A Novel Image Quality Assessment With Globally and Locally Consilient Visual Quality Perception

Sung-Ho Bae and Munchurl Kim, *Senior Member, IEEE*

***Abstract*— Computational models for image quality assessment (IQA) have been developed by exploring effective features that are consistent with the characteristics of a human visual system (HVS) for visual quality perception. In this paper, we first reveal that many existing features used in computational IQA methods can hardly characterize visual quality perception for local image characteristics and various distortion types. To solve this problem, we propose a new IQA method, called the structural contrast-quality index (SC-QI), by adopting a structural contrast index (SCI), which can well characterize local and global visual quality perceptions for various image characteristics with structural-distortion types. In addition to SCI, we devise some other perceptually important features for our SC-QI that can effectively reflect the characteristics of HVS for contrast sensitivity and chrominance component variation. Furthermore, we develop a modified SC-QI, called structural contrast distortion metric (SC-DM), which inherits desirable mathematical properties of valid distance metricability and quasi-convexity. So, it can effectively be used as a distance metric for image quality optimization problems. Extensive experimental results show that both SC-QI and SC-DM can very well characterize the HVS's properties of visual quality perception for local image characteristics and various distortion types, which is a distinctive merit of our methods compared with other IQA methods. As a result, both SC-QI and SC-DM have better performances with a strong consilience of global and local visual quality perception as well as with much lower computation complexity, compared with the state-of-the-art IQA methods. The MATLAB source codes of the proposed SC-QI and SC-DM are publicly available online at https://sites.google.com/site/sunghobaecv/iqa.**

***Index Terms*— Image quality assessment metric, local visual quality, normalized distance metric, structural contrast index.**

I. Introduction

SINCE many image processing and computer vision applications aim at maximizing perceived visual qualities and/or minimizing perceived distortions in images, it is very important to develop more elaborate computational image quality assessment (IQA) methods and to estimate degree of visual qualities for distorted images [1]–[3]. Computational IQA methods can typically be categorized into three folds according to usability of reference images: full-reference (FR), reduced-reference (RR), no-reference (NR) IQA methods [2]. This paper focus on the full reference IQA (FR-IQA) method which predicts the subjective visual quality score (z) of a distorted image ($\mathbf{Y}$) under its given reference image ($\mathbf{X}$). Mathematically, the FR-IQA modeling problem can be formulated as

$$z_i = Q \circ F\left(\mathbf{X}_i, \mathbf{Y}_i | \Theta\right) + \varepsilon_i \tag{1}$$

where $\mathbf{X}_i$, $\mathbf{Y}_i$ are the i-th reference and its distorted image, respectively, and z_i is the subjective quality score for $\mathbf{Y}_i$. $F(\cdot)$ in (1) is an FR-IQA method where Θ is a set of its model parameters and ε_i is a prediction error. Since predicted values of $F(\cdot)$ in (1) and their subjective quality scores often have nonlinear relations, a non-degenerate monotonic and non-linear function $Q(\cdot)$ is applied to map the output of $F(\cdot)$ to z_i in (1) [3].

Typically, mean squared error (MSE) has been considered one of the most commonly used distortion metric not only for the evaluation of image fidelity (or quality) but also for optimization of image processing algorithms. Although MSE is easy to compute and has many mathematically desirable properties of a valid distance metricability, differentiability and convexity, it does not highly correlate with measured subjective visual qualities for distortions [1]–[3]. To overcome this problem, many computational FR-IQA methods have been developed with an aim at coinciding with the subjective visual quality scores [1]–[20].

The Structural SIMilarity (SSIM) index [8] is considered a milestone work in the FR-IQA modeling problem. The SSIM is based on an assumption that human visual system (HVS) importantly extracts structural information from image textures in visual perception and firstly reveal that contrast/structural information in image signals can be important features to represent perceived visual quality characteristics of HVS. After the great success of SSIM [1], finding such effective features that can well characterize contrast and/or structural information in image textures has become one of key issues in enhancing performance of FR-IQA methods [11]–[16]. For example, the first order difference operators (e.g., Scharr operator [21]) and the second order difference operators (e.g., Laplacian operators [16]) have widely been adopted to characterize contrast/ structural information and have often played key roles in recent FR-IQA methods [11]–[16]. More detailed analysis on some representative FR-IQA methods and their features are given in Section II.

This paper also brings up an issue that, although the achievements of recent FR-IQA methods are quite remarkable, the efforts to develop more optimization-friendly FR-IQA

Manuscript received August 21, 2015; revised December 31, 2015; accepted March 15, 2016. Date of publication March 25, 2016; date of current version April 7, 2016. This work was supported by the National Research Foundation of Korea within the Ministry of Education, Science and Technology through the Korean Government under Grant 2014R1A2A2A01006642. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick Le Callet.

The authors are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea (e-mail: s.h.bae@kaist.ac.kr; mkim@ee.kaist.ac.kr).

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/TIP.2016.2545863

methods have scarcely been conducted. If FR-IQA methods have some desirable mathematical properties, i.e., valid distance metricability, differentiability and convexity (or quasi-convexity), then usability of the FR-IQA methods as fidelity functions in optimization problems of image processing applications may significantly be increased. For example, valid distance metricability of FR-IQA methods can help analyze and prove the convergence in many iterative algorithms. Furthermore, differentiability and convexity (or quasi-convexity) properties of FR-IQA methods will assure a global minimum on optimization problems in image processing applications.

Contributions: This paper offers three contributions including the aforementioned issues as followings:

(i) Based upon our best knowledge, we firstly reveal that HVS has different strategies in perceiving visual quality depending jointly on different image texture characteristics and distortion types. Based on this, thorough and systematic analyses of some representative features in the existing FR-IQA methods are performed in terms of their consistencies in the responses for visual quality perception according to different image texture characteristics and distortion types. These analyses draw an essential conclusion that the performance of FR-IQA methods can further be improved by adopting adaptive features that are highly consistent with the global and local visual quality perception for different distortion types and various image texture characteristics.

(ii) We propose a novel FR-IQA method, called Structural Contrast-Quality Index (SC-QI) by adopting a Structural Contrast Index (SCI) which tends to adaptively characterize perceived visual quality for different image texture characteristics with structural-distortions types where the structural-distortion is defined as a distortion due to change in inter dependency among neighbor pixels. In addition to SCI, we introduce some other perceptually important features that can effectively reflect the properties of contrast sensitivity and chrominance component variation which are also incorporated into our SC-QI. Thanks to SCI, SC-QI has a desirable advantage to elaborately quantify global and local visual qualities for various image characteristics with structural-distortion types, which were hardly been observed in other state-of-the-art FR-IQA methods. As a result, SC-QI outperforms the state-of-the art FR-IQA methods with much lower computation complexity.

(iii) We also introduce a modified version of SC-QI, called Structural Contrast-Distortion Metric (SC-DM) which is based on a low level approximation of SC-QI using a normalized distance metric. The in-depth analysis on the mathematical properties of SC-DM reveals that our SC-DM has a perception-friendly framework and is a valid distance metric having quasi-convexity in a feature space. Therefore, SC-DM has much potential to be used as an objective function for optimization problems.

This paper is organized as follows: Section II briefly reviews two important characteristics of HVS popularly used in existing FR-IQA methods and presents overview of important existing FR-IQA methods; Section III analyzes our new observation for visual quality perception and introduces the overall structure and features of the proposed SC-QI; Section IV extends SC-QI to SC-DM and analyzes the perceptual and mathematical properties of SC-DM; In Section V, the performance validations for both SC-QI and SC-DM are presented in terms of prediction accuracy and computation complexity; and Section VI concludes this paper.

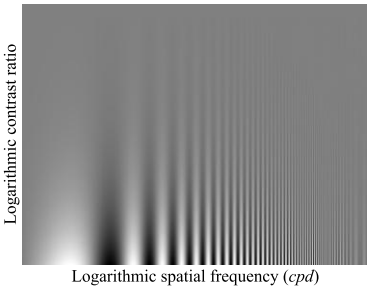


Fig. 1. An example of spatial CSF [35].

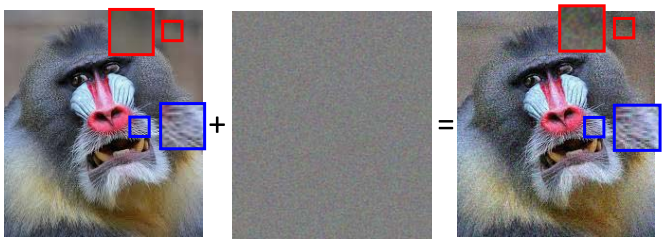


Fig. 2. An example of CM effect.

II. Related Works

A. Two Important Characteristics of HVS

We start by briefly reviewing two important characteristics of HVS, i.e., spatial contrast sensitivity function (CSF) and contrast masking (CM) effect, which have often been applied in some existing FR-IQA methods.

The spatial CSF indicates that HVS has different sensitivities to distortions depending on spatial frequency in cycles per degree (*cpd*). Fig. 1 shows an example of the spatial CSF where a sinusoidal grating pattern is injected into a gray background. In Fig. 1, the spatial frequency of the sinusoidal grating pattern increases along the x-axis, and the contrast ratio of the sinusoidal grating pattern decreases along the y-axis. As shown in Fig. 1, HVS has different sensitivities of visual perception to the changes in the sinusoidal grating pattern depending on spatial frequency, which shows a bandpass property [35].

The CM effect indicates that HVS has different sensitivities to distortions depending on background image texture characteristics. Fig. 2 shows an example of the CM effect where the red and blue boxes define homogeneous and complex texture local image regions, respectively. In Fig. 2, pseudo-additive white Gaussian noise (AGN) distortions (**N**) are injected into an original image (**X**), resulting in the distorted

image $\mathbf{Y}(=\mathbf{X}+\mathbf{N})$. Interestingly, HVS easily perceives the AWGN-distortions in the homogeneous region (e.g., red box in Fig. 2-(c)) and hardly perceives them in the image region of complex texture pattern (e.g., blue box in Fig. 2(c)). This implies that HVS perceives different visual quality to distortions depending on image texture characteristics [23].

It is noticed that modeling CM effect must take into account the degree of image texture complexity. The texture complexity is defined as a quantity that increases as the resulting distortion-visibility threshold increases due to the contrast masking effect of background image texture. A recent psychophysical experiment result revealed that perceived texture complexity is associated with not only the contrast intensity but also the structureness of background texture patterns where the structureness is defined as being inversely proportional to the randomness of image patterns [23]. Based on this observation, a novel texture complexity metric, called 'structural contrast index (SCI)' that is designed to consider both structureness and contrast intensity of image patterns was proposed [23]. In this paper, we found out that SCI can be an effective feature in characterizing perceived visual quality on images with structural-distortion types for various texture characteristics. So, SCI is used as a key feature of our FR-IQA methods.

