Unified Blind Quality Assessment of Compressed Natural, Graphic, and Screen Content Images

Xiongkuo Min, Kede Ma, *Student Member, IEEE*, Ke Gu, Guangtao Zhai, *Member, IEEE*, Zhou Wang, *Fellow, IEEE*, and Weisi Lin, *Fellow, IEEE*

***Abstract*— Digital images in the real world are created by a variety of means and have diverse properties. A photographical natural scene image (NSI) may exhibit substantially different characteristics from a computer graphic image (CGI) or a screen content image (SCI). This casts major challenges to objective image quality assessment, for which existing approaches lack effective mechanisms to capture such content type variations, and thus are difficult to generalize from one type to another. To tackle this problem, we first construct a cross-content-type (CCT) database, which contains 1,320 distorted NSIs, CGIs, and SCIs, compressed using the high efficiency video coding (HEVC) intra coding method and the screen content compression (SCC) extension of HEVC. We then carry out a subjective experiment on the database in a well-controlled laboratory environment. Moreover, we propose a unified content-type adaptive (UCA) blind image quality assessment model that is applicable across content types. A key step in UCA is to incorporate the variations of human perceptual characteristics in viewing different content types through a multi-scale weighting framework. This leads to superior performance on the constructed CCT database. UCA is training-free, implying strong generalizability. To verify this, we test UCA on other databases containing JPEG, MPEG-2, H.264, and HEVC compressed images/videos, and observe that it consistently achieves competitive performance.**

***Index Terms*— Natural scene image, computer graphic image, screen content image, image quality assessment, high efficiency video coding (HEVC), screen content compression (SCC).**

Manuscript received November 26, 2016; revised April 7, 2017, June 6, 2017, and July 16, 2017; accepted July 17, 2017. Date of publication August 2, 2017; date of current version September 1, 2017. This work was supported in part by National Natural Science Foundation of China under Grant 61422112, Grant 61371146, Grant 61521062, and Grant 61527804, in part by the Natural Sciences and Engineering Research Council of Canada, in part by Singapore MoE Tier 1 Project under Grant M4011379 and Grant RG141/14, and in part by the Singapore MoE Tier 2 Project under Grant M4020355.020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stefan Winkler. *(Corresponding author: Guangtao Zhai.)*

X. Min is with the Institute of Image Communication and Network Engineering, Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: minxiongkuo@gmail.com).

K. Ma and Z. Wang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: k29ma@uwaterloo.ca; zhou.wang@uwaterloo.ca).

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

G. Zhai is with the Institute of Image Communication and Network Engineering, Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@sjtu.edu.cn).

W. Lin is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

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.2017.2735192

I. Introduction

With the rapid development of mobile and cloud computing, massive content composed of both natural scenes, graphics, and screen content is being generated everyday [1]–[3]. Different from photographic natural scene images (NSIs) which are captured from real-world scenes, computer graphic images (CGIs) and screen content images (SCIs) are mostly generated artificially by computers. CGIs are digitally synthesized visual content (sometimes simulate real-world scenes), which are widely used in animations, video games, graphic designs, etc. SCIs are generally interfaces presented to users on screens of digital viewing devices. SCIs are often composed of texts, tables, dialogs, and sometimes natural and computer-generated content. SCIs and CGIs have substantially different characteristics from NSIs. For example, SCIs and CGIs tend to have more noise-free smooth areas, high-saturation color content, and extremely sharp texts and edges. Many specialized technologies have been proposed to take care of these differences, especially screen content coding (SCC) [1], [4]. The joint collaborative team on video coding (JCT-VC) also develops HEVC-SCC as an extension of high efficiency video coding (HEVC) [5].

Nevertheless, little work has been dedicated to assess the quality of SCIs, and even less has been devoted to cross-content-type quality assessment. Over the last two decades, a variety of image quality assessment (IQA) measures have been proposed [6]–[11], most of which show good promises on natural content, but several studies have shown that state-of-the-art IQA measures designed for natural content cannot predict the quality of screen content properly [2], [3].

Limited work has been done on quality assessment of SCIs and CGIs. Yang *et al.* [2] constructed a screen IQA database (SIQAD), which contains SCIs corrupted by traditional distortions, *e.g.*, noise, blur, JPEG and JPEG2000 compression. Wang *et al.* [3] conducted a subjective user study on SCIs. Different from traditional distortions studied in SIQAD, they considered the distortions introduced by HEVC and HEVC-SCC compression, and constructed a screen content database (SCD) for SCI quality assessment. Wang *et al.* [12] proposed an SCI quality assessment measure based on visual field adaptation and information content weighting. Gu *et al.* [13] and Ni *et al.* [14], [15] incorporated gradient and edge information in SCI quality assessment. Fang *et al.* [16] designed an uncertainty weighting method to fuse the quality of textual and pictorial regions. Gu *et al.* [17] developed a blind SCI quality model through big data learning. It is worth

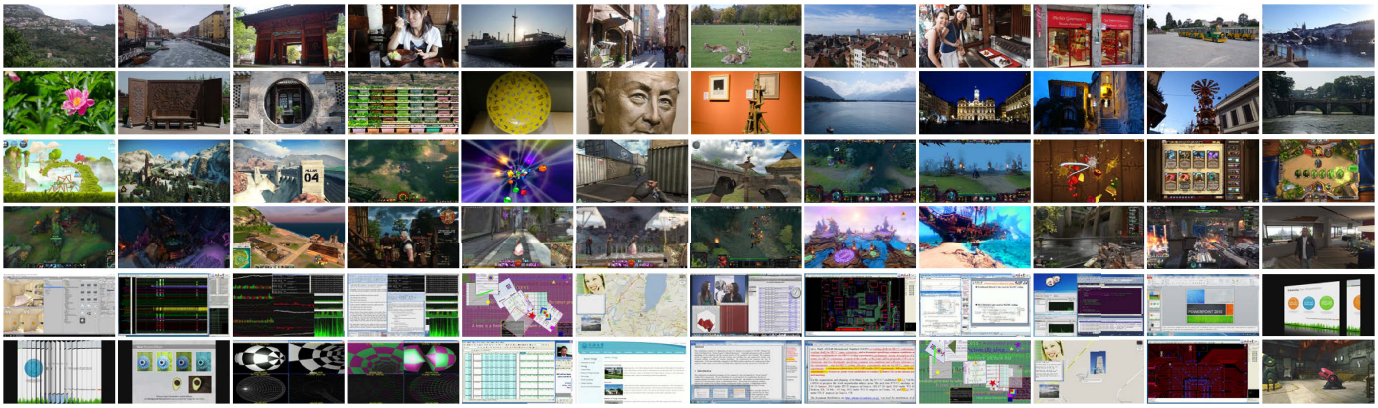


Fig. 1. Reference images in the CCT database. Top two rows: NSIs. Middle two rows: CGIs. Bottom two rows: SCIs.

noting that all these methods were designed and tested on SCIs only, but not in a cross-content-type environment. Very few work considered cross-content-type IQA. Xu *et al.* [18] developed a measure for NSIs, SCIs, and document images which are captured from document files. Min *et al.* [19] proposed a measure for JPEG compressed NSIs and SCIs.

In practice, many causes could lead to image degradations, and the major distortions depend on the application scenarios. For example, photographic NSIs may suffer from degradations caused by noisy sensing, out-of-focus or motion blur, and imperfect lighting conditions. CGIs may suffer from rendering artifacts, such as shadow, noise, and contrast distortion. SCIs sometimes contain NSIs and CGIs at local regions, and may suffer from the degradations happen in both NSIs and CGIs. The major application scenario we have in mind is to enable thin-clients to enjoy the computationally intensive and graphically rich services by instantly transmitting the remote computer interface to the clients. Such interface is a mixture of different types of image/video content (natural, graphical, and screen content), which mostly suffer from practical video compression artifacts. Therefore, in this paper, we consider HEVC and HEVC-SCC compression as main sources of distortions.

HEVC is the latest video coding standard proposed by JCT-VC [20], which also recommends the HEVC-SCC extension, aiming to improve the compression efficiency specifically for screen content [5]. The block-based coding scheme in HEVC/HEVC-SCC may result in blocking artifacts [20], which can significantly degrade the perceptual quality. Many blockiness measures have been proposed especially for JPEG compressed images [19], [21]–[26]. However, those measures do not work well for block-based coding schemes such as HEVC/HEVC-SCC because of the varying block sizes and deblocking filters. Moreover, general-purpose no-reference (NR) IQA measures are also ineffective as will be clear later.

In this paper, we make one of the first attempts to develop a unified content-type adaptive (UCA) blind IQA measure for natural, graphic, and screen content images. We first construct a new IQA database, namely the cross-content-type (CCT) database that contains HEVC/HEVC-SCC compressed NSIs, CGIs, and SCIs, based on which we conduct a subjective experiment. CGIs are generally computer-generated, and are often created to simulate real-world scenes in many practical applications. Thus we take computer graphic content as a separate type in the database.

