt filter, f_v is the vertical Prewitt filter, and the convolutional operation is denoted as $*$. The Prewitt filters are defined as:

$$f_h = \frac{1}{3}\begin{Bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{Bmatrix}, \quad f_v = \frac{1}{3}\begin{Bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{Bmatrix}, \tag{14}$$

According to the orientation similarity between two pixels, the Eq. 11 is rewritten as:

$$\mathcal{I}(x|x_i) = \begin{cases} 1, & \text{if } |\theta(x) - \theta(x_i)| < \mathcal{T} \\ 0, & \text{otherwise,} \end{cases} \tag{15}$$

where excitatory interaction is represented as $'1'$, $'0'$ represents inhibitory interaction, and the similarity threshold is defined as $\mathcal{T}$. In order to obtain the similarity threshold, the subjective viewing test on visual masking effect [36] has been investigated. The researches indicate that the masking effects among nearby gratings are strong if they possess the same orientation. The masking effect becomes marginal when

the orientation difference is larger than the threshold. Thus, in this paper, we set $\mathcal{T} = 6°$.

According to Eq. 15, the OSVP form of a pixel is represented by the arrangement of $'0'$ and $'1'$ within its circularly symmetric neighborhood. In this work, the orientation similarity of the 8-neighborhood $\mathcal{X}$ centered on x organized counterclockwise to obtain the local orientation information of x, which is defined as B,

$$B(x) = (\mathcal{I}(x|x_1), \mathcal{I}(x|x_2), \ldots, \mathcal{I}(x|x_8)), \tag{16}$$

Moreover, going through each pixel in the image and obtaining its local orientation information, the orientational information of the whole image is obtained, which is defined as:

$$O(I) = \begin{bmatrix} B(1,1) & B(1,2) & \ldots & B(1,N) \\ B(2,1) & B(2,2) & \ldots & B(2,N) \\ \ldots & \ldots & \ldots & \ldots \\ B(M,1) & B(M,2) & \ldots & B(M,N) \end{bmatrix} \tag{17}$$

where, M, N are the height and width of an image.

B. FEATURE DISTRIBUTION CALCULATION

Distortions degrade the visual information of an image and lead to the image quality degradation. The amount of visual information in an image can be measured with information entropy. Researches on neuroscience indicate that the visual information can be represented by a series of features. As mentioned above, three features are extracted to represent the visual information of an image. Thus, according to the Eq. 1, the visual information entropy can be rewritten as:

$$\begin{aligned} H(X_1, X_2, X_3) &= \sum_{i=1}^{3} H(X_i|H_{i-1}, \ldots, H_1) \\ &= \frac{1}{3}\sum_{i=1}^{3} H(X_i) + \frac{1}{6}\sum_{i=1}^{3}\sum_{j=1, j\neq i}^{3} H(X_j|X_i) \\ &\quad + \frac{1}{6}\sum_{i=1}^{3}\sum_{j=1, j\neq i}^{3}\sum_{k=1, k\neq i,j}^{3} H(X_k|X_j, X_i). \end{aligned} \tag{18}$$

According to the shannon theory, the event entropy is related to the statistical distribution of the event. According to the Eq. 2, the $H(X_1, X_2, X_3)$ depends on the statistical distribution of each event (i.e., $p(X_i)$, $p(X_j|X_i)$, $p(X_k|X_j, X_i)$).

For analyzing the statistical distribution of each event, the distribution probabilities are required. With the Eq. 3, the intensity of an image is obtained. Since the intensity of an image is continuous, the values of the normalized luminance are discretized and divided into M evenly spaced bins. In this work, we set $M = 23$ by considering the accuracy and the complexity of computation. Moreover, the values of the normalized luminance are specified in $1, 2, \ldots., M$ based on the principle of rounding. Then, the statistical probabilities of the normalized luminance feature are acquired based on counting the number of each bin, e.g., $\{p(X_1) = m | m = 1, 2, \ldots, 23\}$. Similar to the values of the normalized luminance, the values of the gradient in an image based on Eq. 8 are continuous. Thus, the values are discretized and divided into N evenly spaced bins based on same mode, and the $N = 23$ is set. The statistical probabilities of gradient are obtained by counting the number of each bin, i.e.,$\{p(X_2 = n)|n = 1, 2, \ldots, 23\}$. Different from the above two types of values, the values of OSVP are discrete. Even so, there are too many types of OSVPs resulting in computational complexity. Considering the characteristic of rotation invariant, there are only 36 fundamental types of patterns are reserved. Finally, by counting the number of every value, the statistical probabilities of OSVP are acquired, i.e., $\{p(X_3 = k)|k = 1, 2, \ldots, 36\}$. According to the above operations, the representations of visual information on images are discretized and the statistical probabilities of visual features are obtained.