B. Structural Contrast Index

We briefly describe the mathematical formulation of SCI [23] in a self-contained manner. The SCI (τ) is devised to estimate the perceptual complexity of image texture patterns as the ratio of structureness (KT) and contrast intensity (CI) for the coefficients of an $N \times N$ DCT block B as

$$\tau = CI^{\alpha}/KT^{\beta} \tag{2}$$

where α and β are model parameters. KT in (2) reflects the randomness of texture patterns and is defined as the kurtosis of the magnitudes of normalized DCT AC coefficients, which is given by

$$KT = m_4/(m_2)^2 \tag{3}$$

where m_k is the k-th moment of normalized DCT AC coefficients (p) and is defined as

$$m_k = \sum_{\omega \in B, \omega \neq 0} \omega^k p(\omega) \tag{4}$$

ω in (4) is a spatial frequency value in cycles per degree (cpd) for (u, v)-th DCT coefficient and is calculated by $\omega = \delta \cdot \sqrt{u^2+v^2}$ where δ is a constant and is determined by a visual angle per pixel [23]. $p(\omega)$ is the magnitude of a normalized DCT coefficient at ω and is defined as

$$p(\omega) = \left(\varepsilon + |c(\omega)|^{\lambda}\right)/Z \tag{5}$$

where Z is a normalization factor that makes $p(\omega)$ serve as a probability value over ω, and is given by

$$Z = \sum_{\omega \in B, \omega \neq 0} \left(\varepsilon + |c(\omega)|^{\lambda}\right) \tag{6}$$

In (5) and (6), $c(\omega)$ is the DCT coefficient value at ω, ε is a small constant value to avoid unstable results when the denominator in (4) is close to zero, and λ is an adjustment parameter to fit measured experimental results. The contrast intensity in (2) is defined as $CI = m_0/N^2$ where N is the height (= width) of an $N \times N$ DCT block.

In this paper, we use the inverse SCI ($\tau^* = \tau^{-1}$) for FR-IQA modeling such that more distortion-sensitive (i.e., less complex) image texture regions have higher τ^* values. The model parameters of the inverse SCI (τ^*) are empirically set to $\varepsilon = 0.25$, $\lambda = \alpha = \beta = 1$, and we scale τ^* such that the calculation of τ^* is simplified to

$$\tau^* = \frac{\sum_{(u,v)\in B} \left\{\left(u^2+v^2\right)^2 \cdot \left(\varepsilon + |c(u,v)|\right)\right\}}{\sum_{(u,v)\in B} \left\{\left(u^2+v^2\right) \cdot \left(\varepsilon + |c(u,v)|\right)\right\}^2} \tag{7}$$

where $c(u, v)$ is the (u, v)-th DCT coefficient value. Thorough analyses on SCI in terms of the characterization power of perceived visual quality is performed in Section III.

C. Overview of Important FR-IQA Methods

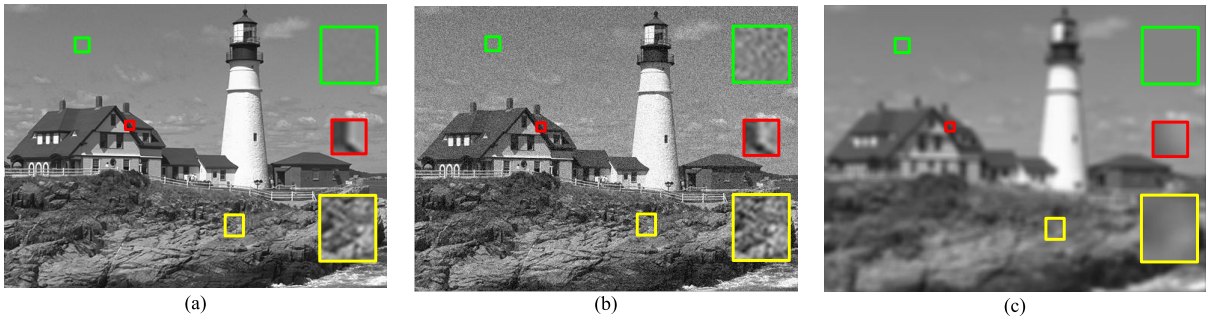
This section presents an overview of the existing FR-IQA methods. We categorize the FR-IQA methods into three folds: (i) visibility characteristic based methods; (ii) contrast/structure feature based methods; and (iii) internal brain-mechanism based methods. Some key FR-IQA methods of each category are described as follows:

1) Visibility Characteristic Based Methods:

- Perceptual Image Distortion (PID) [4], Noise Quality Index (NQM) [5] and Visual Signal-to-Noise Ratio (VSNR) [6] incorporated visibility characteristics of HVS to distortions such as contrast sensitivity functions, luminance adaptation and contrast masking effects. These methods mainly assumed that perceived visibility characteristics and perceived visual quality characteristics of HVS are proportional each other.
- Larson and Chandler [7] assumed that HVS has different strategies between distortions under visibility thresholds and distortions above visibility thresholds (supra-threshold distortions). Based on this, they proposed Most Apparent Distortion (MAD) where perceptual distortions are adaptively calculated by a geometric mean between a distortion visibility-based model and a distortion appearance-based model.

2) Contrast/Structure Feature Based Methods:

- Wang *et al.* [8] proposed SSIM based on an assumption that HVS importantly extracts structural information from image textures in visual perception. The SSIM index incorporates three representative features for luminance, contrast and structural information that are extracted by the average pixel intensity values, the standard deviation of the local image regions and the cross correlation values between two local image regions, respectively. There are many extensions to the SSIM index, such as M-SSIM (Multiscale SSIM) [9] and IW-SSIM (Information content weighted SSIM) [10], etc.



(a) (b) (c)

Fig. 3. One original image $\mathbf{X}$ and its two distorted images ($\mathbf{Y}_1$ and $\mathbf{Y}_2$) of 512×384 pixels which are obtained from the TID2013 IQA database [22]. The image $\mathbf{Y}_1$ (b) is distorted by pseudo-addictive Gaussian noise (AGN), and the image $\mathbf{Y}_2$ in (c) is distorted by pseudo-Gaussian blur (GB) from the original image $\mathbf{X}$. The green, red and yellow boxes bound the collocated homogeneous, edge and complex texture images regions.

- Liu *et al.* [11] assumed that variations of contrast/structural information based on gradient magnitudes in image signals are important in perceived visual quality. In this regard, they proposed a Gradient SIMilarity based FR-IQA method (GSM) which uses four directional high-pass filters to capture the change of contrast/structural information in images.
- Zhang *et al.* [12] proposed a Riesz transform based Feature SIMilarity (RF-SIM) index by adopting first- and second-order Riesz transforms to characterize local structures in images. And visual quality scores are calculated only for edge pixels obtained by the Canny edge operator. This is made based on an assumption that edge information is essential for evaluating visual quality scores.
- Zhang *et al.* [13] also proposed a Feature SIMilarity (FSIM) index where a phase congruency and a gradient magnitude are used as core features to characterize perceived visual quality. The phase congruency and gradient magnitude are calculated by log-scaled values of Gabor filtered coefficients and the Scharr gradient operator [21], respectively. Also, they applied adaptive local weights based on phase congruency such that perceptually more important image regions have higher weights in pooling process of estimated local visual quality values.
- Recently, Zhang *et al.* [14] proposed a Visual Saliency-induced Index (VSI) where a visual saliency index and a gradient magnitude are used as two core features. The visual saliency index is calculated by phase congruency with some simple priors (color temperature and center priors). And the gradient magnitude is obtained by the Scharr gradient operator [21]. The VSI also uses locally adaptive pooling weights that have higher values for more salient local regions.
- Xue *et al.* [15] proposed a very simple perceptual distortion metric, called Perceptual-fidelity Aware Mean Square Error (PAMSE). The PAMSE is calculated by Gaussian smoothened residuals between two images.
- Very recently, Xue *et al.* [16] also proposed a very fast FR-IQA method, called GMSD (Gradient Magnitude Similarity Deviation). This is established based on an assumption that image gradients are more vulnerable to image distortions and different local structures in distorted images from different degrees of degradations. So, the GMSD employs local image quality maps based on the global variation of gradients for overall image quality prediction. The GMSD is not only very fast for computation but also highly correlated with the perceived visual quality.

3) Internal Brain-Mechanism Based Methods:

- Sheikh and Bovik [17] treated the FR-IQA problem as an information fidelity problem. So they proposed a Visual Information Fidelity (VIF) index as an extension to its former method that is an information fidelity criterion index (IFC) [18]. In VIF, information fidelity is quantified by shared information between a reference and its distorted signals.
- Wu *et al.* [19] adopted an unified brain theory to build a FR-IQA method, called a free energy principle indicating that our brain predicts orderly (or structural) visual information and tries to avoid residual disorderly uncertainty. So, they used an autoregressive prediction algorithm such that image signals are decomposed into orderly and disorderly signals. To quantify perceived visual qualities, they calculated visual quality scores for orderly signals and disorderly signals, and combined them into a single predicted visual quality value.

Most of the aforementioned FR-IQA methods explicitly and/or implicitly assumed that contrast/structural information in image signals plays a very important role in perceived visual quality. Therefore, finding effective features to characterize such contrast/structural information may be a key issue in modeling an effective FR-IQA method.

III. Proposed Structural Contrast-Quality Index Based on Structural Contrast Index

A. Our New Observation for Visual Quality Perception

We start by revealing that HVS has different strategies in perceiving visual quality jointly for different image texture characteristics and distortion types. Fig. 3 shows an original image and its two distorted images of 512×384 pixels which are obtained from the TID2013 IQA database [22]. $\mathbf{Y}_1$ in Fig. 3-(b) is a distorted image caused by AGN, and $\mathbf{Y}_2$ in Fig. 3-(c) is a distorted image by pseudo-Gaussian blur (GB) for the original image $\mathbf{X}$ in Fig. 3-(a). The green, red and yellow boxes in Fig. 3-(a), -(b) and -(c) define the

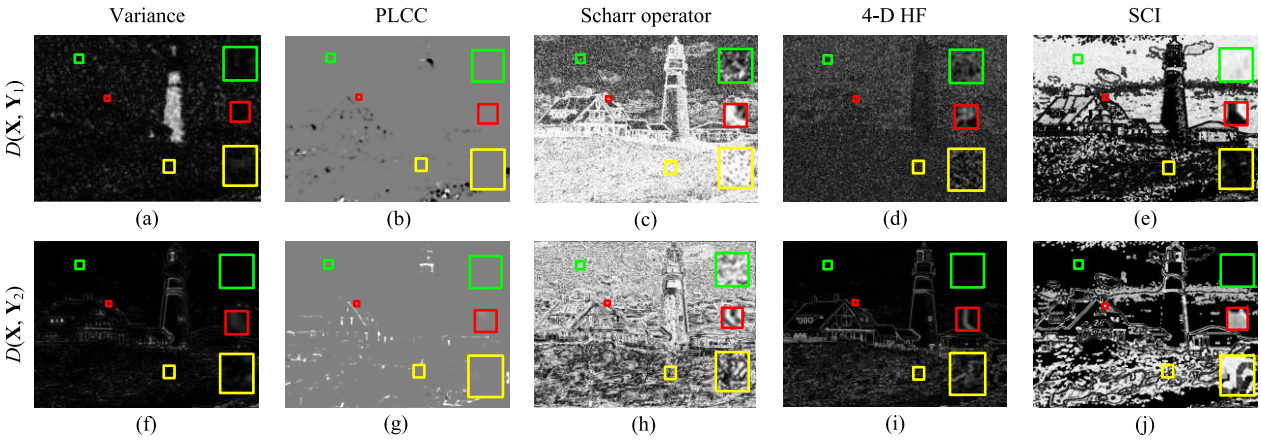


Fig. 4. The distortion maps $D(\mathbf{X}, \mathbf{Y}_1)$ and $D(\mathbf{X}, \mathbf{Y}_2)$ of the AGN-distorted image ($\mathbf{Y}_1$), and the GB-distorted image ($\mathbf{Y}_2$), respectively, for the original image ($\mathbf{X}$). The first and the second rows illustrate $D(\mathbf{X}, \mathbf{Y}_1)$ and $D(\mathbf{X}, \mathbf{Y}_2)$, and each column shows images obtained by the 'variance', 'PLCC', 'Scharr operator', '4D-HF' and 'SCI' features in order. A higher intensity value indicates a relatively larger distortion (difference) in each image.

collocated homogeneous, edge and complex texture image regions, respectively.