Considering that HEVC/HEVC-SCC have flexible block sizes and the human visual system (HVS) has receptive fields of various scales, we extract and integrate block-based corner and edge features at multiple scales adaptively. The relative importance across different scales is determined from the contrast sensitivity function (CSF) and the content dependent characteristics of the HVS, leading to an adaptive multi-scale weighting strategy. Experimental results on the CCT database show that the proposed UCA method well adapts to various content types and achieves competitive performance.

The construction of UCA does not involve any training process and is independent of specific block-based compression algorithms. Therefore, we also test UCA on JPEG compressed images, and MPEG-2, H.264 and HEVC compressed video sequences. Experimental results on public image and video databases show that UCA demonstrates strong generalizability and consistently obtains competitive performance against state-of-the-art NR image/video quality models.

The remainder of this paper is organized as follows. Section II describes the construction of the CCT database and the subjective user study. Section III presents in detail the proposed UCA measure. Experimental results are given in Section IV, where we compare UCA with state-of-the-art NR quality measures. In Section V, we test the generalizability of UCA on other block-based encoders. Section VI concludes this paper.

II. Subjective Quality Assessment of Natural, Graphic and Screen Content Images

Few previous work has been dedicated to cross-content-type IQA. To facilitate the research, we construct the CCT database and conduct a subjective study. We extend the SCD database in [3] by adding two more types: natural and graphic content.

A. Reference and Distorted Images

We include the following 3 types of images in the database.

- NSI: We select 24 high quality NSIs to cover a variety of image content including landscapes and cityscapes,

TABLE I
CONSTRUCTION OF THE CCT DATABASE

Content type	Number of images			Resolution
	Ref.	Dist.		
NSI	24	HEVC	24×11=264	1920×1080 1280×720
CGI	24	HEVC	24×11=264	
		SCC	24×11=264	
SCI	24	HEVC	24×11=264	
		SCC	24×11=264	

indoor and outdoor views, closeups and wide-angle shots. Some images are from [27] and others are captured by ourselves.

- CGI: We collect 24 reference CGIs from 15 computer games via screen snapshots. The chosen games include various types, *e.g.*, action, adventure, and strategy.
- SCI: 23 pristine SCIs are from the SCD database [3] and another one is captured by ourselves through screen snapshot. The selected SCIs provide a good coverage of general computer operation scenes, *e.g.*, web pages, documents, desktops, and softwares.

Fig. 1 illustrates all 72 reference NSIs, CGIs, and SCIs, whose resolutions are either 1920×1080 or 1280×720. We mainly consider the following two distortion types.

- HEVC [20]: All reference NSIs, CGIs, and SCIs are compressed using HEVC intra coding with 11 quality levels from 30 to 50 at an interval of 2. HEVC adopts a 4:2:0 color format, so we downsample all reference images from 4:4:4 to 4:2:0 format before compression.
- HEVC-SCC: As an extension of HEVC, HEVC-SCC provides better compression performance for screen content [5]. The reference CGIs and SCIs are compressed using HEVC-SCC intra coding with 11 quality levels. Since HEVC-SCC supports 4:4:4 format, downsampling is not performed.

For both HEVC and HEVC-SCC, we use the reference software, *i.e.*, HM (HEVC Test Model) [28] along with the provided configuration files. In summary, a total of 72 reference and 1,320 distorted NSIs, CGIs, and SCIs are included in the CCT database, whose details are summarized in Table I.

B. Subjective Testing Methodology

We adopt a single-stimulus strategy in which subjects rate one test image at any time instance directly using a 10-point numerical categorical judgement method. All test images are shown in a random order with a MATLAB interface on a 23-inch LED monitor, which is calibrated according to the recommendations of ITU-R BT.500-13 [29]. A total of 43 subjects (28 males and 15 females) participate in the subjective experiment. Subjects are seated at a viewing distance of around 3 times the screen height in a laboratory environment which has normal indoor illumination levels. The full experiment is divided into 4 sessions, each of which lasts less than 30 minutes. Each subject only takes part in two sessions. We summarize the key points of the subjective testing in Table II.

TABLE II
SUBJECTIVE EXPERIMENT SETUP

Category	Item	Detail
Monitor	Model	Dell U2312HM / LED / 23 in
	Resolution	1920×1080
Methodology	Method	Single-stimulus
	Quality-scale	10-grade category
	Presentation order	Random
Test settings	Sessions	4
	Subjects in total	40 valid / 3 outliers
	Subjects per image	20 valid
	Viewing distance	3 times screen height

C. Data Processing

We employ the outlier detection and subject rejection algorithm suggested in [29]. Specifically, the raw score for an image is considered to be an outlier if it is outside 2 standard deviations (stds) about the mean score of that image for the Gaussian case or outside $\sqrt{20}$ stds for the non-Gaussian case. A subject is removed if more than 5% of his/her evaluations are outliers. As a result, 3 subjects are rejected, and each image is rated by 20 valid subjects. Among all scores given by the remaining valid subjects, about 2.61% of the total subjective evaluations are identified as outliers and are subsequently removed. We calculate the difference mean opinion score (DMOS) as the ground truth. The raw rating r_{ij} (given by the ith subject to the jth image) is first converted to the raw difference d_{ij} by subtracting it from the raw rating of the corresponding reference image. d_{ij} is then converted to a Z-score: $z_{ij} = \frac{d_{ij}-d_i}{\sigma_i}$, where d_i is the mean difference score given by the ith subject, and σ_i is the std. Z-scores are then averaged across subjects: $z_j = \frac{1}{N_j}\sum_{i=1}^{N_j} z_{ij}$, where N_j is the number of valid ratings for the jth image. Finally, averaged Z-scores are linearly rescaled to the range of [0, 100] to obtain the final DMOSs, with higher values indicating worse quality.

III. THE PROPOSED UCA MODEL

A diagram of the proposed UCA method is shown in Fig. 2. UCA is composed of two main steps: a feature extraction process executed at multiple scales and an adaptive multi-scale weighting process that pools the results into a single quality score.

A. Feature Extraction

In spatial domain, corners and edges are presumably the most important image features. They are sensitive to various image distortions. For example, in block-based image/video compression schemes, because of block partitioning and the relatively independent processing between individual blocks, corners and edges near the boundaries of blocks change in different ways as compared to the middle regions of the blocks. More specifically, new corners and edges may be created near block boundaries while genuine corners and edges may be smoothed within blocks. Such disparities increase with the compression level. Therefore, we detect the statistical differences of corners and edges between block boundaries

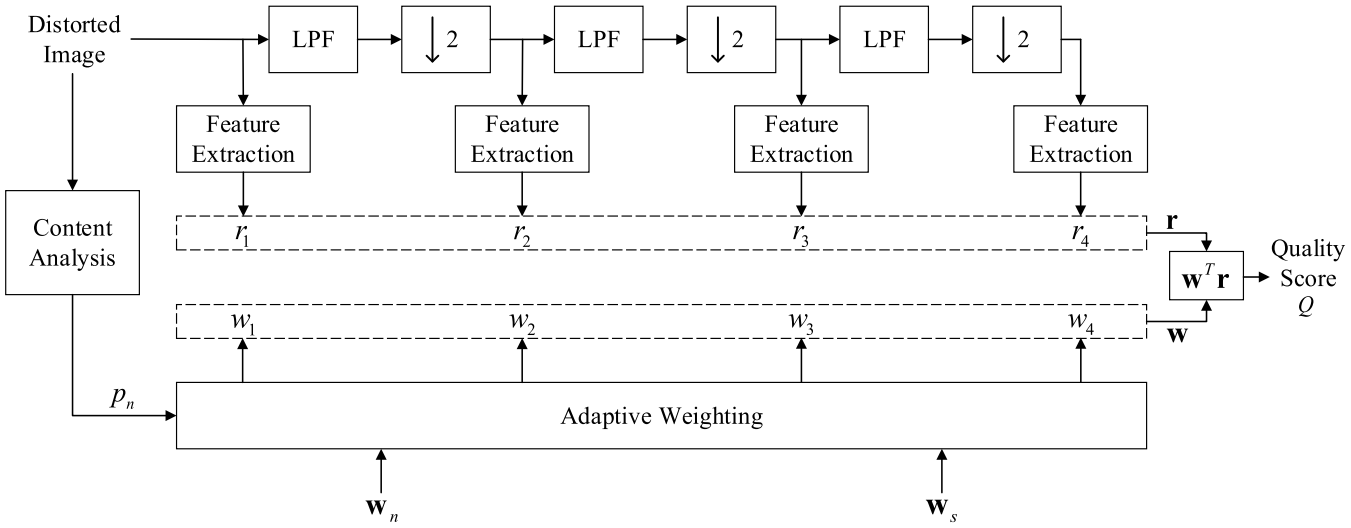


Fig. 2. Diagram of the proposed UCA method. LPF: low-pass filtering. ↓ 2: downsampling by a factor of 2. p_n: the likelihood of a distorted image being NSI. $\mathbf{w}_n$: multi-scale weights for NSI. $\mathbf{w}_s$: multi-scale weights for SCI. $\mathbf{w}$: final multi-scale weights. $\mathbf{r}$: feature vector.

and middle regions. We find this a simple but effective way to gauge the perceptual quality of compressed images.