Moreover, the joint statistical distribution of the features is also counted. For each pixel in an image, the local intensity and the local gradient are calculated according to the Eq. 3 and Eq. 8. Then, the local intensity and the local gradient of the whole image are obtained, which is define as $IG(I)$,

$$\begin{bmatrix} \hat{I}(1,1), G(1,1) & \ldots & \hat{I}(1,N), G(1,N) \\ \hat{I}(2,1), G(2,1) & \ldots & \hat{I}(2,N), G(2,N) \\ \ldots & \ldots & \\ \hat{I}(M,1), G(M,1) & \ldots & \hat{I}(M,N), G(M,N) \end{bmatrix}, \tag{19}$$

where, M, N are the height and width of an image. Discreting each element in the matrix based on the above principle, the $\hat{I}$ and G are joint normalized into $1, 2, \ldots, 23$. By counting the number of each bin, a probability matrix is acquired, i.e., $\{p(X_1 = m, X_2 = n), m = 1, 2, \ldots, 23; n = 1, 2, \ldots, 23\}$. Then, the conditional probability $p(X_1 = m|X_2 = n)$ can be derived as:

$$p(X_1 = m|X_2 = n) = \frac{p(X_1 = m, X_2 = n)}{p(X_2 = n)}, \tag{20}$$

Similarly, the other conditional probabilities (i.e., $p(X_1 = m|X_3 = k)$, $p(X_2 = n|X_1 = m)$, $p(X_2 = n|X_3 = k)$, $p(X_3 = k|X_1 = m)$, $p(X_3 = k|X_2 = n)$) are also counted. Generally speaking, the statistical distribution of each feature (i.e., $p(X_1) = m, m = 1, 2, \ldots, 23$) is far greater than the corresponding conditional probability (i.e., $\{p(X_1 = m|X_2 = n), m = 1, 2, \ldots, 23; n = 1, 2, \ldots, 23\}$). Thus, the measurement of the image quality degradation depends on the statistical distribution of each feature. Moreover, there are a large number of the conditional probabilities which increases the dimension of the statistical features. To tackle these problems, the overall conditional probability $p(X_1 = m|X_2)$ is used to replace the conditional probability $p(X_1 = m|X_2 = n)$, and the overall conditional probability is calculated as:

$$\begin{aligned} p(X_1 = m|X_2) &= \sum_{n=1}^{N} p(X_1 = m|X_2 = n) \\ &= \sum_{n=1}^{N} \frac{p(X_1 = m|X_2 = n)}{p(X_2 = n)}. \end{aligned} \tag{21}$$

As shown in Eq. 18, the conditional probability (i.e., $p(X_1 = m|X_2 = n, X_3 = k)$, $p(X_2 = n|X_1 = m, X_3 = k)$,

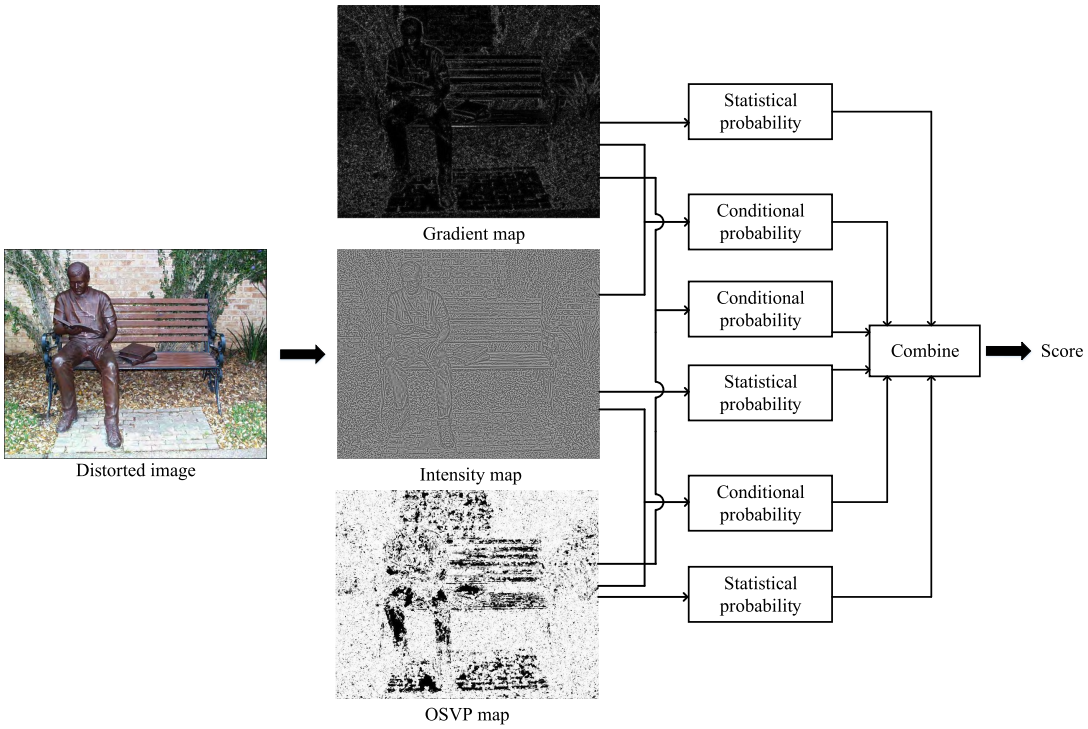


FIGURE 2. Illustration of the proposed method configurations for BIQA.

$p(X_3 = k|X_1 = m, X_2 = n)$) should be taken into account. However, the conditional probabilities of any one event occurs under the other two events are far less than the conditional probability of one event. Therefore, in this work, we take the statistical distribution of each feature and the conditional probability of two features into account to assess the image quality.