These two kinds of distortions differently affect the perceived visual quality of HVS. For AGN distortions, HVS tends to mostly perceive such distortions (or quality degradations) in homogeneous image regions (e.g., green box) while hardly perceiving them in complex texture regions (e.g., yellow box), as shown in Fig. 3-(a) and Fig. 3-(b). On the other hand, HVS perceives GB distortions mostly in complex texture regions (e.g., yellow box) and edge regions (e.g., red box) than the homogeneous regions (e.g., green box) as shown in Fig. 3-(a) and Fig. 3-(c). This implies that HVS perceives distortions depending on their distortion types and image texture characteristics.

B. Analysis on Contrast/Structure Features in FR-IQA Methods

We analyze the characteristics of visual features for contrast/structure information within FR-IQA methods with respect to the two different kinds of distortions (i.e., AGN and GB) in three image regions of different characteristics (homogeneous, edge and complex texture characteristics). To do so, we use the four popular visual features which have widely been used in the existing FR-IQA methods. We briefly describe these four visual features as follows:

- **Variance** [8]–[10]: this is calculated for a local image region to quantify the contrast strength of a local image pattern.
- **Pearson Linear Correlation Coefficient (PLCC)** [8]–[10]: this is jointly calculated between two local image regions to quantify structural similarity between two local regions.
- **Gradient operators** (Sobel, Prewitt and Scharr operators [13]–[16], [21]): Gradient values are obtained via first order difference operators. In this paper, we use the Scharr operator under test [21].
- **Four Directional highpass filter (4D-HF)** [11]: The masks of four different directional highpass filters are convolved for a local image region and the maximum output value among the four highpass-filtered responses is selected as a texture complexity value.

In this qualitative analysis, we also use SCI in [23] to see its effectiveness of adaptive characterization power for visual quality perception with respect to different image texture types for the two kinds of distortion types.

It is known that most FR-IQA methods have estimated distortions using a difference (or relative difference) of feature/signal values obtained from an original image and its distorted image [2]–[20]. Likewise, in the paper, we compare the difference of feature values between the original image $\mathbf{X}$ and its distorted images $\mathbf{Y}_1$ and $\mathbf{Y}_2$ to see the estimated distortions by the features for the perceived visual qualities. We apply a simple difference operation to see local distortions estimated by feature values. The absolute difference between two feature values are calculated by $|\phi(\mathbf{x})-\phi(\mathbf{y})|$, $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$, where $\phi(\cdot)$ is a feature function (e.g., a gradient operator). However, since PLCC is jointly calculated from two local image signals ($\mathbf{x}$, $\mathbf{y}$), we simply calculate distortions by taking a negative sign to the output value of PLCC.

Fig. 4 shows the distortion maps of $\mathbf{Y}_1$, and $\mathbf{Y}_2$ for $\mathbf{X}$, denoted by $D(\mathbf{X}, \mathbf{Y}_1)$ and $D(\mathbf{X}, \mathbf{Y}_2)$, respectively. The pixel intensity values in each distortion map is scaled to be [0, 255] for better visibility purpose. The first and second rows in Fig. 4 illustrate the distortion maps for $\mathbf{Y}_1$ and $\mathbf{Y}_2$, respectively, and the images in the column order are obtained by the variance, PLCC, Scharr operator, 4D-HF and SCI features, respectively. Note that a higher intensity value indicates a relatively larger distortion (difference) value. It is shown in Fig. 4-(a), -(b), -(f) and -(g) that the variance and PLCC features yield similar distortions for homogeneous, edge and complex texture regions in both AGN- and GB-distorted images, hence weakly reflecting the image texture characteristics. In case of Scharr operator, the estimated distortions for both $\mathbf{Y}_1$ and $\mathbf{Y}_2$ show slightly negative correlations with the perceived visual qualities as shown in Fig. 4-(c) and -(h). For the 4-D HF feature, the estimated distortions for both $\mathbf{Y}_1$ and $\mathbf{Y}_2$ exhibit slightly positive correlations with the perceived distortions as shown in Fig. 4-(d) and -(i).

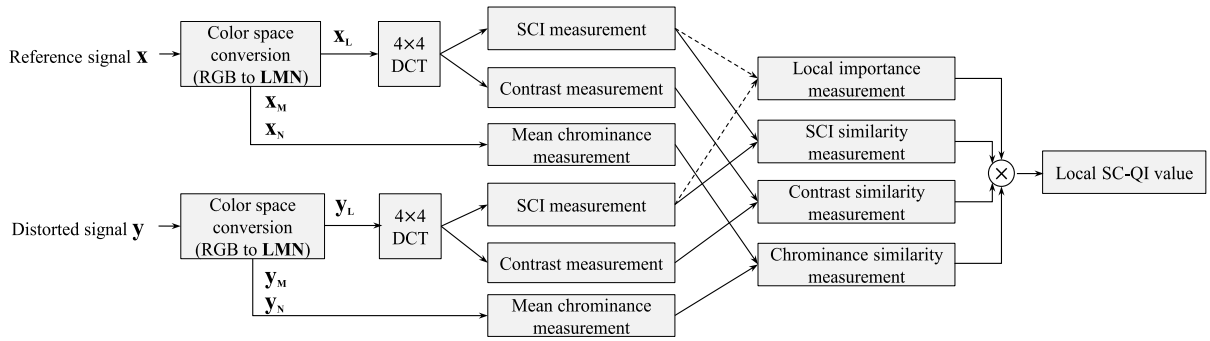


Fig. 5. A block diagram of local SC-QI computation between a local image signal $\mathbf{x}$ and its local distorted image signal $\mathbf{y}$.

It can be seen in Fig. 4-(e) and -(j) that SCI can effectively reflect the perceived distortions for both $\mathbf{Y}_1$ and $\mathbf{Y}_2$. The SCI yields large distortion values in homogeneous regions and small distortion values in complex texture regions, as shown in Fig. 4-(e), which is well agreed with the perceived distortion of HVS for $\mathbf{Y}_1$. Also, it produces small distortion values in homogeneous regions and large distortion values in edge and complex texture regions for $\mathbf{Y}_2$, as shown in Fig. 4-(j), which is also well correlated with the visual quality perception of HVS. These consistent responses of SCI with respect to HVS imply that SCI can be an effective feature to well reflect the visual quality perception of HVS for different local image characteristics with different structural-distortion types. In this regard, we incorporate SCI as a primary feature for contrast/structural information into our SC-QI and SC-DM which then have adaptive characterization power for visual quality perception.

It is noticed that, although SCI-distortion maps broadly show good agreement with visual quality perception in a subjective manner of qualitative analysis, they sometimes fail to precisely reflect visual quality perception. Especially, SCI can hardly characterize chromatic-distortions and some non-structural distortions, e.g., Mean shift (intensity shift) distortion. This is because SCI is designed to only consider structural distortions of luminance components.

C. Proposed Structural Contrast-Quality Index

Fig. 5 shows a block diagram of local SC-QI computation between a reference local image signal $\mathbf{x} \in \mathbf{X}$ within a whole original image $\mathbf{X}$ and its local distorted image signal $\mathbf{y} \in \mathbf{Y}$ within a whole distorted image $\mathbf{Y}$. A local SC-QI value is calculated for each 4×4 image block between $\mathbf{x}$ and $\mathbf{y}$. First, the pixel intensity values for signals $\mathbf{x}$ and $\mathbf{y}$ are normalized to be in the range of [0, 1]. And, if $\mathbf{x}$ and $\mathbf{y}$ are color image signals in RGB color spaces, they are converted into the LMN color space to decorrelate luminance (L) and chrominance (M, N) components [24] as

$$\begin{bmatrix} \mathrm{L} \\ \mathrm{M} \\ \mathrm{N} \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{bmatrix} \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \tag{8}$$

We define $\mathbf{x}_\mathrm{L}$, $\mathbf{x}_\mathrm{M}$ and $\mathbf{x}_\mathrm{N}$ (or, $\mathbf{y}_\mathrm{L}$, $\mathbf{y}_\mathrm{M}$, $\mathbf{y}_\mathrm{N}$) as local image blocks of the luminance and two chrominance components in $\mathbf{X}$ (or $\mathbf{Y}$), respectively. Only $\mathbf{x}_\mathrm{L}$ are transformed into DCT coefficients to calculate feature values of SCI and frequency-dependent contrast energies. We adopt a similarity measure form [8] which has popularly been used in many FR-IQA methods. A local SC-QI value between $\mathbf{x}$ and $\mathbf{y}$ is calculated by the multiplication of six similarity measures s_k, $k = 1, \ldots, 6$ as

$$f(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^{6} s_k \tag{9}$$

where all the similarity measures have the same form as

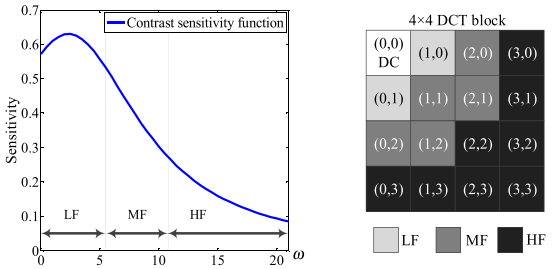
$$s_k\left(\phi_{\mathbf{x}(k)}, \phi_{\mathbf{y}(k)} | \boldsymbol{\theta}_k, \upsilon_k\right) = \left(\frac{2\phi_{\mathbf{x}(k)}\phi_{\mathbf{y}(k)} + \boldsymbol{\theta}_k}{\phi^2_{\mathbf{x}(k)} + \phi^2_{\mathbf{y}(k)} + \boldsymbol{\theta}_k} \right)^{\upsilon_k} \tag{10}$$

In (10), $\phi_{\mathbf{x}(k)} \in \mathbf{R}^1$ is the feature function of $\mathbf{x} \in \mathbf{R}^{N \times N}$ where $N = 4$ for the k-th similarity measure, and $\boldsymbol{\theta}_k$ is adopted to avoid unstable results when the denominator is close to zero. Also, $\boldsymbol{\theta}_k$ and υ_k in (10) serve as model parameters to control the descending gradient sensitivity of the similarity measure form. The product form of the multiple similarity measures in (9) has a meaning that the perceived local visual quality is low when at least one similarity measure among the multiple similarity measures is low. The SCI values between $\mathbf{x}_\mathbf{L}$ and $\mathbf{y}_\mathbf{L}$ are used in the first similarity measure (s_1) where $\phi_{\mathbf{x}(1)}$ is equal to $\tau^*(\mathbf{x}_\mathbf{L})$ in (7).