1) Corner Feature: We adopt Shi and Tomasi's minimum eigenvalue method [30] to detect corners. Given an image $\mathbf{I}(x, y)$, corners are detected by maximizing the weighted sum of squared differences

$$\mathbf{D}(u, v) = \sum_{x,y} \mathbf{a}(x, y)[\mathbf{I}(x + u, y + v) - \mathbf{I}(x, y)]^2, \quad (1)$$

where (u, v) is the spatial shift, and $\mathbf{a}(x, y)$ is a weighting window. Applying a Taylor series expansion to $\mathbf{I}(x+u, y+v)$, $\mathbf{D}(u, v)$ can be approximated by

$$\mathbf{D}(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} \mathbf{H} \begin{bmatrix} u \\ v \end{bmatrix}, \quad (2)$$

where $\mathbf{H}$ is a Harris matrix

$$\mathbf{H} = \sum_{x,y} \mathbf{w}(x, y) \begin{bmatrix} \mathbf{I}_x\mathbf{I}_x & \mathbf{I}_x\mathbf{I}_y \\ \mathbf{I}_x\mathbf{I}_y & \mathbf{I}_y\mathbf{I}_y \end{bmatrix}. \quad (3)$$

Presumably, a corner should have larger differences in $\mathbf{D}(u, v)$ along all directions defined by (u, v), which means the Harris matrix $\mathbf{H}$ should have two large eigenvalues. Therefore, a corner metric is given by

$$\mathbf{H}_\lambda = \min(\lambda_1, \lambda_2), \quad (4)$$

where λ_1 and λ_2 are eigenvalues of $\mathbf{H}$. Then a corner map can be created by

$$\mathbf{C} = (c_{ij})_{h\times w} = \mathrm{BW}(\mathbf{H}_\lambda), \quad (5)$$

where BW is a threshold-based binarization function. $\mathbf{C}$ is the final binary corner map, in which $c_{ij} = 1$ indicates that a corner is detected at position (i, j). h and w denote the image height and width, respectively.

To single out the corners distributed at the block boundaries, we create a mask map $\mathbf{M} = (m_{ij})_{h\times w}$, whose elements are

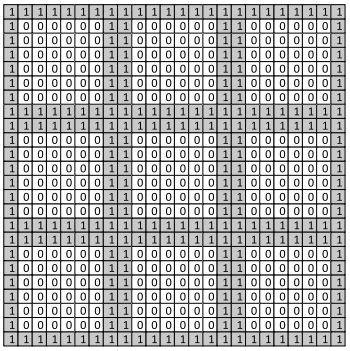


Fig. 3. Illustration of the block boundary mask $\mathbf{M} = (m_{ij})_{h\times w}$ for a 24×24 image with 8×8 blocks. Gray pixels denote $m_{ij} = 1$ and white pixels denote $m_{ij} = 0$.

defined as

$$m_{ij} = \begin{cases} 1 & \text{if } \mathrm{mod}(i, N) < 2 \text{ or } \mathrm{mod}(j, N) < 2 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where i and j are the row and column indices, mod calculates the remainder, and N is the block size. Fig. 3 gives an illustration of the mask $\mathbf{M}$ for a 24×24 image with 8×8 blocks. Then corners identified at the block boundaries can be represented by

$$\mathbf{C}' = (c'_{ij})_{h\times w} = \mathbf{C} \circ \mathbf{M} = (c_{ij} \cdot m_{ij})_{h\times w}, \quad (7)$$

where $\circ$ denotes the Hadamard product. Our corner feature is computed as the ratio of the corners at the block boundaries

$$r_c = \frac{\sum_{i,j} c'_{ij}}{\sum_{i,j} c_{ij}}. \quad (8)$$

2) Edge Feature: Given an image $\mathbf{I}$, its gradient magnitude can be calculated as

$$\mathbf{G} = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}, \quad (9)$$

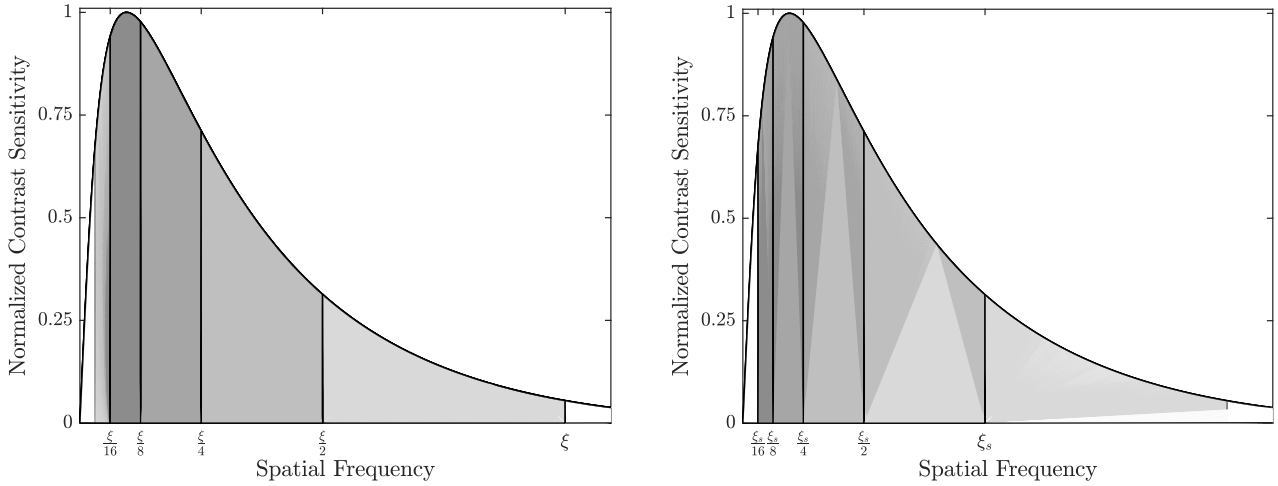


Fig. 4. CSFs and multi-scale weights for natural and screen content. (a) Multi-scale weights for natural content. (b) Multi-scale weights for screen content.

where $\mathbf{G}_x$ and $\mathbf{G}_y$ are the partial derivatives along horizontal and vertical directions, respectively. An edge map is created by

$$\mathbf{E} = (e_{ij})_{h\times w} = \mathrm{BW}(\mathbf{G}), \tag{10}$$

where $\mathbf{E}$ is a binary map and $e_{ij} = 1$ indicates that an edge pixel is detected at (i, j). We use the same mask map $\mathbf{M}$ defined in Eq. (6) to identify edges at the block boundaries

$$\mathbf{E}' = (e'_{ij})_{h\times w} = \mathbf{E} \circ \mathbf{M} = (e_{ij} \cdot m_{ij})_{h\times w}. \tag{11}$$

The edge feature is given by the ratio of the edges at the block boundaries

$$r_e = \frac{\sum_{i,j} e'_{ij}}{\sum_{i,j} e_{ij}}. \tag{12}$$

3) Overall Quality Feature: As illustrated in Fig. 3, block boundaries occupy $R = \frac{4(N-1)}{N^2}$ of an image. Presumably, a high quality image free of block-based compression should have r_c and r_e values close to R. With the increase of the compression level, a large proportion of corners and edges are detected near block boundaries, and thus r_c and r_e increase. We compute the overall quality/distortion feature as the product of r_c and r_e normalized by R^2 such that high quality images have a feature value close to 1

$$r = \frac{r_c r_e}{R^2} = \frac{N^4 r_c r_e}{16(N-1)^2}. \tag{13}$$

The feature extraction process is executed at 4 scales, forming a feature vector $\mathbf{r} = [r_1, r_2, r_3, r_4]^T$, where r_i is the overall quality feature at the ith scale derived from Eq. (13).

B. Adaptive Multi-Scale Weighting

Many existing work [31]–[33] has shown that multi-scale analysis improves quality prediction performance. Here we propose a multi-scale framework and focus on the impact of the human psychological behaviors and the visual sensitivity characteristics across scales.

1) Multi-Scale Weights for Natural Content: We divide an image into 4 scales by iterative low-pass filtering and downsampling. Quality features are extracted at each scale and integrated linearly using a set of multi-scale weights. Following the strategies adopted in [32] and [33], we derive the weights of each scale from the CSF of the HVS.

We first define a viewing resolution factor ξ as

$$\xi = \frac{\pi \cdot d \cdot n}{180 \cdot h_s \cdot 2}, \tag{14}$$

where the unit of ξ is cycles per degree of visual angle (cpd), d is the viewing distance (inch), h_s is the height of the screen (inch), and n denotes the number of pixels in the vertical direction of the screen. We use ξ to divide the spatial frequency range for each scale, which covers one section of the CSF formulated by [34]

$$S(u) = \frac{5200e^{(-0.0016u^2(1+100/L)^{0.08})}}{\sqrt{(1+\frac{144}{X_0^2}+0.64u^2)(\frac{63}{L^{0.83}}+\frac{1}{1-e^{(-0.02u^2)}})}}, \tag{15}$$

where u, L, and X_0^2 indicate spatial frequency (cpd), luminance (cd/m^2), and angular object area (squared degrees), respectively.