C. QUALITY DEGRADATION ASSESSMENT

Distortions decrease the visual contents of an image resulting in the image quality decreases. From the perspective of information theory, distortions degrade the amount of the visual information which can be measured by the visual information entropy. Inspired by the researches on the neuroscience, the visual contents are represented by a series of features. In this work, the local intensity, local gradient and local orientation selectivity of an image are extracted for visual information representation. Thus, the image quality is assessed based on joint feature entropy. By deducing the joint feature entropy, the joint visual entropy of images is directly related to the statistical distribution and the overall conditional probability. And thus, the six overall conditional probabilities (i.e., $p(X_1 = m|X_2)$, $p(X_1 = m|X_3)$, $p(X_2 = n|X_1)$, $p(X_2 = n|X_3)$, $p(X_3 = k|X_2)$, $p(X_3 = k|X_1)$) and the statistical distributions (i.e., $p(X_1), p(X_2), p(X_3)$) are combined and a feature set ($\mathcal{F}$=$\{p(X_1 = m|X_2), p(X_1 = m|X_3), p(X_2 = n|X_1), p(X_2 = n|X_3), p(X_3 = k|X_2), p(X_3 = k|X_1), p(X_1), p(X_2), p(X_3)\}$) is obtained for image quality assessment.

In order to assess the image quality, a mapping function is learned between the vector space $\mathcal{F}$ and the subjective quality scores $\mathcal{Q}$ (i.e., MOS, DMOS) by using a regression module. In this work, a classical support vector regression (SVR) is adopted to image quality assessment for regression [37]–[39].

$$\mathcal{R} = SVR_{learn}(\mathcal{F}, \mathcal{Q}). \tag{22}$$

The quality of a distorted image $\mathcal{I}_d$ can be predicted when the mapping function is determined,

$$\hat{\mathcal{Q}}(\mathcal{I}_d) = SVR_{predict}(\mathcal{F}(\mathcal{I}_d), \mathcal{R}), \tag{23}$$

where $\mathcal{F}(\mathcal{I}_d)$ represents the feature of the distorted image $\mathcal{I}_d$, and $\hat{\mathcal{Q}}(\mathcal{I}_d)$ represents its predicted quality score. The Fig. 2 shows the proposed method's architecture.

IV. EXPERIMENTS

In this section, the databases and evaluation criteria that used in the experiment are firstly given. Secondly, the experimental setup is explained. Then, the performances of the proposed method are displayed by comparing to the other existing state-of-the-art BIQA and FR-IQA methods. Next, the robustness of the proposed method is proved through cross-validation experiments on different databases. Finally, the efficiency of the proposed method is illustrated.

A. DATABASE AND EVALUATION CRITERIA

In this work, the proposed method is evaluated on three quality-annotated IQA databases, including LIVE [40], CSIQ [41], and TID2013 [42]. The LIVE database was the first successful quality-annotated image database, and

TABLE 1. Performance evaluation for different databases, and the best performance BIQA method is emphasized with bold.

IQM	LIVE			CSIQ			TID2013		
	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE
PSNR [44]	0.884	0.881	12.7383	0.922	0.907	0.119	0.924	0.914	0.567
MS-SSIM [45]	0.956	0.952	8.203	**0.954**	**0.958**	**0.081**	0.919	0.934	0.497
OG-IQA [46]	0.950	0.952	9.527	-	-	-	-	-	-
IL-NIQE [23]	0.919	0.920	10.616	0.887	0.884	0.131	0.883	0.875	0.673
NIQE [47]	0.926	0.925	10.209	0.901	0.910	0.115	0.845	0.835	0.769
BRISQE [48]	0.953	0.949	7.363	0.902	0.927	0.104	0.893	0.909	0.581
CORINA [49]	0.938	0.937	9.645	0.676	0.750	0.172	0.434	0.552	1.035
DIIVINE [21]	0.928	0.926	8.886	0.876	0.896	0.124	0.908	0.923	0.538
DIQA [50]	**0.975**	**0.977**	-	0.884	0.915	-	0.825	0.850	-
BIECOM [51]	0.958	0.960	-	0.815	0.823	-	0.717	0.762	-
Proposed	0.971	0.974	**6.158**	0.948	0.918	0.089	**0.960**	**0.949**	**0.396**

used widely. The LIVE contains 779 distorted images based on 29 source reference images subject to 5 different types of distortions at different levels. The CSIQ database contains 886 images generated by 30 reference images degraded by 6 types of distortions under 5 levels. The TID2013 image quality database includes 3000 distorted images which is the largest quality-annotated database. This database was generated by 25 source references with 24 different distortion types, and each distortion has 5 distortion levels.

In order to measure the performances of the proposed method, three evaluation criteria are adopted, which are SRCC (the Spearman rank order correlation coefficient), PLCC (the Pearson linear coefficient), and RMSE (the root mean squared error) [43]. In those criteria, the relationship of the predicted qualities (the quality scores evaluated from the proposed method) and the ground truth scores (MOS or DMOS) are analyzed. Those criteria display the prediction monotonicity, the prediction consistency and the prediction accuracy of the proposed method, respectively. Moreover, a better IQA method will have a larger SRCC value and PLCC value. Conversely, a better IQA method will have a smaller RMSE value.