We also introduce three features to reflect the contrast sensitivity function (CSF) of HVS that has different sensitivities to distortions depending on spatial frequency [25]. To reflect the CSF into the SC-QI design, we devise three similarity measures (s_2, s_3, s_4) taking the form in (10) by comparing contrast energy values in low frequency (LF), middle frequency (MF) and high frequency (HF) regions in 4×4 DCT blocks, respectively, for $\mathbf{x}_\mathbf{L}$ and $\mathbf{y}_\mathbf{L}$. The contrast energy values $\phi_{\mathbf{x}(2)}$, $\phi_{\mathbf{x}(3)}$ and $\phi_{\mathbf{x}(4)}$ for s_2, s_3 and s_4 in (10), respectively, are calculated as

$$\phi_{\mathbf{x}(k)} = \sum_{(u,v) \in R_k} p(u, v) \tag{11}$$

where $p(u, v)$ is the normalized magnitude of a DCT coefficient at (u, v). Note that $p(u, v)$ can be converted to (5) by its spatial frequency value $\omega = \delta \cdot \sqrt{u^2 + v^2}$. Also, in (11), $k = 2$, 3 and 4. R_2, R_3 and R_4 in (11) indicate the LF, MF and HF regions, respectively. We divide these frequency regions



(a) (b)

Fig. 6. Spatial CSF in LF, MF and HF regions for a 4×4 DCT block: (a) classification of frequency regions depending on the CSF obtained in [25]; (b) Three frequency regions in a 4×4 DCT block.

depending on their sensitivities to distortions according to the psychophysical experiment results in [25]. Fig. 6 shows the spatial CSF in the LF, MF and HF regions of a 4×4 DCT block. We found out that the exponent υ_k, $k = 1, 2, 3, 4$ in (10) can help slight performance improvement of SC-QI but can introduce additional computation complexity except when $\upsilon_k = 1$. So, we set $\upsilon_k = 1$, $k = 1, 2, 3, 4$ for computation simplicity in our SC-QI.

It is known that variations of chrominance components (M and N signals in (8)) also importantly affect perceived visual quality in color images [13], [14]. To reflect this effect into SC-QI, we devise last two similarity measures s_5 and s_6 by comparing two average chrominance values, $\phi_{\mathbf{x}(5)} = E[\mathbf{x}_{\mathrm{M}}]$ and $\phi_{\mathbf{x}(6)} = E[\mathbf{x}_{\mathrm{N}}]$, where $E[\cdot]$ is the expectation operator. The model parameter values of $\boldsymbol{\theta}_k$ and υ_k in s_5 and s_6 are identically set in SC-QI (i.e., $\boldsymbol{\theta}_5 = \boldsymbol{\theta}_6$ and $\upsilon_5 = \upsilon_6$), due to their similar effects on perceived quality of HVS [13], [14].

After the local SC-QI value is estimated between each pair of $\mathbf{x}^{(j)}$ and $\mathbf{y}^{(j)}$ for all local image blocks ($j = 1, \ldots, J$) where J is the total number of the local image blocks in $\mathbf{X}$ and $\mathbf{Y}$, their weighted average is taken in a pooling stage to produce the global (whole) perceptual quality value for $\mathbf{Y}$ compared to $\mathbf{X}$. We define the global perceptual visual quality value as

$$F(\mathbf{X}, \mathbf{Y}) = E[f] = \frac{1}{W} \sum_{j=1}^{J} w\left(\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\right) f\left(\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\right) \quad (12)$$

where $w(\mathbf{x}, \mathbf{y})$ is the local weight based on visual priors with respect to a local importance (e.g., degree of image content information [9], phase congruency [13], visual saliency index [14], etc.) and W in (12) is a normalization factor and is the summation of all $w(\mathbf{x}, \mathbf{y})$ values over all J local image blocks.

In this paper, we adopt a distortion-sensitivity of local signals as the local weight in (12). This is based on an assumption that more distortion-sensitive regions (i.e., less complex texture regions) play more important roles in visual quality perception. This assumption is reasonable, since distortions in sensitive regions can easily be detected than other regions [24], [26], [27]. We utilize the inverse SCI ($\tau^* = \phi_{\mathbf{x}(1)}$ or $\phi_{\mathbf{y}(1)}$) as the local weight in the pooling stage of the local SC-QI values. The local weight for $\mathbf{x}$ is obtained as

$$w_{\mathbf{x}} = 0.2 + \phi_{\mathbf{x}}^3 \quad (13)$$

Since more sensitive regions are easily detectable for distortion than less sensitive regions, the higher one of the two $w_{\mathbf{x}}$ and $w_{\mathbf{y}}$ values in collocated regions, $\mathbf{x}$ and $\mathbf{y}$, is used as the final local weight $w(\mathbf{x}, \mathbf{y})$ for the local image region. So, $w(\mathbf{x}, \mathbf{y})$ in (12) is calculated as a softmax operator between $w_{\mathbf{x}}$ and $w_{\mathbf{y}}$, which is expressed as

$$w_{\mathbf{x},\mathbf{y}} = (A + B)^{-1} \left(A \cdot w_{\mathbf{x}} + B \cdot w_{\mathbf{y}}\right) \quad (14)$$

where $A = \exp(\chi \cdot w_{\mathbf{x}})$ and $B = \exp(\chi \cdot w_{\mathbf{y}})$, and χ is a scale factor and is set to 100 which is high enough to be a good approximate of the max operator.

The six model parameters of SC-QI, i.e., ($\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_5$ and υ_5) are found with TID2013 IQA database [22]. To avoid the over-fitting problem of the model parameters for some specific training datasets, 2-fold cross validation is performed with each pair of two subsets of TID2013 database where the initial model parameter values are randomly set, and all the distorted images in the database are randomly split into two subsets for the corresponding cross validation. In our case, the model parameters of SC-QI are optimized using a genetic algorithm [28]. In order to avoid possible content bias, this 2-fold cross validation is performed 100 times for the database. The final model parameters are determined by taking medians of all the estimated parameter values obtained from 100 times trials [29]. The final parameter values of SC-QI are $\boldsymbol{\theta}_1 = 8.7$, $\boldsymbol{\theta}_2 = 0.6$, $\boldsymbol{\theta}_3 = 2 \times 10^3$ $\boldsymbol{\theta}_4 = 1.7$, $\boldsymbol{\theta}_5 = 6.3 \times 10^{-3}$ and $\upsilon_5 = 7.3 \times 10^{-3}$.

IV. Proposed Structural Contrast Distortion Metric With Normalized Distance Metrics

Inspired by the works in [20] and [30], our SC-QI is extended to SC-DM using a normalized root mean squared error (NRMSE) such that SC-DM obtains a desirable mathematical property of valid distance metricability with quasi-convexity in a feature space. The NRMSE is expressed as

$$d(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) = \frac{\|\mathbf{x} - \mathbf{y}\|_2}{\sqrt{\|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 + \boldsymbol{\theta}}} \quad (15)$$

where $\boldsymbol{\theta}$ is a model parameter to avoid unstable results when the denominator is close to zero. To extend SC-QI to SC-DM, we adopt NRMSE for a distortion measure instead of a similarity measure such as

$$d\left(\phi_{\mathbf{x}}, \phi_{\mathbf{y}}|\boldsymbol{\theta}\right) := \mathrm{NRMSE}\left(\phi_{\mathbf{x}}, \phi_{\mathbf{y}}|\boldsymbol{\theta}\right) = \frac{\left|\phi_{\mathbf{x}} - \phi_{\mathbf{y}}\right|}{\sqrt{\phi_{\mathbf{x}}^2 + \phi_{\mathbf{y}}^2 + \boldsymbol{\theta}}} \quad (16)$$

where $\phi_{\mathbf{x}} \in \mathbf{R}_+^1$and $\phi_{\mathbf{y}} \in \mathbf{R}_+^1$are non-negative real values. The local SC-DM for $\mathbf{x}$ and $\mathbf{y}$ is formulated as the sum of squared six distortion measures (i.e., NRMSEs) as

$$f(\mathbf{x}, \mathbf{y}) = \|\mathbf{d}\|_2^2 \quad (17)$$

where $\mathbf{d} = \{d_1, d_2, \ldots, d_6\}$. d_1, d_2, d_3 and d_4 utilize the same features as s_1, s_2, s_3 and s_4 described in Section III. For d_5 and d_6, the features in s_5 and s_6 are slightly modified

to have non-negative real values as $\phi_{\mathbf{x}(5)} = E[\mathbf{x}_{\mathrm{M}}] + 0.35$ and $\phi_{\mathbf{x}(6)} = E[\mathbf{x}_{\mathrm{N}}] + 0.6$ where the minimum values of M and N channel components are -0.35 and -0.6, respectively. In the experiments, we use the six model parameters of SC-DM where $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\theta}_3$ and $\boldsymbol{\theta}_4$ are the same as those of SC-QI, and $\boldsymbol{\theta}_5$ and $\boldsymbol{\theta}_6$ are empirically set as $\boldsymbol{\theta}_5 = \boldsymbol{\theta}_6 = 2$.

The same weights and pooling method for SC-QI are employed in SC-DM. Therefore, the only changes from the SC-QI to SC-DM are the distortion measure form in (16), the local SC-DM measure form in (17), and the features for d_5 and d_6. The NRMSE offers two advantages to SC-DM as follows:

(i) The NRMSE in (16) reflects the Weber's law in HVS, indicating that the sensitivity of HVS to signal differences depends on the current stimulus [8]. Thanks to this, NRMSE is capable of yielding much higher correlations with perceived visual distortions than the L^2-norm.

(ii) The NRMSE based local SC-DM in (17) is a valid difference metric. Furthermore, it is quasi-convex and differentiable, which can widen the applicability of SC-DM in image quality optimization problems. More detailed proofs and mathematical validations of SC-DM as a valid distance metric with differentiable quasi-convex function are given in Appendix.

Regarding the first advantage in (i), we prove that NRMSE clearly reflects the Weber's law that is often modeled as

$$c = |\Delta|/x \tag{18}$$

where c is a constant value, and $\Delta = x - y$ for $x,\ y \in \mathbf{R}^1_+$ (= a set of nonnegative real number) is the smallest visibility threshold of HVS for the difference between an original signal (or stimulus) value x and its distorted signal value y. The Weber's law model in (18) indicates that visibility threshold Δ increases as the original signal value x increases. We can easily show that NRMSE has a constant value when Δ/x is fixed and $x^2 \gg \boldsymbol{\theta}$ (i.e., $\boldsymbol{\theta}/x^2 \approx 0$). That is, substituting y with $x + \Delta$ into NRMSE in (15) and rearranging it yield

$$d\,(x, y|\boldsymbol{\theta}) = \frac{(\Delta/x)}{\sqrt{1 + (1 + \Delta/x)^2}} \tag{19}$$

This indicates that Weber's law is an underlying principle in behavior of NRMSE under $x^2 \gg \boldsymbol{\theta}$.