As illustrated in Fig. 4(a), each scale corresponds to one spatial frequency range. For example, the finest scale corresponds to $[\frac{\xi}{2}, \xi]$ and so on. The weight for each scale is calculated as the area under the CSF within the corresponding frequency covering range

$$w_i = \frac{1}{Z}\int_{\frac{\xi}{2^i}}^{\frac{\xi}{2^{i-1}}} S(u)\mathrm{d}u, \quad i \in \{1, 2, 3, 4\}, \tag{16}$$

where i from 1 to 4 corresponds the finest to coarsest scale, respectively. Z is a normalization factor such that $\sum_i w_i = 1$.

2) Multi-Scale Weights for Screen Content: Previous studies in the literature [12], [35], [36] suggest that humans have significantly different behaviors when viewing natural scene and screen content. Unlike natural scene perception, when viewing screen content, the HVS tends to be more focused

so as to perform tasks better suited to extracting the type of information conveyed in screen content, such as reading texts. Rayner [35], [36] reviewed the viewing behaviors under different conditions. Two important aspects are saccade length and perceptual span. The saccade length in natural scene perception is generally 2 times as many as screen content perception. The perceptual span of natural scene perception is also larger (which is around 2 if under the same task of recognition). Moreover, the attentional selection and shift are often object-based and the "object size" of natural scene approximately doubles that of screen content.

To account for such behaviors in our model and to achieve equivalence in viewing natural scene and screen content, we choose to set screen content's equivalent viewing distance as half of its actual viewing distance ($\xi_s = \frac{\xi}{2}$). Replacing ξ with ξ_s in Eq. (16), we derive a group of multi-scale weights $\mathbf{w}_s$ for screen content. As illustrated in Fig. 4(b), more weights are given to finer scales, which is expected since observers are more focused at high-frequency details.

3) Adaptive Weighting: In practice, it is sometimes difficult to make a crisp judgement about if an image whether a natural scene or screen content, and sometimes the image is a mixture of both. This is often the case for computer graphic images that mimic natural scenes. Here we propose an adaptive weighting strategy that bridges between the two cases

$$\mathbf{w} = p_n \cdot \mathbf{w}_n + (1 - p_n) \cdot \mathbf{w}_s, \tag{17}$$

where $\mathbf{w}_n$ and $\mathbf{w}_s$ are the multi-scale weight vectors for NSIs and SCIs, respectively, and p_n is the likelihood that the image is an NSI.

It remains to determine p_n. This requires certain image features that are sensitive to image types but robust to distortion types and levels. We find the variance of local variance (VOLV) serves this purpose well. In particular, we compute the variance of local image patches. We then compute the variance of all local variances in the image. To test how effective VOLV is at differentiating image types, we collect another set of high quality images including 200 NSIs and 100 SCIs, which are not included in the CCT database, and compress them using HEVC/HEVC-SCC with 11 compression levels. Fig. 5 illustrates the histograms of VOLV of 2,200 compressed NSIs and 2,200 compressed SCIs, from which we see that a good separation is achieved.

Moreover, we find that the distributions of both content types generally follow the gamma distribution, whose probability density function (PDF) is given by

$$f(\upsilon; \alpha, \beta) = \frac{\beta^\alpha \upsilon^{\alpha-1} e^{-\upsilon\beta}}{\Gamma(\alpha)}, \quad \text{for } \upsilon \geq 0 \text{ and } \alpha, \beta > 0, \tag{18}$$

where υ is the VOLV value, α, β are two shape parameters, and $\Gamma(\cdot)$ is the gamma function. The fitted curves are also given in Fig. 5. We compute p_n as

$$\begin{aligned} p_n &= p(t = 0|\upsilon) = \frac{p(t = 0, \upsilon)}{p(\upsilon)} \\ &= \frac{p(\upsilon|t = 0) \cdot p(t = 0)}{p(\upsilon|t = 0) \cdot p(t = 0) + p(\upsilon|t = 1) \cdot p(t = 1)}, \end{aligned} \tag{19}$$

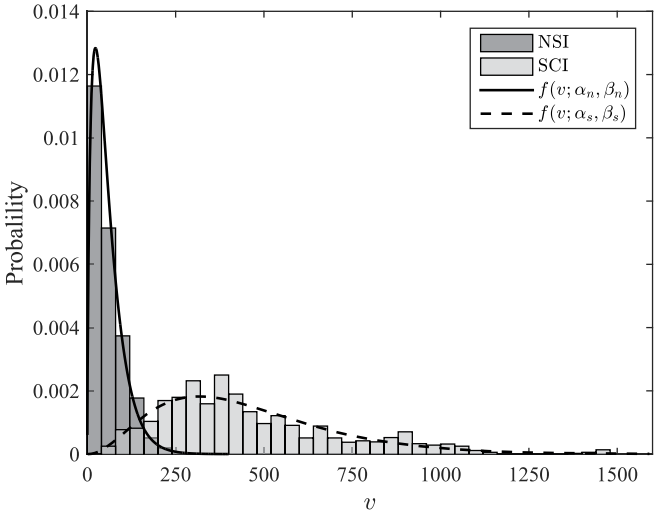


Fig. 5. Histograms of variance of the local variance. Natural scene and screen content are well separated and the histograms are well fitted by the gamma probability density function.

where $t = 0$ and $t = 1$ indicate that the content type is an NSI and SCI, respectively.

Since we do not have any prior knowledge about the content type, we assume $p(t = 0) = p(t = 1) = 0.5$. Moreover, we assume that the VOLV value υ for each content type follows a gamma PDF

$$p(\upsilon|t = 0) = f(\upsilon; \alpha_n, \beta_n), \tag{20}$$

$$p(\upsilon|t = 1) = f(\upsilon; \alpha_s, \beta_s), \tag{21}$$

where (α_n, β_n) and (α_s, β_s) are fitting parameters for NSIs and SCIs, respectively. As a result,

$$p_n = \frac{f(\upsilon; \alpha_n, \beta_n)}{f(\upsilon; \alpha_n, \beta_n) + f(\upsilon; \alpha_s, \beta_s)}. \tag{22}$$

Although we do not apply an explicit content type classification in our method, we test the discrimination ability of p_n on the CCT database. Specifically, we classify the content type through hard-thresholding, *i.e.*, an image is classified as NSI if $p_n \geq 0.5$. We have accuracies of 92.8% and 88.3% on NSIs and SCIs, respectively. On CGIs, 87.5% of the images are classified as NSIs because they simulate natural scenes.

Finally, the overall quality Q is the weighted sum given by

$$Q = \mathbf{w}^T \mathbf{r}. \tag{23}$$

C. Implementation Details

In the current implementation, UCA works with gray-scale images only. To detect corners, we apply a 3×3 Gaussian filter with a std of 0.5 as a preprocessing. The threshold in Eq. (5) is set to

$$\tau = 0.0005 \cdot \max(\mathbf{H}_\lambda), \tag{24}$$

where $\max(\cdot)$ finds the maximum of the corner metric matrix $\mathbf{H}_\lambda$ given by Eq. (4). For edge detection, we adopt the Prewitt operation. The corresponding threshold is set to 2. Similar

TABLE III
THE DETAILS OF COMPETING NR IQA MODELS. NSS: NATURAL SCENE STATISTICS. DCT: DISCRETE COSINE TRANSFORMATION. DWT: DISCRETE WAVELET TRANSFORM

Index	Model	Type	Feature extraction methodology	Training required
A	DIIVINE [37]	General-purpose	Spatial NSS	YES
B	BLIINDS-II [38]		DCT domain NSS	YES
C	BRISQUE [39]		Spatial NSS	YES
D	NFERM [40]		Free energy principle	YES
E	CORNIA [41]		Unsupervised learning of linear filters	YES
F	HOSA [18]		High order statistics	YES
G	NIQE [42]		Spatial NSS	YES
H	IL-NIQE [43]		Spatial NSS	YES
I	LPSI [44]		Local pattern statistics	NO
J	Wang2000 [21]	Distortion-specific (blockiness)	Fourier magnitude pattern detection	NO
K	Bovik [22]		Block-DCT discontinuity detection	NO
L	Wang2002 [23]		Spatial activity variation	YES
M	Perra [24]		Edge detection	NO
N	Li [25]		Spatial regularity detection	YES
O	RMB [26]		Tchebichef moments	NO
P	BIBLE [45]	Distortion-specific (blur)	Discrete orthogonal moments	NO
Q	CPBD [46]		Cumulative probability of blur detection	NO
R	FISH [47]		Log-energies of DWT subbands	NO
S	S_3 [48]		Spectral and spatial properties	NO
T	LPC-SI [49]		Local phase coherence	NO
U	UCA	Distortion-specific (Block-based compression)	Corner and edge detection / adaptive weighting	NO

performance is obtained using other edge detectors, such as Sobel and Canny [50] methods.