B. EXPERIMENTAL SETUP

In order to verify the effectiveness of the proposed method, the proposed method is evaluated on the three databases mentioned above. When using SVR for quality prediction, a training procedure is required in the regression module. Similar to the most of the SVR based quality prediction [50], [51], an $80\% - 20\%$ training-testing procedure is used. In each database, 80% original scenes are randomly selected, and their corresponding distorted images are used for training, the left distorted images for testing. Moreover, in order to eliminate the bias caused by the data separation, the training-testing procedure is repeated for 1000 times, and the median performance is used for the final results.

C. PERFORMANCE EVALUATION

In this section, the performances of the proposed method are illustrated. Though the databases consist of different distortion types, there exist four types of common distortions. Thus, we choose those four common types of distortions for comparison. In order to demonstrate the performance, the proposed method is compared with some outstanding BIQA methods and two classical FR IQA methods. The results of three databases are listed in Table 1. As shown in Table 1, the PSNR and MS-SSIM are the most common FR IQA methods, and the rest are NR IQA methods. Moreover, the DIQA and BIECOM are the NR IQA methods based on convolutional neural networks. The Table 1 indicates that the proposed method obtains superior performances on TID2013 in terms of PLCC, SRCC, RMSE. However, the performances are slightly worse than the DIQA on LIVE. Although the performance of MS-SSIM is slightly better than the proposed method on CSIQ, the proposed method is the best among the existing BIQA methods.

There exist four types of common distortions in those databases, which are JP2K, JPEG, WN, Gblur. Thus, the comparisons of those four common types of distortions are implemented in this work. The performances of these IQA methods on those databases are also verified by three metrics directly. The performances on LIVE database are shown in Table 2, it is apparent that the proposed method performs highly consistent with the subjective perception. Moreover, the proposed method performs the best on three types of distortions (i.e., JP2K, JPEG, WN) among those BIQA methods, and slightly worse on Gblur. The performances of the proposed method on CSIQ database are shown in Table 3, and the results achieve 3 of 12 (3 criteria $\times$ 4 distortion type) best performances among these BIQA methods. In other performances, the experimental results of the proposed method are very similar to other BIQA methods. The performances on TID2013 database are listed in Table 4. It is apparent that the proposed method performs the best on three types of distortions (i.e., JP2K, JPEG, Gblur) when compared to other BIQA methods, and slightly worse on the WN. In summary, the proposed method gains 22 of 36 (3 database $\times$ 3 criteria $\times$ 4 distortion type) best performances among those BIQA methods. In additional, the MEON [48] has obtained the best performance on individual distortion type among the existing methods based on Convolutional Neural

TABLE 2. Performance comparison on individual distortion type of LIVE database, and the best performance BIQA method is emphasized with bold.

Distortion	Crit.	Proposed	IL-NIQE [23]	NIQE [47]	BRISQUE [48]
JP2K	PLCC	**0.968**	0.919	0.940	0.955
	SRCC	**0.956**	0.899	0.924	0.947
	RMSE	**6.187**	9.827	8.551	7.134
JPEG	PLCC	**0.979**	0.968	0.962	0.944
	SRCC	**0.966**	0.943	0.945	0.926
	RMSE	**6.333**	8.011	8.666	8.039
WN	PLCC	**0.991**	0.988	0.981	0.969
	SRCC	**0.985**	0.980	0.972	0.981
	RMSE	**3.770**	4.326	5.375	5.402
Gblur	PLCC	0.960	0.944	**0.964**	0.956
	SRCC	0.949	0.926	0.939	**0.962**
	RMSE	**5.109**	5.972	5.515	6.353

TABLE 3. Performance comparison on individual distortion type of CSIQ database, and the best performance BIQA method is emphasized with bold.

Distortion	Crit.	Proposed	IL-NIQE [23]	NIQE [47]	BRISQUE [48]	MEON [54]
JP2K	PLCC	0.948	0.894	**0.950**	0.903	0.931
	SRCC	**0.929**	0.928	0.923	0.877	0.898
	RMSE	**0.098**	0.138	0.098	0.135	-
JPEG	PLCC	0.964	0.790	0.953	0.959	**0.979**
	SRCC	0.928	0.899	0.889	0.925	**0.948**
	RMSE	**0.0811**	0.201	0.090	0.086	-
WN	PLCC	0.944	0.806	0.791	**0.960**	0.958
	SRCC	0.935	0.865	0.780	**0.952**	0.951
	RMSE	0.057	0.103	0.102	**0.050**	-
Gblur	PLCC	0.930	0.877	0.937	0.922	**0.946**
	SRCC	0.900	0.865	0.901	0.901	**0.918**
	RMSE	0.104	0.135	**0.098**	0.11	-

TABLE 4. Performance comparison on individual distortion type of TID2013 database, and the best performance BIQA method is emphasized with bold.