Our local SC-DM form in (17) takes the advantage of its quasi-convexity that assures a global minimum on any convex set of a function domain [31]. To make convenient analysis for the sum of squared NRMSE in (17), we define the squared NRMSE as a normalized mean square error (NMSE) as

$$d^2\,(x, y|\boldsymbol{\theta}) := \mathrm{NMSE}\,(x, y|\boldsymbol{\theta}) = \frac{(x - y)^2}{x^2 + y^2 + \boldsymbol{\theta}} \tag{20}$$

where $\boldsymbol{\theta} \geq 0$ and $x,\ y \in \mathbf{R}^1_+$.

Fig. 7 shows an example of NMSE of x for fixed y values ($y = 1$, 2 and 3) at $\boldsymbol{\theta} = 1$. This shows that NMSE has the minimal values at $x = y$ and is quasi-convex for x.

To analyze the behavior of SC-DM in the form of a sum of NMSE components, we compare equi-distortion contours of the MSE and NMSE. Fig. 8 shows the equi-distortion contours of MSE and NMSE for the case that an image block consists of two signal components $\mathbf{x} = \{x_1, x_2\}$. There are three image blocks, each of which has three equi-distortion contours that have the same distortion levels (e.g., NMSE = 0.1, 0.2 and 0.3).

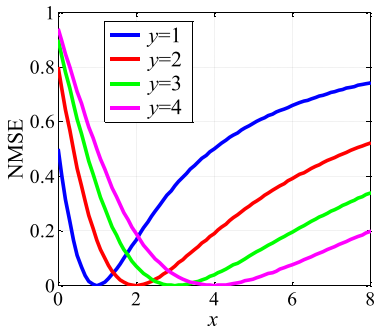


Fig. 7. NMSE plots versus x for fixed y values ($y = 1$, 2 and 3) at $\boldsymbol{\theta} = 1$.

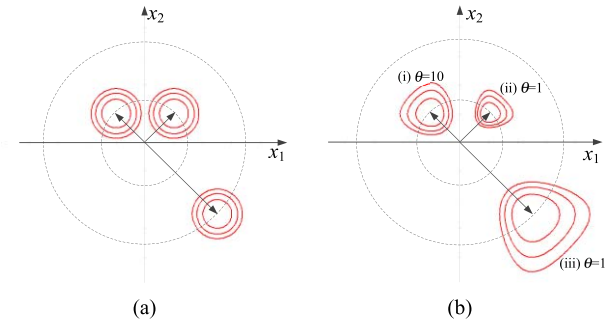


Fig. 8. The equi-distortion contours of MSE and NMSE for an image block consisting of two signal components $\mathbf{x} = \{x_1, x_2\}$. (a) MSE. (b) NMSE.

As shown in Fig. 8-(a), the shapes of equi-distortion contours of MSE are invariant to both the magnitude of the original signal components and the direction of distorted signal components, thus resulting in isotropic distortion contours at any position. On the other hand, as shown in the three cases (i), (ii) and (iii) in Fig.8-(b), the NMSE has anisotropic contours of equi-distortion depending on the magnitudes of original signal components and their distorted signal component. This clearly reflects Weber's law in (18). We also analyze the behaviors of NMSE according to the model parameters $\boldsymbol{\theta}$. The two NMSEs for the cases (i) and (ii) in Fig. 8-(b) have different model parameters of $\boldsymbol{\theta} = \{1, 1\}$ and $\boldsymbol{\theta} = \{10, 10\}$, respectively, for the same magnitude of signal components. This implies that larger model parameter values make NMSE less sensitive to the distortions than smaller ones. So, the model parameters $\boldsymbol{\theta}$ in the NMSE play two important roles of avoiding unstable results when the denominator goes to zero and controlling the sensitivity to each component (distortion) of the distortion measure in (17).

V. Experimental Results

A. Experimental Setup

Table I shows the information of the eight publicly available IQA datasets [14]. To verify the effectiveness of

TABLE I
INFORMATION OF EIGHT PUBLICLY AVAILABLE IQA DATASETS

Dataset	Reference Img. No.	Distorted Img. No.	Distortion Types	Subjects No.
TID2013	**25**	**3000**	**24**	**917**
TID2008	**25**	**1700**	**17**	**838**
CSIQ	**30**	**866**	**6**	**35**
LIVE	**29**	**779**	**5**	**161**
IVC	10	185	4	15
Toyama	14	168	2	16
A57	3	54	6	7
WIQ	7	80	5	60

TABLE II
DISTORTION TYPES IN EACH DATASET

T3*	T8*	C*	L*	Distortion types
✓	✓	✓	✓	Additive Gaussian noise
✓	✓	-	-	Additive noise (more intensive color components)
✓	✓	-	-	Spatially correlated noise
✓	✓	-	-	Masked noise
✓	✓	-	-	High frequency noise
✓	✓	-	-	Impulse noise
✓	✓	-	-	Quantization noise
✓	✓	✓	✓	Gaussian blur
✓	✓	-	-	Image denoising
✓	✓	✓	✓	JPEG compression
✓	✓	✓	✓	JPEG2000 compression
✓	✓	-	-	JPEG transmission errors
✓	✓	-	-	JPEG2000 transmission errors
✓	✓	-	-	Non eccentricity pattern noise
✓	✓	-	-	Local block-wise distortions of different intensity
✓	✓	-	-	Mean shift (intensity shift)
✓	✓	✓	-	Contrast change
✓	-	-	-	Change of color saturation
✓	-	-	-	Multiplicative Gaussian noise
✓	-	-	-	Comfort noise
✓	-	-	-	Lossy compression of noisy images
✓	-	-	-	Image color quantization with dither
✓	-	-	-	Chromatic aberrations
✓	-	-	-	Sparse sampling and reconstruction
-	-	-	✓	Fast Fading Rayleigh

*T3: TID2013, T8: TID2008, C: CSIQ, L: LIVE

both SC-QI and SC-DM compared to the state-of-the-art FR-IQA methods, we perform our experiments on the four largest databases, i.e., TID2013 [22], TID2008 [32], CSIQ [7] and LIVE [33] that contain total 6,345 distorted images with their corresponding subjective visual quality scores obtained from total 1,951 subjects. Table II shows distortion types in each database.

In our experiments, total twelve FR-IQA methods, i.e., PSNR, PAMSE [16], VSNR [6], SSIM [8], GSM [11], M-SSIM [9], IW-SSIM [10], RF-SIM [12], GMSD [15], IGM [19], FSIM_C [13] and VSI [14] are compared with SC-QI and SC-DM in terms of prediction accuracy and computation complexity. To show the prediction accuracy, we use four performance measures, i.e., Spearman Rank-Order Correlation coefficient (SROC) and Kendall Rank-Order Correlation coefficients (KROC), Pearson Linear Correlation Coefficient (PLCC) and Root Mean Square Error (RMSE). The SROC, KROC and PLCC measures present monotonicity between measured subjective visual quality scores and their estimated values by the FR-IQA methods, while RMSE presents prediction error. Usually SROC and KROC are considered representative performance measures in FR-IQA methods [3]. In order to measure PLCC and RMSE values, we adopt a logistic regression in [34] as the non-linear mapping function to map the estimated values of FR-IQA methods to their measured subjective visual quality scores. The overall performance is measured for all the four test datasets by taking weighted averages for the numbers of the test images on each database. It is noticed that overall RMSE cannot offer meaningful measure in the experiments, because each IQA database has different scales for measured subjective visual quality scores (e.g., LIVE contains the measured subjective quality scores ranging from 0 to 100, while CSIQ takes values between 0 and 1.). So, we do not measure overall RMSE in our experiments.

Likewise in the most existing FR-IQA methods [8]–[15], [19], we adopted downsizing process for input images by a factor of 2 using a 2×2 uniform kernel before our proposed SC-QI and SC-DM are performed. We found out that this downsizing process helps reduce computational complexity while scarcely degrading the performances of the SC-QI and SC-DM. To detect the blockiness artifacts in distorted images, the local SC-QI and SC-DM values are calculated using a sliding window approach with every one pixel shift [8].

B. Performance Comparison in Prediction Accuracy

Table III shows the overall performance for the FR-IQA methods under comparison in terms of prediction accuracy, where the first-, second- and third-ranked performances are highlighted in blue, red and black bolds, respectively.

Overall, both the proposed SC-QI and SC-DM outperform all other compared FR-IQA methods in terms of overall SROC, KROC and PLCC. This implies that SCI is an effective feature to reflect the perceptual behavior of HVS for visual quality perception. It is noticed that SC-QI slightly outperforms SC-DM in prediction performance. This is mainly due to the fact that the local quality (or distortion) measure forms are different for SC-QI and SC-DM. The local SC-DM is calculated by summing all the squared distortion measures, while the local SC-QI is obtained in terms of the multiplication of all similarity measures. One theory in psychophysics, called the probabilistic summation theory [35], may explain why the multiplication of all similarity measures is more appropriate than the summation of all the squared distortion measures. The probabilistic summation theory indicates that HVS perceives a distortion if at least one signal component in a subband domain is visibly distorted among various signals components. So, a multiplicative model with multiple similarity measures can be more consistent with the intrinsic behaviors of HVS than a summation model. However, it is noticed that the performance difference between SC-QI and SC-DM is small (about 0.3%- and 0.4%-point differences in overall SROC and KROC, respectively). Compared to VSI, our SC-QI (SC-DM) shows better prediction performances with 0.6(0.3)%-point, 1.3 (0.9)%-point and 0.6 (0.4)%-point higher overall SROC, KROC and PLCC values, respectively.