The parameters in computing the viewing resolution factor in Eq. (14) and the CSF in Eq. (15) are set based on our subjective experiment. These are $d = 35$, $n = 1,080$, $h \approx 11.3$, $L = 200$, and $X_0^2 \approx 606$. The final multi-scale weights for natural and screen content are $\mathbf{w}_n = [0.2066, 0.3329, 0.2855, 0.1749]^T$ and $\mathbf{w}_s = [0.3858, 0.3309, 0.2026, 0.0807]^T$, respectively. To compute VOLV, we adopt a 7×7 Gaussian filter. The fitted parameters in the gamma PDFs for NSIs and SCIs are $\alpha_n = 1.6876$, $\beta_n = 33.3924$ and $\alpha_s = 3.2516$, $\beta_s = 140.6982$, respectively.

IV. EXPERIMENTAL RESULTS

A. Experimental Settings

The proposed UCA method is tested on the CCT database, which contains a total of 1,320 compressed images divided into 5 subsets, as listed in Table I. Following the practices in [51], we first map the predicted scores nonlinearly using a five-parameter logistic function

$$q(s) = \epsilon_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\epsilon_2 (s - \epsilon_3)}} \right) + \epsilon_4 s + \epsilon_5, \qquad (25)$$

where $\{\epsilon_i | i = 1, 2, \ldots, 5\}$ are parameters to be fitted. We then compare the mapped scores with DMOSs using Spearman rank-order correlation coefficient (SRCC) and Pearson linear correlation coefficient (PLCC). A higher SRCC or PLCC value indicates better performance.

We choose a variety of objective NR IQA models to cover different design methodologies. These include state-of-the-art general-purpose models DIIVINE [37], BLIINDS-II [38], BRISQUE [39], NFERM [40], CORNIA [41], HOSA [18], NIQE [42], IL-NIQE [43], and LPSI [44]. Since HEVC/HEVC-SCC compressed images mainly suffer from blocking and blurring artifacts, we also include top performing distortion-specific (blockiness and blur) NR models, including Wang2000 [21], Bovik [22], Wang2002 [23], Perra [24], Li [25], RMB [26], BIBLE [45], CPBD [46], FISH [47], S_3 [48], and LPC-SI [49]. A summary of the competing models is given in Table III. We implement our own versions of Wang2000 [21], Bovik [22], and Perra [24], while the implementations of the rest algorithms are obtained from the original authors.

For methods that involve training (DIIVINE [37], BLIINDS-II [38], BRISQUE [39], and NFERM [40]), we retrain them using the CCT database. Specifically, given a target set (one of the five subsets or the whole CCT database), we randomly select 80% reference images along with their distorted versions as the training set and leave the rest 20% as the test set. Note that images corresponding to the same reference image are assigned to the same set to ensure the complete independence of the training and test data. If the target set is the whole database, we conduct such random separation for every subset to keep a balance among all subsets. We repeat this procedure 1,000 times and report the median SRCC and PLCC results. In order to make a fair comparison, we also apply this splitting procedure for the training-free methods. In other words, results only on the test sets are reported.

B. Performance Comparison

Table IV summarizes the test results, from which we have several useful findings. First, training-based models exhibit advantages over training-free models. Second, models based on blockiness and blur estimation may perform reasonably in one subset, but fail to align different content types well, leading to a performance drop on the whole database. Third, there is a clear trend of performance drop from NSIs and CGIs to SCIs, which is the case for almost all models. As a result, the SCI subset appears to be extremely difficult

TABLE IV

MEDIAN SRCC AND PLCC PERFORMANCE ACROSS 1,000 TRAILS ON 5 SUBSETS AND THE WHOLE CCT DATABASE. TIME: SECONDS/IMAGE

Index	Model	NSI		CGI				SCI				All		Time
		HEVC		HEVC		SCC		HEVC		SCC				
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	
A	DIIVINE	0.6923	0.7381	0.6740	0.7184	0.5833	0.6749	0.4308	0.5473	0.3737	0.5291	0.5083	0.5551	26.431
B	BLIINDS-II	0.8543	0.8743	0.8057	0.8462	0.8457	0.8649	0.6856	0.7429	0.5267	0.6446	0.8229	0.8277	79.410
C	BRISQUE	0.4847	0.5716	0.6629	0.7272	0.6745	0.7379	0.5935	0.6943	0.5458	0.6752	0.6164	0.6591	0.6262
D	NFERM	0.7450	0.7894	0.5850	0.6764	0.6260	0.7081	0.6425	0.7109	0.4413	0.5702	0.6969	0.7260	82.603
E	CORNIA	0.7096	0.7225	0.7796	0.8110	0.7713	0.8014	0.2543	0.5082	0.2585	0.5241	0.4871	0.5811	3.5770
F	HOSA	0.8202	0.8348	0.8067	0.8182	0.7922	0.8167	0.0030	0.4051	0.0503	0.4028	0.2967	0.5148	0.2600
G	NIQE	0.7209	0.7580	0.7543	0.7928	0.7380	0.7933	0.2373	0.5016	0.2052	0.4487	0.2635	0.5479	0.3840
H	IL-NIQE	0.6187	0.6989	0.7262	0.7705	0.7196	0.7694	0.2815	0.4651	0.2875	0.4531	0.1660	0.5007	3.6150
I	LPSI	0.7931	0.8364	0.8093	0.8536	0.7833	0.8448	0.2952	0.5452	0.1115	0.3627	0.4578	0.5130	0.0509
J	Wang2000	0.0853	0.3336	0.1436	0.3848	0.1853	0.4008	0.2586	0.4607	0.3203	0.4819	0.0580	0.2800	0.3618
K	Bovik	0.3802	0.4444	0.4764	0.4985	0.3337	0.4304	0.3175	0.4819	0.3111	0.4777	0.0638	0.2993	1.7952
L	Wang2002	0.7106	0.7430	0.7841	0.8087	0.7644	0.8217	0.1583	0.5410	0.1709	0.5393	0.2805	0.5741	**0.0398**
M	Perra	0.3775	0.5078	0.5573	0.6320	0.5543	0.6192	0.3606	0.5938	0.3278	0.5677	0.0762	0.4313	0.1277
N	Li	0.5838	0.6169	0.6525	0.6785	0.5873	0.6331	0.1996	0.4930	0.2569	0.4912	0.3948	0.4880	0.5372
O	RMB	0.5665	0.6448	0.6856	0.7067	0.7009	0.7222	0.1282	0.4650	0.1541	0.4585	0.1780	0.4604	2.6324
P	BIBLE	0.7662	0.7891	0.7240	0.7528	0.6572	0.7239	0.5062	0.6403	0.2847	0.5856	0.3677	0.4623	5.1047
Q	CPBD	0.7579	0.7794	0.7564	0.7821	0.7495	0.7877	0.7077	0.7564	0.6322	0.7349	0.4143	0.5326	0.3077
R	FISH	0.6557	0.7097	0.6287	0.6978	0.5804	0.6926	0.5275	0.6221	0.4729	0.5633	0.3610	0.4470	0.0480
S	S_3	0.5473	0.6402	0.2940	0.5075	0.2607	0.4994	0.2294	0.4752	0.1659	0.4864	0.1353	0.3051	23.592
T	LPC-SI	0.1716	0.4102	0.2759	0.4665	0.2593	0.4548	0.0742	0.4263	0.0918	0.4604	0.1264	0.2608	1.5491
U	UCA	**0.9043**	**0.9170**	**0.9052**	**0.9149**	**0.8815**	**0.9025**	**0.8803**	**0.8997**	**0.8667**	**0.8985**	**0.8387**	**0.8400**	0.2531

TABLE V

SRCC PERFORMANCE OF DIFFERENT FEATURES AND WEIGHTING STRATEGIES

Settings	NSI	CGI		SCI		All
	HEVC	HEVC	SCC	HEVC	SCC	
UCA_c	0.8685	0.8492	0.8284	0.8247	0.7869	0.7896
UCA_e	0.8625	0.8678	0.8383	0.7606	0.7007	0.7958
UCA	0.8857	0.8789	0.8598	0.8630	0.8516	0.8345
$\mathbf{w}_{c_1}$	0.8040	0.8652	0.8363	0.8169	0.7832	0.7961
$\mathbf{w}_{c_2}$	0.8737	0.8769	0.8540	0.8689	0.8384	0.8233
$\mathbf{w}_{c_3}$	0.8895	0.8817	0.8610	0.8610	0.8490	0.8247
$\mathbf{w}_{c_4}$	0.8866	0.8795	0.8595	0.8602	0.8524	0.8242
$\mathbf{w}_{c_5}$	0.8809	0.8763	0.8575	0.8502	0.8372	0.8198
$\mathbf{w}_{c_6}$	0.8857	0.8789	0.8598	0.8630	0.8516	0.8345

TABLE VI

SRCC PERFORMANCE AS A FUNCTION OF THE THRESHOLD τ

Feature	τ	NSI	CGI		SCI		All
		HEVC	HEVC	SCC	HEVC	SCC	
Corner	0.0001	0.8537	0.8656	0.8466	0.8377	0.8383	0.8220
	0.0003	0.8812	0.8763	0.8563	0.8603	0.8520	0.8328
	0.0005	0.8857	0.8789	0.8598	0.8630	0.8516	0.8345
	0.0007	0.8874	0.8819	0.8608	0.8616	0.8465	0.8349
	0.0009	0.8887	0.8831	0.8588	0.8599	0.8435	0.8345
	0.0011	0.8898	0.8834	0.8581	0.8560	0.8420	0.8334
Edge	1.2	0.8636	0.8792	0.8620	0.8684	0.8597	0.8306
	1.6	0.8813	0.8810	0.8613	0.8661	0.8570	0.8351
	2.0	0.8857	0.8789	0.8598	0.8630	0.8516	0.8345
	2.4	0.8883	0.8771	0.8560	0.8597	0.8453	0.8335
	2.8	0.8878	0.8741	0.8515	0.8556	0.8380	0.8299
	3.2	0.8873	0.8709	0.8470	0.8536	0.8322	0.8274

for most models. Finally, the proposed UCA model, which is training-free, achieves the best performance in all cases, and creates the least performance drop from NSI to SCI subsets.