Distortion	Crit.	Proposed	IL-NIQE [23]	NIQE [47]	BRISQUE [48]	MEON [54]
JP2K	PLCC	**0.964**	0.915	0.925	0.916	0.924
	SRCC	**0.949**	0.911	0.905	0.902	0.911
	RMSE	**0.446**	0.680	0.639	0.684	-
JPEG	PLCC	**0.974**	0.911	0.939	0.931	0.969
	SRCC	**0.933**	0.884	0.883	0.876	0.919
	RMSE	**0.335**	0.628	0.517	0.543	-
WN	PLCC	0.93	0.901	0.862	**0.936**	0.911
	SRCC	0.930	0.892	0.849	**0.932**	0.908
	RMSE	0.254	0.315	0.357	**0.253**	-
Gblur	PLCC	**0.945**	0.861	0.868	0.889	0.899
	SRCC	**0.947**	0.862	0.845	0.894	0.891
	RMSE	**0.397**	0.631	0.611	0.565	-

Network (CNN). We can not achieve the performances in the replication experiment. Thus, the existing results in the experiments of the MEON are adopted as the comparison in this work. Unfortunately, they did not provide the results on LIVE database which lead to the methods compared with the proposed approach are different on different databases.

D. CROSS-DATABASE EVALUATION

The performances of the proposed method on each database have been shown in the former subsection. In order to reflect the generalization capability of the proposed method, the cross-validation among the three databases is adopted. Although the numbers and types of distortions for the three

TABLE 5. Performance comparison on TID2013 and CSIQ when trained on LIVE.

IQM	CSIQ			TID2013		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
Proposed	**0.940**	**0.916**	**0.096**	**0.910**	**0.912**	**0.581**
IL-NIQE [23]	0.906	0.880	0.119	0.873	0.877	0.683
NIQE [47]	0.890	0.866	0.128	0.822	0.814	0.795
BRISQUE [48]	0.840	0.826	0.153	0.721	0.726	0.969
CBIQ [55]	0.835	0.842	0.155	0.811	0.817	0.819
DIIVINE [21]	0.875	0.854	0.137	0.859	0.849	0.714

TABLE 6. Performance comparison on TID2013 and LIVE when trained on CSIQ.

IQM	LIVE			TID2013		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
Proposed	**0.936**	**0.947**	**9.581**	**0.917**	**0.902**	**0.560**
IL-NIQE [23]	0.910	0.918	11.180	0.873	0.877	0.685
NIQE [47]	0.917	0.918	10.721	0.822	0.814	0.795
BRISQUE [48]	0.643	0.632	12.252	0.583	0.570	1.135
CBIQ [55]	0.828	0.811	11.974	0.851	0.803	0.733
DIIVINE [21]	0.522	0.520	13.654	0.812	0.764	0.814

TABLE 7. Performance comparison on LIVE and CSIQ when trained on TID2013.

IQM	LIVE			CSIQ		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
Proposed	0.873	0.884	13.25	**0.910**	**0.877**	**0.118**
IL-NIQE [23]	0.913	0.915	10.99	0.906	0.880	0.119
NIQE [47]	**0.917**	**0.918**	**10.72**	0.890	0.866	0.128
BRISQUE [48]	0.789	0.795	11.82	0.839	0.808	0.153
CBIQ [55]	0.663	0.617	11.98	0.824	0.794	0.159
DIIVINE [21]	0.627	0.621	12.46	0.658	0.641	0.212

databases are different, they contain four common distortion types. Thus, in order to demonstrate the robustness of the proposed method, the cross-database evaluation is applied to the four types distortion. In other words, for the three databases, one of them is chosen for training, and the rest two for testing.

The Table 5 lists the performances on CSIQ and TID2013 databases when training on LIVE database. As can be seen, the proposed method performs best in terms of other BIQA methods in four types of distortions. Moreover, the results that training on CSIQ database and testing on LIVE and TID2013 databases are shown in Table 6. From the Table 6, we can see that the performances on LIVE and TID2013 databases are optimal in all criteria. Lastly, the Table 7 shows the results that training on TID2013 database and testing on LIVE and CSIQ databases. The performances of the proposed method on CSIQ outperform other methods apparently, and the performances on LIVE database are a slightly worse than the best one (NIQE) as shown in Table 7.

E. EXPERIMENTAL ANALYSIS

In this work, we propose a new BIQA method to assess the image quality. When the image is distorted, the visual contents of image will degrade. From the perspective of the information theory, the amount of the visual information can be measured by the visual information entropy. In order to measure and analyze the visual information, the representations of the visual information of an image are required. Inspired by researches on the neuroscience, the visual information can be represented by the local features (i.e., local intensity, local orientation selectivity, local gradient). Moreover, according to the shannon information theory, the joint visual feature entropy is related to the statistical distribution or the conditional probability of the visual features. In order to assess the image quality by the degradation of the statistical distribution, the statistics of the visual information in the natural scene without distortion should follow some kind of statistical distribution. To reveal the characteristic, we give an example that displays the statistical distribution (i.e., $p(X_1)$, $p(X_3|X_1)$) on three images. The characteristic is shown in Fig. 3, the (a), (b), (c) represent the different original images, and (d) shows the statistical distributions of these images. As can be seen, the statistical distributions of different original images have same statistical nature which provides a possibility to predict the image quality by measuring the visual information entropy.