TABLE III
THE OVERALL PERFORMANCE OF THE FOURTEEN FR-IQA METHODS

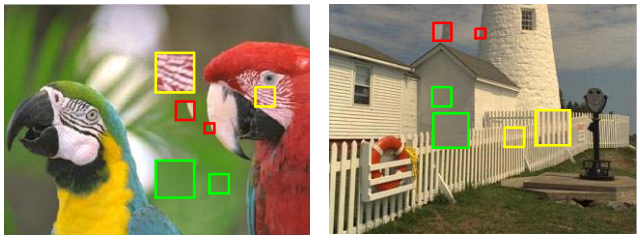
Dataset	Measure	PSNR	PAMSE	VSNR	SSIM	GSM	MSSIM	IWSSIM	RF-SIM	GMSD	IGM	$FSIM_C$	VSI	SC-DM	SC-QI
TID 2013	SROC	0.6869	0.6869	0.7385	0.7417	0.8019	0.7853	0.7779	0.8019	0.8044	0.8096	0.8510	**0.8965**	**0.9003**	**0.9052**
	KROC	0.4958	0.4958	0.5482	0.5588	0.6064	0.6058	0.5977	0.6157	0.6339	0.6394	0.6665	**0.7183**	**0.7270**	**0.7327**
	PLCC	0.6748	0.6748	0.7330	0.7895	0.8229	0.8367	0.8319	0.8333	0.8590	0.8563	0.8769	**0.9000**	**0.9049**	**0.9071**
	RMSE	0.9149	0.9149	0.8432	0.7608	0.7044	0.6789	0.6880	0.6852	0.6346	0.6403	0.5959	**0.5404**	**0.5277**	**0.5219**
TID 2008	SROC	0.5245	0.5245	0.7171	0.7749	0.7908	0.8549	0.8559	0.8659	0.8907	0.8902	0.8840	**0.8979**	**0.9021**	**0.9051**
	KROC	0.3696	0.3696	0.5309	0.5768	0.5877	0.6586	0.6636	0.6771	0.7092	0.7104	0.6991	**0.7123**	**0.7252**	**0.7294**
	PLCC	0.5309	0.5309	0.6899	0.7732	0.7845	0.8473	0.8579	0.8648	0.8788	0.8858	0.8762	**0.8762**	**0.8861**	**0.8899**
	RMSE	1.1372	1.1372	0.9715	0.8511	0.8321	0.7127	0.6895	0.6738	0.6404	0.6227	0.6468	**0.6466**	**0.6220**	**0.6120**
CSIQ	SROC	0.8389	0.8389	0.8311	0.8756	0.9112	0.9121	0.9213	0.9292	**0.9570**	0.9403	0.9310	0.9423	**0.9423**	**0.9434**
	KROC	0.6356	0.6356	0.6404	0.6907	0.7308	0.7391	0.7529	0.7640	**0.8129**	**0.7879**	0.7690	0.7857	0.7863	**0.7870**
	PLCC	0.8276	0.8276	0.8196	0.8613	0.8995	0.8983	0.9144	0.9165	**0.9541**	**0.9281**	0.9192	**0.9279**	0.9261	0.9268
	RMSE	0.1474	0.1474	0.1504	0.1334	0.1147	0.1153	0.1063	0.1050	**0.0786**	**0.0978**	0.1034	**0.0979**	0.0991	0.0986
LIVE	SROC	0.8999	0.8999	0.8346	0.9460	0.9407	0.9535	**0.9604**	0.9500	0.9546	**0.9605**	**0.9599**	0.9464	0.9475	0.9480
	KROC	0.7348	0.7348	0.6668	0.8057	0.7955	0.8224	**0.8379**	0.8144	0.8236	**0.8355**	**0.8366**	0.8000	0.8092	0.8098
	PLCC	0.7921	0.7921	0.8889	0.9385	0.9419	0.9444	**0.9515**	**0.9615**	0.9511	**0.9678**	0.9503	0.9431	0.9376	0.9373
	RMSE	14.1093	14.1093	10.5916	7.9838	7.7672	7.6027	**7.1116**	**6.3540**	7.1374	**5.8198**	7.2002	7.6856	8.0388	8.0590
Overall	SROC	0.6968	0.6968	0.7596	0.7987	0.8343	0.8453	0.8445	0.8576	0.8695	0.8704	0.8865	**0.9104**	**0.9134**	**0.9166**
	KROC	0.5174	0.5174	0.5737	0.6179	0.6464	0.6696	0.6714	0.6811	0.7056	0.7069	0.7141	**0.7379**	**0.7467**	**0.7506**
	PLCC	0.6753	0.6753	0.7566	0.8171	0.8409	0.8638	0.8675	0.8717	0.8906	0.8902	0.8933	**0.9040**	**0.9077**	**0.9098**

- For each raw, the first-, second- and third- ranked performances are highlighted in blue, red and black bolds, respectively.

TABLE IV
THE STATISTICAL TEST RESULTS FOR Δ_{VSI} VS. $\Delta_{SC\text{-}QI}$ (Δ_{VSI} VS. $\Delta_{SC\text{-}DM}$)

Data sets	p-value from Shapiro-Wilk tests (If p-value > α (= 0.05), normal distribution	p-value from Wilcoxon rank-sum tests (If p-value < α (= 0.05), significantly different)	Results of Statistical Tests
T3[1)]	0 (0)	0.0352 (0.2853)	*SD* (*NSD*)[2)]
T8[1)]	0 (0)	0.0013 (0.0029)	*SD* (*SD*)
C[1)]	0 (0)	0.2166 (0.2395)	*NSD* (*NSD*)
L[1)]	0 (0)	0.7778 (0.8371)	*NSD* (*NSD*)

[1)] T3: TID2013, T8: TID2008, C: CSIQ, L: LIVE
[2)] *NSD*: not significantly different, *SD*: significantly different



(a) $\mathbf{X}_1$ (b) $\mathbf{X}_2$

Fig. 9. Two original images $\mathbf{X}_1$ and $\mathbf{X}_2$ obtained from the TID2013 dataset [22].

In order to see the statistical significance for the performance improvement on the proposed FR-IQA methods compared to the state-of-the-art ones, we performed statistical tests between VSI and the proposed methods. For this, we define absolute difference values ($\mathbf{\Delta} = |\mathbf{z} - \hat{\mathbf{z}}|$) between measured MOS values ($\mathbf{z}$) and estimated MOS values ($\hat{\mathbf{z}}$) by an FR-IQA method where lower $\mathbf{\Delta}$ values indicate better prediction performances.

For simplicity purpose, we define the sets of absolute difference values obtained from VSI, SC-DM and SC-QI as $\mathbf{\Delta}_{VSI}$, $\mathbf{\Delta}_{SC\text{-}DM}$ and $\mathbf{\Delta}_{SC\text{-}QI}$.

We firstly check the normality of the residuals between $\mathbf{\Delta}_{VSI}$ and $\mathbf{\Delta}_{SC\text{-}QI}$ ($\mathbf{\Delta}_{SC\text{-}DM}$) using Shapiro-Wilk tests for each test IQA database. Since there is no case with a normal distribution (all p-values are approximately zero) for the residuals of each dataset, Wilcoxon rank-sum tests were performed. We performed Wilcoxon rank-sum tests for the residuals between $\mathbf{\Delta}_{VSI}$ and $\mathbf{\Delta}_{SC\text{-}QI}$ ($\mathbf{\Delta}_{SC\text{-}DM}$) with 5% significance level ($\alpha = 0.05$). Table IV summarizes the statistical testing results.

It is shown in Table IV that $\mathbf{\Delta}_{VSI}$ and $\mathbf{\Delta}_{SC\text{-}QI}$ turned out to be not significantly different (*NSD*) each other for CSIQ and LIVE, and are significantly different (*SD*) for TID2013 and TID2008. That is, the performance differences between VSI and SC-QI are statistically significant for TID2013 and TID2008 and are not statistically significant for CSIQ and LIVE. It should be noted that TID2013 and TID2008 contains many more distorted images with 24 and 17 distortion types, respectively, while CSIQ and LIVE are relatively small datasets having distorted images with only 6 and 5 distortion types, respectively. So, it can be concluded that our statistical tests confirmed the statistical significance of performance improvement on our SC-QI compared to the state-of-the-art one (VSI). While $\mathbf{\Delta}_{VSI}$ and $\mathbf{\Delta}_{SC\text{-}DM}$ turned out to be *NSD* for TID2013, CSIQ and LIVE, and are *SD* for TID2008. That is, the performance differences between VSI and SC-DM are significant only for TID2008 but are not significant for TID2013, CSIQ and LIVE. This result indicates that performance of VSI is somewhat comparable to SC-DM. Such reasonably high performance of VSI comes from the fact that it uses an important visual prior such as visual saliency information which enables to allocate higher local weights for visually salient regions during a local quality pooling stage. Some high-ranked FR-IQA methods ($FSIM_C$ and VSI) can hardly characterize localized visual quality perception for different image texture characteristics and distortion types, thus limiting

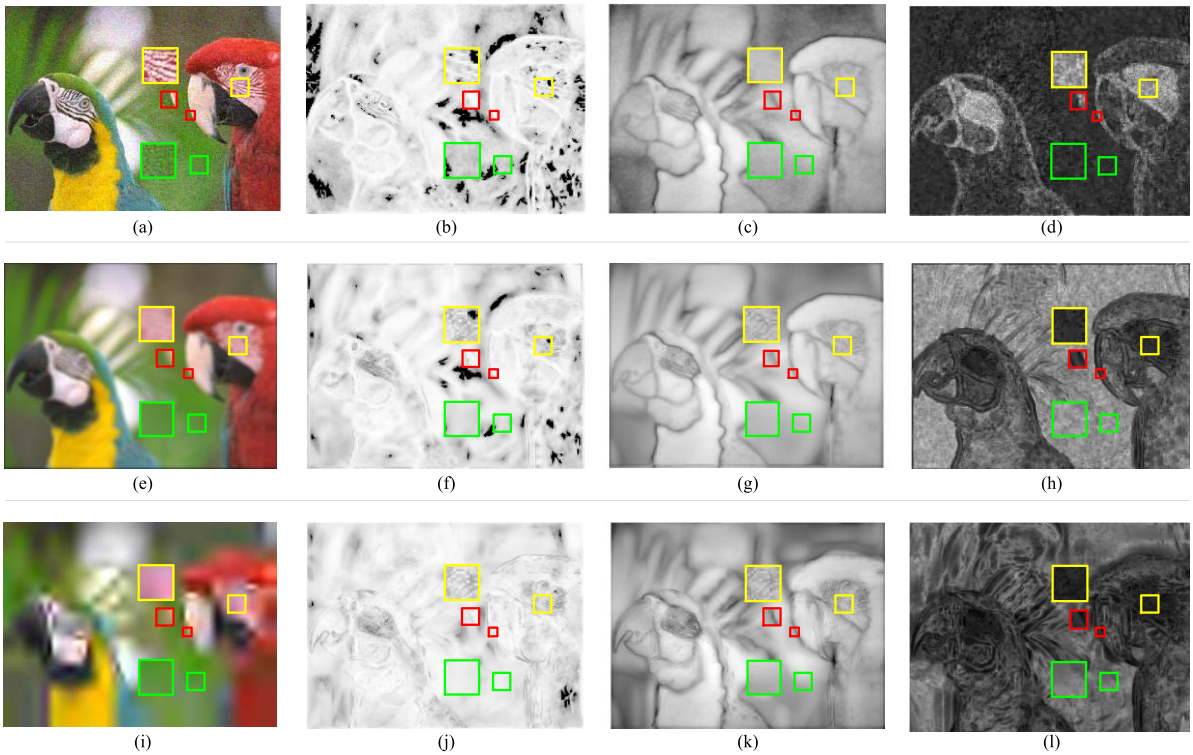


Fig. 10. Distorted images by AGN (first row), GB (second row) and J2KC (third row) from the original images $\mathbf{X}_1$ in Fig. 9. The FSIM$_C$, VSI and reversed SC-DM maps in each row are computed from $\mathbf{X}_1$ with AGN, GB and J2KC distorted images in order. (a) AGN distortion of $\mathbf{X}_1$. (b) Local FSIM$_C$ map. (c) Local VSI map. (d) Local SC-DM map (reversed). (e) GB distortion of $\mathbf{X}_1$. (f) Local FSIM$_C$ map. (g) Local VSI map. (h) Local SC-DM map (reversed). (i) J2KC distortion of $\mathbf{X}_1$. (j) Local FSIMC map. (k) Local VSI map. (l) Local SC-DM map (reversed).

its applicability only for image quality assessment tasks on whole images. While, our SC-QI and SC-DM can provide elaborate local visual quality perception maps which are highly correlated with our visual quality perception for various image texture characteristics and distortion types. Also, SC-DM has mathematically desirable properties whereas the top-ranked FR-IQA methods are neither valid distance metrics nor convex (or quasi-convex) functions. Since most of the image quality optimizations in image processing applications perform in the unit of a local image patch (e.g., [36]) rather than a whole image, our SC-QI and SC-DM have much potential to be applied in such image processing applications to estimate local visual quality values.