C. Statistical Significance Analysis

Fig. 6 illustrates the mean and std of the SRCC values across the 1,000 trials for all algorithms listed in Table IV. The mean performance follows a similar trend as in Table IV. Besides, the std of UCA is smaller than competing models, which suggests that UCA performs more consistently.

To testify whether the differences between any two models are statistically significant, we perform a t-test [52] on SRCC values obtained by different models. After comparing every possible pairs of models, the results are summarized in Fig. 7, where a white/black block indicates that the row algorithm is statistically better/worse than the column algorithm, while a gray block indicates that the row and column algorithms are statistically indistinguishable. It can be observed that UCA is statistically superior to all competing algorithms on all 5 subsets and the whole CCT database.

D. Ablation Experiment

We conduct a series of ablation experiments to single out the core contributors of UCA. First, we only adopt corner features to obtain a model denoted by UCA_c. A similar model can be obtained with edge features only, denoted by UCA_e. SRCC performance results of UCA_c, UCA_e, and UCA are listed in Table VI, from which we observe that feature combination achieves the best performance. Second, we test six different weighting strategies. These are $\mathbf{w}_{c_1} = [1, 0, 0, 0]$, $\mathbf{w}_{c_2} = [0.25, 0.25, 0.25, 0.25]$, $\mathbf{w}_{c_3} = \mathbf{w}_s$, $\mathbf{w}_{c_4} = 0.5 \cdot \mathbf{w}_n + 0.5 \cdot \mathbf{w}_s$, $\mathbf{w}_{c_5} = \mathbf{w}_n$, and $\mathbf{w}_{c_6} = p_n \cdot \mathbf{w}_n + (1 - p_n) \cdot \mathbf{w}_s$ (adopted by UCA), where the subscript c_i is the weighting strategy index.

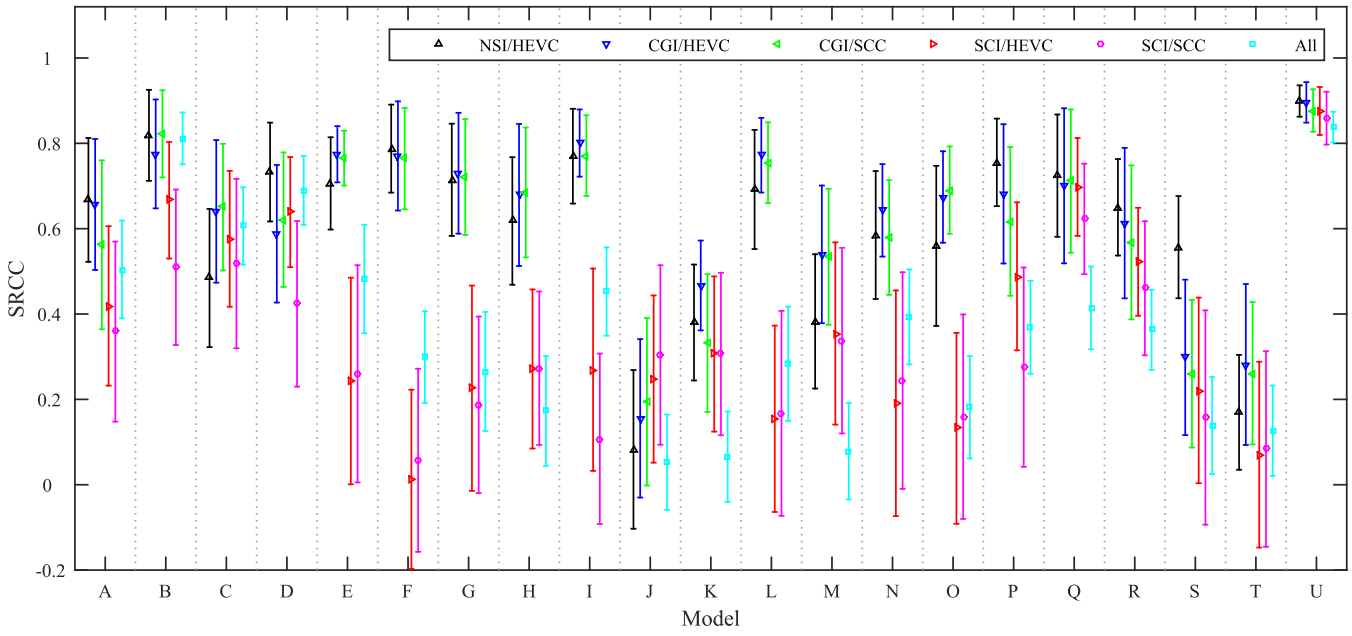


Fig. 6. Mean SRCC values and standard error bars for all competing algorithms across 1,000 train-test trials on 5 subsets and the whole CCT database. A-U are model indices given in Tables III & IV.

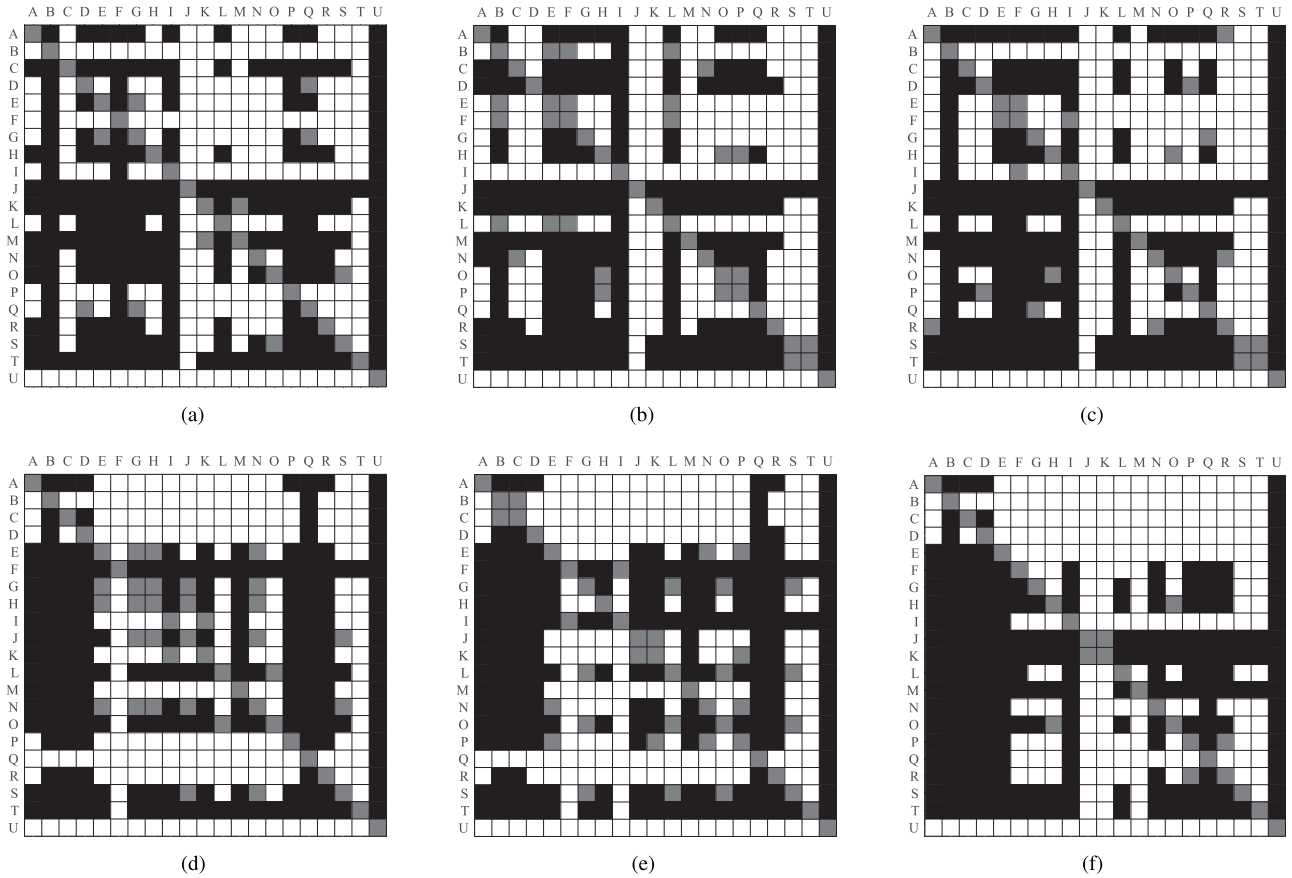


Fig. 7. Statistical significance test results on 5 subsets and the whole CCT database. A white/black block indicates that the row model is statistically better/worse than the column model. A gray block indicates that the row and column models are statistically indistinguishable. A-U are model indices given in Tables III & IV. (a) NSI/HEVC. (b) CGI/HEVC. (c) CGI/SCC. (d) SCI/HEVC. (e) SCI/SCC. (f) All.