Different levels of distortions produce individual degradations on visual contents, and distortions will generate changes

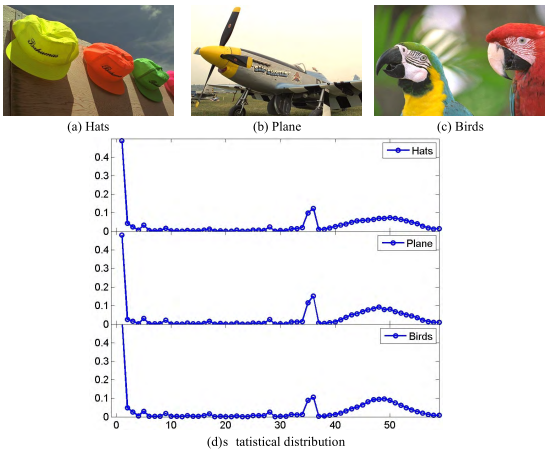


FIGURE 3. An illustration about statistical distributions of different original images. (a) Hats. (b) Plane. (c) Birds. (d) Statistical distribution.

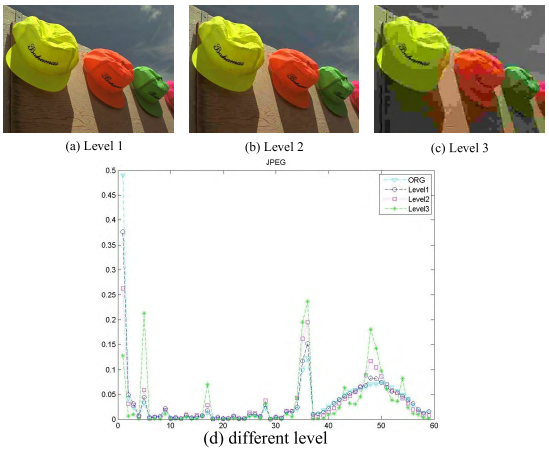


FIGURE 4. An illustration about statistical distributions of different levels of distortions on a same image. (a) Level 1. (b) Level 2. (c) Level 3. (d) Different level.

on these statistical distributions. As shown in Fig. 4, the original hat scene is degraded by different levels of JPEG distortions, the (a), (b), (c) distorted by the distortion of JPEG gradually, and (d) displays the statistical distributions of these images. As can be seen, different levels of distortions generate different changes on statistical distributions (i.e., $p(X_1)$, $p(X_3|X_1)$), which proves that these statistical probabilities can efficiently represent the quality degradation on different levels of distortions.

Moreover, the different types of distortions generate individual degradation on visual information, and different types of distortions will degrade the statistical distributions dissimilarly. In the Fig. 5, (a), (b), (c), (d) are distorted by JPEG, WN, BLUR, and JP2K, and (e) displays the different statistical distributions of a same image distorted by different distortions. As can be seen, it proves that these statistical probabilities (i.e., $p(X_1)$, $p(X_3|X_1)$) can efficiently represent the quality degradation on different types of distortions.

Since the image quality predicted by the statistical distributions of three local features, it is meaningful to understand the contribution of each feature to the whole prediction performance. In our experiments, we test the

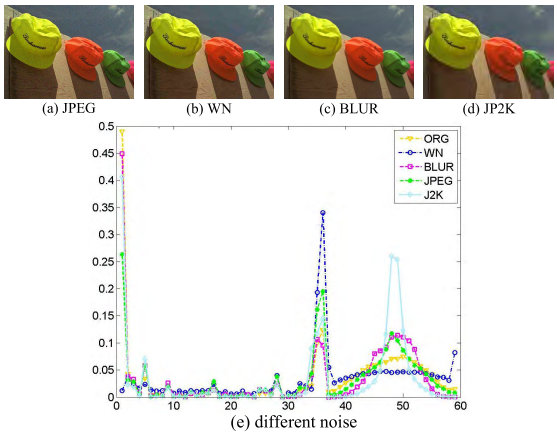


FIGURE 5. An illustration about statistical distributions of different types of distortions on a same image. (a) JPEG. (b) WN. (c) BLUR. (d) JP2K. (e) Different noise.

statistical distributions of different features on LIVE database by using 80% − 20% training-testing procedure. The Table 8 lists the PLCC, SRCC and RMSE results on the LIVE database. In this table, I, G and O represent the feature of the local intensity, the local gradient and the local orientation selectivity about an image. As can be seen, each individual feature contributes to the experimental results. And compared to the other statistical distributions, the proposed method based on the statistical distribution of all the features performs best. It also proves the rationality of the proposed method.

TABLE 8. Performance evaluation on LIVE database with different statistical distributions.

IQM	PLCC	SRCC	RMSE
I	0.891	0.878	12.225
G	0.794	0.749	16.480
O	0.932	0.930	9.853
I+G	0.923	0.917	10.371
I+O	0.950	0.948	8.445
G+O	0.950	0.948	8.489
Proposed(All)	**0.974**	**0.971**	**6.158**

In this work, we obtain the experimental results by the regression approach SVR. In order to eliminate the bias caused by the regression algorithm, other common regression algorithms (i.e. random forest regression (RFR), back propagation neural network regression (BPNNR), adaptive boosting regression (ABR), decision tree regression (DTR)) are adopted to compare with SVR. Experiments are carried out on LIVE database by using 80% − 20% training-testing procedure. The Table 9 lists the PLCC, SRCC and RMSE results produced by different regression algorithms. As can be seen, for the performances regressed by the RFR and ABR algorithms, there is no significant difference between them and the performances generated by SVR. The performances of the BPNNR algorithm are slightly worse than SVR algorithm. However, the performances of DTR algorithm are worse than the SVR algorithm obviously. Although the proposed method dose not obtain the best results on all regression

TABLE 9. Perforamce evaluation on LIVE database with different regression approaches.