C. Analysis on SC-DM for Local Visual Quality Perception

We thoroughly analyze the performance consistency of our SC-DM and SC-QI with respect to local visual quality perception of HVS for different image texture characteristics and distortion types. For this, the quality maps of the two top-ranked FR-IQA methods, FSIM$_C$ and VSI, except SC-DM and SC-QI in Table III, are demonstrated for performance comparison. Since SC-DM and SC-QI have very similar local visual quality maps, we representatively demonstrate local visual quality maps for SC-DM.

We illustrate local visual quality maps of FR-IQA methods for three different kinds of distortions with AGN, GB and JPEG2000 Compression noise (J2KC), which are popularly observed in many image processing applications. Since local SC-DM estimates perceived distortions which are inversely proportional to the perceived visual qualities, we demonstrate reversed local SC-DM maps for the convenient comparison with other FR-IQA methods. So, higher pixel intensities in visual quality maps indicate higher visual quality values.

In our experiments, we use two reference images selected from TID2013 database [22]. Fig. 9 shows the two reference images $\mathbf{X}_1$ and $\mathbf{X}_2$ of 512×384 pixels. In our experiments, the green, red and yellow boxes in figures define the homogeneous, edge and complex texture image regions, respectively.

Fig. 10 and Fig. 11 show the distorted images of $\mathbf{X}_1$ and $\mathbf{X}_2$ by AGN, GB and J2KC distortions with their scaled local visual quality maps of the three FR-IQA methods (FSIM$_C$, VSI, SC-DM). For the AGN-distorted images, perceived distortions are quite visible in homogeneous image regions while they are less visible in complex texture regions as shown in Fig. 10-(a) and Fig. 11-(a). It can be seen in Fig. 10-(b), -(c), -(d) and Fig. 11-(b), -(c), -(d) that SC-DM is very consistent with the perceived distortions. That is, SC-DM estimates lower visual quality values in homogeneous regions and higher visual quality values in complex texture regions, while FSIM$_C$ and VSI can hardly capture the distortions in homogeneous image regions.

Also, for the GB-distorted images, perceived distortions are mostly affected from the blurred edges and texture regions while they are less visible in homogeneous regions

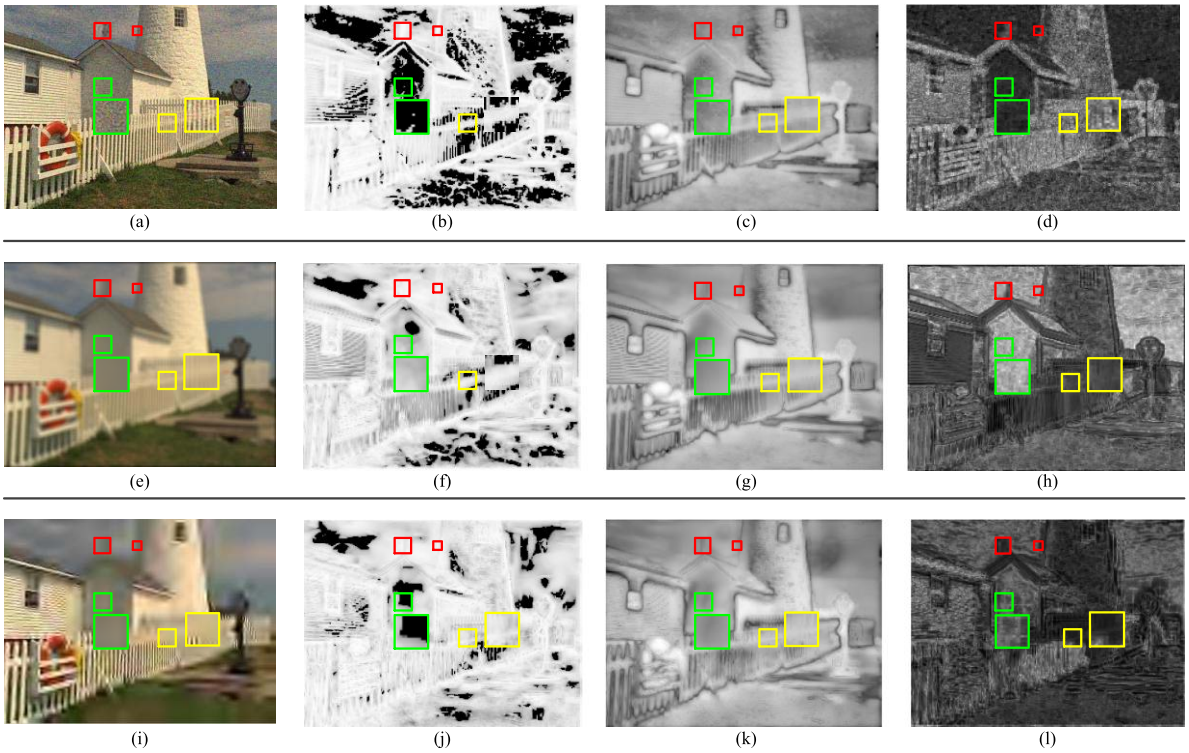


Fig. 11. Distorted images by AGN (first row), GB (second row) and J2KC (third row) from the original images $\mathbf{X}_2$ in Fig. 9. The FSIM$_C$, VSI and reversed SC-DM maps in each row are computed from $\mathbf{X}_2$ with AGN, GB and J2KC distorted images in order. (a) AGN distortion of $\mathbf{X}_2$. (b) Local FSIM$_C$ map. (c) Local VSI map. (d) Local SC-DM map (reversed). (e) GB distortion of $\mathbf{X}_2$. (f) Local FSIM$_C$ map. (g) Local VSI map. (h) Local SC-DM map (reversed). (i) J2KC distortion of $\mathbf{X}_2$. (j) Local FSIM$_C$ map. (k) Local VSI map. (l) Local SC-DM map (reversed).

as shown in Fig. 10-(e) and Fig. 11-(e). It can be seen in Fig. 10-(f), -(g), -(h) and Fig. 11-(f), -(g), -(h) that SC-DM appropriately predicts lower visual quality values in edge and complex texture image regions and higher visual quality values in homogeneous image regions, which is well agreed with visual quality perception of HVS. However, FSIM$_C$ and VSI can scarcely reflect the considerable quality degradations in complex texture image regions, which is not consilient with local visual quality perception of HVS. It is also observed in Fig. 10-(f) and Fig. 11-(f) that FSIM$_C$ has negative correlations with perceived visual quality for edge image regions, i.e., FSM$_C$ predicts high visual quality values for blurred edge regions. Similar results are observed in the J2CK distorted images as shown in Fig. 10-(i), -(j), -(k), -(l) and Fig. 11-(i), -(j), -(k), -(l).

In conclusion, SC-DM and SC-QI have a distinct merit of adaptive characterization power for global and local visual quality perception depending various image characteristics and distortion types.

D. Performance Comparison in Computation Complexity

The computation complexity of each FR-IQA method is also measured in terms of average frames per second (*fps*). For this, all distorted images in TID2013 dataset are used. This experiment is performed on a 3.2 GHz Intel Core i7$^{\text{TM}}$ processor with 24GB RAM. Table V tabulates the average running speeds of the fourteen FR-IQA methods under comparison.

TABLE V
AVERAGE RUNNING SPEED OF THE FOURTEEN FR-IQA METHODS IN FRAMES PER SECOND (fps)

Method	*fps*	Method	*fps*
PSNR	526.51	SC-DM	17.45
PAMSE	129.37	SC-QI	16.98
GMSD	119.82	VSNR	13.15
SSIM	74.70	VSI	8.08
M-SSIM	29.81	IW-SSIM	4.15
GSM	28.29	FSIM$_C$	6.86
RF-SIM	26.26	IGM	0.13

It is shown in Table V that PSNR, PAMSE, GMSD and SSIM are much faster than SC-QI and SC-DM. However, their performances are fairly worse even than the top-3 recent FR-IQA methods, i.e., IGM, FSIM$_C$ and VSI. So, we compared the running speeds of SC-QI and SC-DM with the top-3 FR-IQA methods (IGM, FSIM$_C$, and VSI). Since our proposed methods and the top-3 FR-IQA methods have the same downsizing process, the running speeds of them are compared for the test images of the downsized resolutions.

Fig. 12 shows the comparison for the top-3 state-of-the-art FR-IQA methods and the proposed SC-QI and SC-DM in terms of prediction accuracy (overall SROC and KROC in Table III) and computation complexity (*fps*). As shown in Fig. 12, it is clear that SC-QI and SC-DM not only outperform the prediction performance but also run much faster than the top-3 FR-IQA methods.

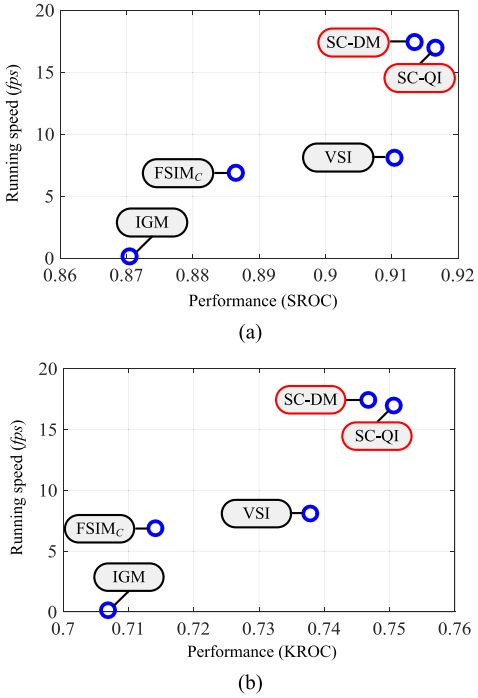


Fig. 12. Comparison for the top 3 FR-IQA methods and the proposed SCIQ and SC-DM in terms of estimation accuracy (overall SROC and KROC) and computation complexity (*fps*): (a) SROC and (b) KROC.

VI. Conclusion

In this paper, we firstly reveal that HVS has different strategies in measuring perceived visual quality depending on different image texture characteristics and distortion types. Based on this, we propose a novel FR-IQA index, called SC-QI, which incorporates a very effective feature, i.e., SCI, to adaptively characterize local visual quality perception with respect to different image texture characteristics with structural-distortion types. We also further extend SC-QI to SC-DM based on a normalized distance metric such that SC-DM obtains desirable mathematical properties of valid distance metricability and quasi-convexity. To verify the performance of the proposed SC-QI and SC-DM, qualitative and quantitive experiments were extensively performed on large IQA datasets. The experimental results show that SC-QI and SC-DM can very elaborately predict global and local visual quality scores, thus yielding better prediction performance with fast computation compared to the state-of-the-art FR-IQA methods. Since many image quality optimizations in image processing applications performed in units of local image patches, our SC-QI and SC-DM can have great merit to be applied into such image processing applications.