The SRCC results are listed in Table VI, from which we observe that multi-scale weighting improves the performance and the CSF-based weighting is more effective than uniform weighting. Moreover, the proposed adaptive weighting is more effective than fixed weighting, especially on CGI and SCI subsets.

TABLE VII
PERFORMANCE COMPARISON ON JPEG IQA DATABASES

Index	Model	LIVE [53]		CSIQ [54]		TID2013 [55]		MICT [56]		SIQAD [2]		Average	
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
A	DIIVINE [37]	-	-	0.7996	0.8239	0.6288	0.6643	0.7023	0.7087	0.0818	0.1460	0.5531	0.5857
B	BLIINDS-II [38]	-	-	0.9004	0.9401	0.8643	0.9112	0.9062	0.8999	0.4107	0.4292	0.7704	0.7951
C	BRISQUE [39]	-	-	0.9040	0.9456	0.8453	0.8994	0.8695	0.8737	0.2671	0.2884	0.7251	0.7518
D	NFERM [40]	-	-	0.9223	0.9678	0.8720	0.9612	0.8642	0.8755	0.4168	0.4221	0.7688	0.8067
E	CORNIA [41]	-	-	-	-	0.8972	0.9354	0.7630	0.7774	0.2496	0.3215	0.6366	0.6781
F	HOSA [18]	-	-	0.9153	0.9569	0.8957	0.9181	0.8768	0.8920	0.3990	0.4017	0.7717	0.7922
G	NIQE [42]	0.9422	0.9516	0.8826	0.9360	0.8435	0.8904	0.8381	0.8456	0.4447	0.4476	0.7902	0.8142
H	IL-NIQE [43]	0.9407	0.9590	0.8993	0.9542	0.8346	0.8996	0.7069	0.7153	0.2874	0.3811	0.7388	0.7818
I	LPSI [44]	0.9677	0.9748	0.9502	0.9693	0.9123	0.9536	0.9182	0.9226	0.7149	0.7302	0.8926	0.9101
J	Wang2000 [21]	0.9068	0.9217	0.8688	0.8830	0.8734	0.8777	0.6341	0.6852	0.6779	0.6814	0.7922	0.8098
K	Bovik [22]	0.9476	0.9532	0.9454	0.9716	0.8821	0.9311	0.8140	0.8512	0.4655	0.4587	0.8109	0.8332
L	Wang2002 [23]	0.9735	0.9787	0.9551	0.9799	0.9267	0.9530	0.8829	0.8875	0.7452	0.7412	0.8967	0.9081
M	Perra [24]	0.8688	0.8712	0.8517	0.8927	0.7601	0.8170	0.7262	0.7525	0.2401	0.3108	0.6894	0.7288
N	Li [25]	-	-	0.9424	0.9612	0.8644	0.9156	0.8915	0.8945	0.6261	0.6206	0.8311	0.8480
O	RMB [26]	0.9689	0.9727	0.9533	0.9794	0.9134	0.9512	0.8802	0.8832	0.7491	0.7397	0.8930	0.8052
U	UCA	0.9466	0.9452	0.9285	0.9441	0.8548	0.8926	0.9111	0.9198	0.6925	0.6892	0.8667	0.8782

TABLE VIII
PERFORMANCE COMPARISON ON MPEG-2, H.264 AND HEVC VQA DATABASES

Model	VQEG [57]		CSIQ [58]				IVP [59]				LIVE [60]				Average	
	H.264		H.264		HEVC		H.264		MPEG-2		H.264		MPEG-2			
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
BRISQUE	0.3552	0.3855	0.6885	0.6768	0.4569	0.6053	0.2255	0.5421	0.5991	0.6293	0.1799	0.2169	0.2795	0.3103	0.3978	0.4809
NIQE	0.5082	0.6033	0.6386	0.6544	0.4368	0.5233	0.3008	0.3607	0.5964	0.6983	0.1285	0.4255	0.2832	0.3295	0.4132	0.5136
HOSA	0.6840	0.7060	0.7017	0.6979	0.5609	0.6065	0.4510	0.5792	0.7762	0.8046	0.1604	0.3315	0.3832	0.6291	0.5311	0.6221
LPSI	0.4603	0.5241	0.6605	0.6657	0.2064	0.2034	0.5463	0.6202	0.8572	0.8988	0.2362	0.3775	0.5934	0.6074	0.5086	0.5567
Wang2000	0.2575	0.3833	0.7004	0.7048	0.2039	0.3880	0.5983	0.5648	0.6013	0.5833	0.2927	0.4071	0.3872	0.4383	0.4345	0.4957
Wang2002	0.5518	0.6032	0.6582	0.6317	0.1501	0.2666	0.5685	0.5538	0.7086	0.7383	0.1906	0.4286	0.3827	0.4433	0.4586	0.5236
Perra	0.1808	0.3700	0.2265	0.2425	0.0054	0.0899	0.0311	0.3122	0.5978	0.7037	0.1745	0.3628	0.3723	0.4977	0.2269	0.3684
Li	0.2780	0.3868	0.4548	0.3723	0.0870	0.2723	0.3856	0.4899	0.4091	0.4411	0.3850	0.5698	0.3778	0.5930	0.3396	0.4465
CPBD	0.6108	0.6540	0.4723	0.5038	0.4190	0.4178	0.4876	0.5244	0.4100	0.5985	0.0220	0.3399	0.0569	0.3951	0.3541	0.4905
FISH	0.2413	0.2894	0.3748	0.4244	0.4595	0.5149	0.1779	0.3449	0.1003	0.1399	0.1884	0.2798	0.1900	0.2485	0.2475	0.3203
V-BLIINDS	0.0543	0.4091	0.3951	0.4843	0.0970	0.2818	0.0582	0.1984	0.1453	0.2601	-	-	-	-	0.1500	0.3267
VIIDEO	0.1571	0.3109	0.7153	0.7252	0.3467	0.3673	0.1621	0.4289	0.2685	0.2086	-	-	-	-	0.3299	0.4082
UCA	0.5995	0.6095	0.7035	0.7302	0.6641	0.6905	0.7036	0.7045	0.6432	0.6290	0.3653	0.4826	0.4677	0.4400	0.5924	0.6123

E. Parameter Sensitivity

Most parameters in UCA are inherited from previous publications or can be inferred from the subjective testing environment. Here, we only test the sensitivity of UCA with respect to two critical parameters—the thresholds of the binarization functions in Eq. (5) and Eq. (10), which specify the minimum accepted quality of the extracted corners and edges, respectively. We vary the corner threshold from 0.0001 to 0.0011, with a step of 0.0002, and fix the edge threshold at 2. Similarly we vary the edge threshold from 1.2 to 3.2 with a step of 0.4, while keeping the corner threshold fixed at 0.0005. From Table V, we see that the SRCC performance remains stable within a significantly wide range.

F. Computational Complexity

To empirically compare the computational complexity of state-of-the-art NR IQA models with UCA, we report their execution time for 100 images with a fixed resolution of 1280×720 on a computer with 4.00 GHz Intel Core i7-6700K CPU and 32 GB RAM. From Table IV, it appears that UCA is among the models with the lowest computational complexity.

V. GENERALIZABILITY OF UCA

The proposed UCA method is designed for block-based image/video compression. Therefore, it is interesting to observe how UCA generalizes to other block-based image/video compression methods and databases. We test it using publicly available JPEG image compression and MPEG-2/H.264/HEVC video compression databases.

A. JPEG Compressed Images

We test UCA on the JPEG subsets of 5 widely used IQA databases: LIVE [53], CSIQ [54], TID2013 [55], MICT [56], and SIQAD [2], which contain 175, 150, 125, 84, and 140 JPEG compressed images, respectively. All reference images are excluded. All models listed in Table III except the blur-specific ones are compared. Table VII lists the performance comparison results. It needs to be aware that many algorithms are trained or tuned based on some of these databases. For example, DIIVINE [37], BLIINDS-II [38], BRISQUE [39], NFERM [40], HOSA [18], Wang2002 [23], and Li [25] are trained on LIVE, and CORNIA [41] is trained on both LIVE and CSIQ. We do not list their performance on

the training databases in Table VII. It can be observed that UCA achieves high performance, competitive with state-of-the-art NR quality measures.