Regression	PLCC	SRCC	RMSE
RFR	0.969	0.968	6.584
BPNNR	0.952	0.950	8.121
ABR	0.969	0.966	6.475
DTR	0.931	0.923	9.670
SVR	0.974	0.971	6.158

algorithms, the proposed method has a good performance in most of these algorithms. Thus, we can draw the conclusion that the performances of the proposed method do not depend on a particular regression algorithm, and it also corroborates the efficacy and the efficiency of the proposed method.

V. CONCLUSION

In this work, a novel BIQA method based on the joint entropy degradation is proposed. Distortions degrade the visual information of an image. Moreover, the visual contents of an image suffer from individual degradations by different types and different levels of distortions, which makes it difficult to measure the quality degradation. From the perspective of information theory, the amount of visual information that an image contains can be measured by the visual information entropy. Inspired by researches on the neuroscience, the visual information can be represented by a series of features. Thus, we have proposed to measure the degradations of the visual feature entropy for image quality.

By deducing the visual feature entropy, the amount of visual information is related to the statistical probabilities and conditional probabilities of the visual features. Then, the relationship between the visual information and the distribution probabilities of features are analyzed. Finally, the quality of an image is predicted by measuring the changes on the distribution probabilities of the visual features. The experimental results show that the proposed method outperforms other state-of-the-art BIQA methods, and has a good generalization capability on cross-database evaluation.

ACKNOWLEDGEMENT

The preliminary experimental results were published in IEEE International Conference on Multimedia Big Data (BigMM) [53]. In this work, the experimental process and the experimental results have been extended. The authors would like to thank the Associate Editor and the anonymous reviewers for their constructive comments on this paper.

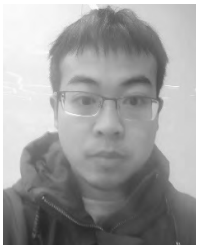
REFERENCES

[1] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, May 2011.
[2] C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 517–526, Aug. 2010.
[3] R. A. Manap and L. Shao, "Non-distortion-specific no-reference image quality assessment: A survey," *Inf. Sci.*, vol. 301, pp. 141–160, Apr. 2015.
[4] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synthesis Lectures Image Video Multimedia Process.*, vol. 2, no. 1, p. 156, 2006.
[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] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.
[7] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.
[8] 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.
[9] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, Sep. 2014.
[10] M. Oszust, "Optimized filtering with binary descriptor for blind image quality assessment," *IEEE Access*, vol. 6, pp. 42917–42929, 2018.
[11] L. Li *et al.*, "Color image quality assessment based on sparse representation and reconstruction residual," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 550–560, Jul. 2016.
[12] L. Li, Y. Yan, Z. Lu, J. Wu, K. Gu, and S. Wang, "No-reference quality assessment of deblurred images based on natural scene statistics," *IEEE Access*, vol. 5, pp. 2163–2171, 2017.
[13] P. Joshi and S. Prakash, "Continuous wavelet transform based no-reference image quality assessment for blur and noise distortions," *IEEE Access*, vol. 6, pp. 33871–33882, 2018.
[14] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.
[15] F. Pan *et al.*, "A locally-adaptive algorithm for measuring blocking artifacts in images and videos," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2004, p. III-925,
[16] F. Gao, D. Tao, X. Gao, and X. Li, "Learning to rank for blind image quality assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2275–2290, Oct. 2015.
[17] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipiQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.
[18] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Objective quality assessment and perceptual compression of screen content images," *IEEE Comput. Graph. Appl.*, vol. 38, no. 1, pp. 47–58, Jan./Feb. 2018.
[19] G. Ginesu, F. Massidda, and D. D. Giusto, "A multi-factors approach for image quality assessment based on a human visual system model," *Signal Process., Image Commun.*, vol. 21, no. 4, pp. 316–333, Apr. 2006.
[20] C. Yong, F. Hao, and L. Huanlin, "No-reference image quality assessment and application based on spatial domain coding," *IEEE Access*, vol. 6, pp. 60456–60466, 2018.
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
[26] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Royal Soc. London B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, 1980.
[27] D. L. Ruderman, "The statistics of natural images," *Network, Comput. neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.
[28] S. A. Coleman, B. W. Scotney, and S. Suganthan, "Multi-scale edge detection on range and intensity images," *Pattern Recognit.*, vol. 44, no. 4, pp. 821–838, 2011.
[29] E. Nezhadarya and R. K. Ward, "An efficient method for robust gradient estimation of RGB color images," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 701–704.
[30] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
[31] B. Chapman, K. R. Zahs, and M. P. Stryker, "Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex," *J. Neurosci.*, vol. 11, no. 5, pp. 1347–1358, 1991.