Appendix

We first show that NRMSE in (16) for $x,\ y \in \mathbf{R}_+^1$ is a valid distance metric which is referred to [20].

Theorem 1: The NRMSE is a metric for $\boldsymbol{\theta} > 0$.

Proof: It is easy to verify that

1. $d(x, y) = d(y, x)$
2. $d(x, y) \geq 0$ for $\forall x,\ y \in \mathbf{R}_+^1$
3. $d(x, y) = 0$, if and only if $x = y$

It remains to prove the triangular inequality such that

$$d(x, y) + d(y, z) \geq d(x, z) \tag{21}$$

where $x,\ y,\ z \in \mathbf{R}_+^1$. Without loss of generality, we assume that $x \leq y$ such that (21) is rearranged as

$$\frac{\sqrt{x^2+z^2+\boldsymbol{\theta}}}{\sqrt{x^2+y^2+\boldsymbol{\theta}}} \cdot \frac{|x-y|}{|x-z|} + \frac{\sqrt{x^2+z^2+\boldsymbol{\theta}}}{\sqrt{y^2+z^2+\boldsymbol{\theta}}} \cdot \frac{|y-z|}{|x-z|} \geq 1 \tag{22}$$

(22) can be proved by considering three cases as follows:

Case 1) $y \leq x \leq z$. The left second term in (22) holds

$$\frac{\sqrt{x^2+z^2+\boldsymbol{\theta}}}{\sqrt{y^2+z^2+\boldsymbol{\theta}}} \geq 1, \quad \text{and} \ \frac{|y-z|}{|x-z|} \geq 1 \tag{23}$$

So, the triangular inequality holds when $y \leq x \leq z$.

Case 2) $x \leq z \leq y$. The left first term in (22) holds

$$\frac{\sqrt{x^2+z^2+\boldsymbol{\theta}}}{\sqrt{x^2+y^2+\boldsymbol{\theta}}} \geq 1, \quad \text{and} \ \frac{|x-y|}{|x-z|} \geq 1 \tag{24}$$

So, the triangular inequality holds when $y \leq x \leq z$.

Case 3) $x \leq y \leq z$. For two constants $a \geq 0$ and $c \geq 0$, we define a convex function

$$f(x) := \sqrt{x^4 + 2cx^2 + a} \tag{25}$$

where the convexity of $f(x)$ can easily be verified by checking the positive sign of its second derivative. By definition of convexity, we have

$$f(y) \leq \lambda \cdot f(x) + (1-\lambda) \cdot f(z) \tag{26}$$

The constants, a and λ, are chosen to have

$$a := x^2y^2 + y^2z^2 + x^2z^2 + c\left(x^2+y^2+z^2\right) + c^2 \tag{27}$$

$$\lambda = |z-y|/|x-z| \tag{28}$$

Finally, we obtain

$$\frac{|x-y|}{\sqrt{x^2+y^2+\boldsymbol{\theta}}} + \frac{|y-z|}{\sqrt{y^2+z^2+\boldsymbol{\theta}}} \geq \frac{|x-z|}{\sqrt{x^2+z^2+\boldsymbol{\theta}}} \tag{29}$$

Therefore, the proof is complete. ■

We now verify that the L^2-norm of vectors consisting of NRMSE elements, i.e., $||\mathbf{d}||_2$, belongs to a metric space. For this, the following increasing property is needed:

Definition 1: For any $\mathbf{x} \in \mathbf{R}_+^K$ and $\mathbf{y} \in \mathbf{R}_+^K$, an L^2-norm in $\mathbf{R}_+^K$ satisfies the following increasing property:

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}+\mathbf{y}\|_2 \tag{30}$$

The following theorem shows that $||\mathbf{d}||_2$ is also a metric.

Theorem 2: For any $\mathbf{x} \in \mathbf{R}_+^K$ and $\mathbf{y} \in \mathbf{R}_+^K$, $||\mathbf{d}||_2$ is a metric in $\mathbf{R}_+^K$.

Proof: It is easy to verify that

1. $\|\mathbf{d}(\mathbf{x}, \mathbf{y})\|_2 = \|\mathbf{d}(\mathbf{y}, \mathbf{x})\|_2$
2. $\|\mathbf{d}(\mathbf{x}, \mathbf{y})\|_2 \geq 0$ for $\forall\ \mathbf{x}, \mathbf{y} \in \mathbf{R}_+^K$
3. $\|\mathbf{d}(\mathbf{x}, \mathbf{y})\|_2 = 0$, if and only all $\mathbf{x} = \mathbf{y}$

It remains to prove the triangular inequality. Since $d_1, \ldots, d_6$ are all metrics, every distortion component of $\mathbf{d}$ satisfies

the triangular inequality. From the increasing property in Definition 3, we conclude that

$$\begin{aligned}\|\mathbf{d}(\mathbf{x},\mathbf{z})\|_2 &\le \|\mathbf{d}(\mathbf{x},\mathbf{y})+\mathbf{d}(\mathbf{y},\mathbf{z})\|_2 \\ &\le \|\mathbf{d}(\mathbf{x},\mathbf{y})\|_2+\|\mathbf{d}(\mathbf{y},\mathbf{z})\|_2 \end{aligned} \tag{31}$$

Therefore, the proof is complete. ■

We now show that NMSE in (20) is a quasi-convex function. For the convenience of the reader, we describe the definition of quasi-convexity. We start by defining convexity of a set.

Definition 2: A subset X of a vector space is convex for $\forall x,\ y \in X$ and for $\lambda \in (0, 1)$ if it satisfies

$$\lambda x + (1-\lambda)\, y \in X \tag{32}$$

Definition 3: For a convex set X, a function $f\colon X \to \mathbf{R}$ is referred to be quasi-convex if its h-sublevel set X_h satisfies a convex set for all $h \in \text{Range}(f)$, where the h-sublevel set is defined as

$$X_h = \{x \in X \text{ and } f(x) \le h\} \tag{33}$$

Theorem 3: For a fixed y, NMSE is quasi-convex for $H = \{x, y \in \mathbf{R}_+^1\}$.

Proof: We start by seeing the range of NMSE. Substituting x with $\Delta + y$ and a rearrangement yields

$$\text{NMSE}(x, y|\boldsymbol{\theta}) = \frac{1}{(1+y/\Delta)^2 + (y/\Delta)^2 + \boldsymbol{\theta}} \tag{34}$$

This shows that NMSE has the maximum value of $(1+\boldsymbol{\theta})^{-1}$ for $\Delta = \infty$ or $-y$. So, the NMSE has the maximum range of [0, 1) when $\boldsymbol{\theta} > 0$. Therefore, it is sufficient that the NMSE is quasi-convex for NMSE < 1. We first establish an inequality to see that the set of points x under NMSE < 1 is a convex set as

$$\text{NMSE}(x, y|\boldsymbol{\theta}) \le h \tag{35}$$

where $h \in [0, 1)$. This inequality becomes

$$(x - y/(1-h))^2 \le y^2((1-h)^{-2} - 1) + \boldsymbol{\theta} h/(1-h) \tag{36}$$

This inequality holds that the set of points x in (36) is a round, indicating that the set is a convex set. In the case for $h = 1$, we have the inequality as

$$x \ge -y\boldsymbol{\theta}/2 \tag{37}$$

Since $\boldsymbol{\theta} \in \mathbf{R}_+^1$, NMSE holds quasi-convexity for $x,\ y \in \mathbf{R}_+^1$. ■

It is straightforward to see that the local SC-DM in (17) is also quasi-convex, because each squared distortion measures in (17) is the form of NMSE which is quasi-convex.

REFERENCES

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

[2] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Rafael, CA, USA: Morgan & Claypool, 2006.

[3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 1477–1480.

[4] P. C. Teo and D. J. Heeger, "Perceptual image distortion," *Proc. SPIE*, vol. 2179, pp. 127–141, May 1994.

[5] 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.

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

[7] 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, pp. 011006:1–011006:21, Jan. 2010.

[8] 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.

[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2003, pp. 1398–1402.

[10] 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.

[11] 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.

[12] L. Zhang, D. Zhang, and X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 321–324.

[13] 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.

[14] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.

[15] 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.

[16] W. Xue, X. Mou, L. Zhang, and X. Feng, "Perceptual fidelity aware mean squared error," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 705–712.

[17] 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.

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

[19] 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.

[20] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1488–1499, Apr. 2012.

[21] B. Jähne, H. Haußecker, and P. Geißler, *Handbook of Computer Vision and Applications*. New York, NY, USA: Academic, 1999.

[22] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th Eur. Workshop Vis. Inf. Process.*, Jun. 2013, pp. 106–111.

[23] S. H. Bae and M. Kim, "A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3227–3240, Aug. 2014.

[24] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 12, pp. 1338–1350, Dec. 2001.

[25] S.-H. Bae and M. Kim, "A novel DCT-based JND model for luminance adaptation effect in DCT frequency," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 893–896, Sep. 2013.

[26] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.

[27] S.-H. Bae and M. Kim, "A new DCT-based JND model of monochrome images for contrast masking effects with texture complexity and frequency," in *Proc. IEEE Int. Conf. Image Proc.*, Melbourne, VIC, Australia, Sep. 2013, pp. 431–434.

[28] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*. NewYork, NY, USA: Wiley, 1996.

[29] F. Zhang and D. Bull, "A perception-based hybrid model for video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, May 2015, doi: 10.1109/TCSVT.2015.2428551.

[30] L. Chen, "A new weighted metric: The relative metric I," *J. Math. Anal. Appl.*, vol. 274, no. 1, pp. 38–58, 2002.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
[32] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, Oct. 2009.
[33] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[34] (2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment VQEG*. [Online]. Available: http://www.vqeg.org
[35] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vis. Res.*, vol. 21, no. 3, pp. 409–418, Mar. 1981.
[36] J. Kim, S. H. Bae, and M. Kim, "An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1786–1800, Nov. 2015.

Sung-Ho Bae received the B.S. degree from Kyung-Hee University, Suwon, South Korea, in 2011, and the M.S. degree from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2012, where he is currently pursuing the Ph.D. degree. His research interests include machine learning and computer vision, visual perception characteristics, perceptual video coding, and visual quality assessment on HDR/UHD video and super-resolution.

Munchurl Kim (M'99–SM'13) received the B.E. degree in electronics from Kyungpook National University, Daegu, South Korea, in 1989, and the M.E. and Ph.D. degrees in electrical and computer engineering from the University of Florida, Gainesville, in 1992 and 1996, respectively. After his graduation, he joined the Electronics and Telecommunications Research Institute, Daejeon, South Korea, as a Senior Research Staff Member, where he led the Realistic Broadcasting Media Research Team. In 2001, he was an Assistant Professor with the School of Engineering, Information and Communications University, Daejeon. Since 2009, he has been with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, where he is currently a Full Professor. He had been involved with scalable video coding and High Efficiency Video Coding in JCT-VC standardization activities of ITU-T VCEG and ISO/IEC MPEG. His current research interests include high performance video coding, perceptual video coding, visual quality assessments on 3-D/UHD video, computational photography, machine learning, and pattern recognition.