[32] T. W. Troyer, A. E. Krukowski, N. J. Priebe, and K. D. Miller, "Contrast-invariant orientation tuning in cat visual cortex: Thalamocortical input tuning and correlation-based intracortical connectivity," *J. Neurosci.*, vol. 18, no. 15, pp. 5908–5927, 1998.
[33] J. A. Cardin, L. A. Palmer, and D. Contreras, "Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex," *J. Neurosci.*, vol. 27, no. 39, pp. 10333–10344, 2007.
[34] D. Ferster and K. D. Miller, "Neural mechanisms of orientation selectivity in the visual cortex," *Annu. Rev. Neurosci.*, vol. 23, no. 1, pp. 441–471, 2000.
[35] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.
[36] F. W. Campbell and J. J. Kulikowski, "Orientational selectivity of the human visual system," *J. Physiol.*, vol. 187, no. 2, pp. 437–445, 1966.
[37] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 2013.
[38] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.
[39] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[40] H. R. Sheikh. (2003). *Image and Video Quality Assessment Research at Live*. [Online]. Available: http://live.ece.utexas.edu/research/quality
[41] E. C. Larson and D. M. Chandler. *Categorical Image Quality (CSIQ) Database*, 2010.
[42] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Euro. Workshop Vis. Inf. Process.*, Jun. 2013, pp. 106–111.
[43] V. Q. E. Group *et al.* (2003). *Final Report From the VQEG on the Validation of Objective Models of Video Quality Assessment, Pase II*. [Online]. Available: http://www.vpeg.org
[44] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.
[45] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Process., Image Commun.*, vol. 40, pp. 1–15, Jan. 2016.
[46] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[47] P. Ye, J. Kumar, L. Kang, and D. Doermann. "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1098–1105,
[48] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.
[49] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel.Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
[50] X. Xie, Y. Zhang, J. Wu, G. Shi, and W. Dong, "Bag-of-words feature representation for blind image quality assessment with local quantized pattern," *Neurocomputing*, vol. 266, pp. 176–187, Nov. 2017.
[51] 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.
[52] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.
[53] J. Wu, M. Zhang, X. Xie, G. Shi, and Z. Sun, "Joint entropy degradation based blind image quality assessment," in *Proc. IEEE 4th Int. Conf. Multimedia Big Data (BigMM)*, Sep. 2018, pp. 1–6.

WEIPING JI received the B.S. degree from the Jiangsu University of Science and Technology, Zhenjiang, China, in 2013, and the M.S. degree from the Nanjing University of Information Science and Technology, Nanjing, China. She is currently pursuing the Ph.D. degree with Xidian University, Xi'an, China. Her research interests include visual perceptual modeling and image quality assessment.

JINJIAN WU received the B.Sc. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From 2011 to 2013, he was a Research Assistant with Nanyang Technological University, Singapore. From 2013 to 2014, he was a Postdoctoral Research Fellow with Nanyang Technological University. From 2013 to 2015, he was a Lecturer with Xidian University. Since 2015, he has been an Associate Professor with the School of Electronic Engineering, Xidian University. His research interests include visual perceptual modeling, saliency estimation, quality evaluation, and just noticeable difference estimation. He was a recipient of the Best Student Paper of ISCAS, 2013. He has served as the Special Section Chair for IEEE Visual Communications and Image Processing, 2017, a Section Chair/Organizer/TPC member for ICME2014-2015, PCM2015-2016, ICIP2015, and QoMEX2016.



MAN ZHANG received the B.S. degree from the Wuhan University of Science and Technology, Wuhan, China, in 2015, and the M.S. degree from Xidian University, Xi'an, China.

ZUOZHI LIU received the B.S. degree in applied mathematics from Binzhou University, Binzhou, China, in 2010, the M.S. degree in fundamental mathematics from Northwest University, Xi'an, China, in 2013, and the Ph.D. degree in intelligent information processing from Xidian University, Xi'an, in 2018. Since 2018, he has been a Lecturer with the School of Mathematics and Statistics, Guizhou University of Finance and Economics. His research interests include intelligent information processing, sparse representation, and neural networks.

GUANGMING SHI received the B.Sc. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively, where he joined the School of Electronic Engineering, in 1988. From 1994 to 1996, he was a Research Assistant with the Department of Electronic Engineering, The University of Hong Kong, Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University. In 2004, he was the Head of the National Instruction Base of Electrician and Electronic. In 2004, he was with the Department of Electronic Engineering, University of Illinois at Urbana–Champaign, Urbana, where he is currently the Deputy Director with the School of Electronic Engineering and the Academic Leader in circuits and systems. He has authored or co-authored more than 60 research papers. His current research interests include compressed sensing, theory and design of multirate filter banks, image denoising, low-bit-rate image and video coding, and implementation of algorithms for intelligent signal processing.

XUEMEI XIE received the M.S. degree in electronic engineering from Xidian University, China, in 1996, and the Ph.D. degree in electrical and electronic engineering from The University of Hong Kong, in 2004. She is currently with the School of Electronic Engineering, Xidian University. Her research interests include digital signal processing, multirate filter bank, and wavelet transform.