B. MPEG-2, H.264, and HEVC Compressed Videos

We test UCA on MPEG-2, H.264, and HEVC compressed videos using the relevant subsets of 4 public video quality assessment (VQA) databases—VQEGHD3 [57], CSIQ [58], IVP [59], and LIVE [60]. The VQEGHD3 subset contains 54 H.264 coded video sequences of 6 bit rates. The CSIQ subset contains 36×2 video sequences of 3 distortion levels, compressed by H.264 and HEVC, respectively. The IVP subset contains 40 sequences of 4 levels compressed by H.264 and 30 sequences of 3 levels compressed by MPEG-2, respectively. Finally, the LIVE subset contain 40×2 sequences of 4 levels, compressed by H.264 and MPEG-2, respectively.

Many NR methods under comparison are designed for images only, for which we simply average the scores given to all video frames as an overall quality evaluation. We incorporate two NR VQA models, namely V-BLIINDS [61] and VIIDEO [62] into comparison. Note that the two models are trained or tuned on LIVE, whose results are not listed in Table VIII. Compared with state-of-the-art distortion-specific and general-purpose IQA/VQA algorithms, UCA is quite competitive in all block-based coding schemes, especially for HEVC.

VI. Conclusion

In this paper, we propose a cross-content-type NR IQA measure for compressed natural, graphic, and screen content images by integrating a novel perceptually motivated content-adaptive multi-scale weighting strategy. The proposed UCA method is validated on a first-of-its-kind CCT database, which contains a total of 1,320 HEVC/HEVC-SCC compressed NSIs, CGIs, and SCIs. Experimental results show that UCA not only significantly outperforms state-of-the-art general-purpose and distortion-specific NR quality measures but also exhibits stronger robustness across both content and content type variations. Experimental results on several other IQA/VQA databases further demonstrate the generalizability of UCA.

The current UCA is dedicated to block-based image and video compression, largely due to its wide usage in practice. Nevertheless, the design principle of content-adaptive multi-scale weighting is general. In the future, we plan to extend it to other IQA/VQA application scenarios. The proposed UCA method may be applied to color images by measuring all color channels in certain color spaces. In typical image/video compression methods, all color channels are distorted in a similar fashion, and thus UCA should work effectively. However, specific color distortions may occur when the distortion is not restricted to compression. In those cases, UCA needs to be improved.

References

[1] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, "Screen content coding based on HEVC framework," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1316–1326, Aug. 2014.

[2] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408–4421, Nov. 2015.

[3] S. Wang *et al.*, "Subjective and objective quality assessment of compressed screen content images," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 532–543, Dec. 2016.

[4] T. Lin, P. Zhang, S. Wang, K. Zhou, and X. Chen, "Mixed chroma sampling-rate High Efficiency Video Coding for full-chroma screen content," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 173–185, Jan. 2013.

[5] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 50–62, Jan. 2016.

[6] 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.

[7] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

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

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

[10] 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.

[11] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.

[12] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Perceptual screen content image quality assessment and compression," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 1434–1438.

[13] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.

[14] Z. Ni, L. Ma, H. Zeng, C. Cai, and K.-K. Ma, "Gradient direction for screen content image quality assessment," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1394–1398, Oct. 2016.

[15] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "ESIM: Edge similarity for screen content image quality assessment," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4818–4831, Oct. 2017.

[16] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, and Z. Guo, "Objective quality assessment of screen content images by uncertainty weighting," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2016–2027, Apr. 2017.

[17] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, and A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.

[18] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sep. 2016.

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

[20] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[21] Z. Wang, A. C. Bovik, and B. L. Evan, "Blind measurement of blocking artifacts in images," in *Proc. Int. Conf. Image Process.*, Sep. 2000, pp. 981–984.

[22] A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2001, pp. 1725–1728.

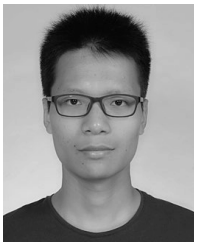
[23] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. Int. Conf. Image Process.*, Sep. 2002, pp. I-477–I-480.

[24] C. Perra, F. Massidda, and D. D. Giusto, "Image blockiness evaluation based on Sobel operator," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, pp. 389–392.

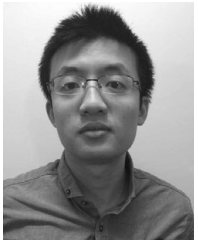
[25] L. Li, W. Lin, and H. Zhu, "Learning structural regularity for evaluating blocking artifacts in JPEG images," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 918–922, Aug. 2014.

[26] L. Li, H. Zhu, G. Yang, and J. Qian, "Referenceless measure of blocking artifacts by Tchebichef kernel analysis," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 122–125, Jan. 2014.

[27] H. Nemoto, P. Hanhart, P. Korshunov, and T. Ebrahimi, "Ultra-eye: UHD and HD images eye tracking dataset," in *Proc. 6th Int. Workshop Quality Multimedia Exper.*, Sep. 2014, pp. 39–40.
[28] *HEVC Reference Software*. [Online]. Available: https://hevc.hhi.fraunhofer.de/
[29] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-13, Jan. 2012.
[30] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1994, pp. 593–600.
[31] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.
[32] A. Rehman, K. Zeng, and Z. Wang, "Display device-adapted video quality-of-experience assessment," *Proc. SPIE*, vol. 9394, p. 939406, Mar. 2015.
[33] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.
[34] P. G. J. Barten, "Formula for the contrast sensitivity of the human eye," *Proc. SPIE*, vol. 5294, pp. 231–238, 2003.
[35] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychol. Bull.*, vol. 124, no. 3, pp. 372–422, Nov. 1998.
[36] K. Rayner, "Eye movements and attention in reading, scene perception, and visual search," *Quart. J. Experim. Psychol.*, vol. 62, no. 8, pp. 1457–1506, 2009.
[37] 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.
[38] 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.
[39] 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.
[40] 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.
[41] 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.
[42] 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.
[43] 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.
[44] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 339–343.
[45] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.
[46] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678–2683, Sep. 2011.
[47] P. V. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 423–426, Jul. 2012.
[48] C. T. Vu, T. D. Phan, and D. M. Chandler, "S_3: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 934–945, Mar. 2012.
[49] R. Hassen, Z. Wang, and M. M. A. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2798–2810, Jul. 2013.
[50] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[51] 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.
[52] D. J. Sheskin, *Handbook Parametric Nonparametric Statistical Procedures*. Boca Raton, FL, USA: CRC Press, 2003.
[53] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, *Image and Video Quality Assessment Research at LIVE*. [Online]. Available: http://live.ece.utexas.edu/research/quality
[54] 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, p. 011006, Jan. 2010.
[55] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015
[56] Y. Horita, K. Shibata, Y. Kawayoke, and Z. P. Sazzad, *MICT Image Quality Evaluation Database*. [Online]. Available: http://mict.eng.utoyama.ac.jp/mictdb.html
[57] VQEG. (2010). *Report on the Validation of Video Quality Models for High Definition Video Content*. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/hdtv/hdtv.aspx
[58] P. V. Vu and D. M. Chandler, "ViS3: An algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *J. Electron. Imag.*, vol. 23, no. 1, p. 013016, Feb. 2014.
[59] F. Zhang, S. Li, L. Ma, Y. C. Wong, and K. N. Ngan. *IVP Subjective Quality Video Database*. [Online]. Available: http://ivp.ee.cuhk.edu.hk/research/database/subjective
[60] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.
[61] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, Mar. 2014.
[62] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, Jan. 2016.



Xiongkuo Min received the B.E. degree from Wuhan University, Wuhan, China, in 2013. He is currently pursuing the Ph.D. degree with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai, China. From 2016 to 2017, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image quality assessment, visual attention modeling, and perceptual signal processing. He received the Best Student Paper Award of ICME 2016.

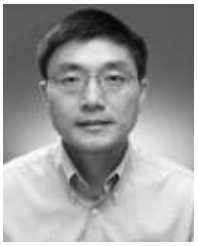


Kede Ma (S'13) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interests include perceptual image processing and computational photography.

Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. His research interests include quality assessment, contrast enhancement, visual saliency detection, and air quality prediction. He received the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and the excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He is the Leading Special Session Organizer in VCIP2016 and ICIP2017. He is an Associated Editor of the IEEE ACCESS. He is the Reviewer for the IEEE T-NNLS, T-IP, TCYB, TIE, T-MM, T-CSVT, T-BC, J-STSP, SPL, ACCESS, the *Information Sciences*, the *Neurocomputing*, SPIC, JVCI, DSP, MTAP, and ELL. He has reviewed over 50 journal papers each year.

Guangtao Zhai (M'10) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich-Alexander University of Erlangen–Nuremberg, Germany. He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. His research interests include multimedia signal processing and perceptual signal processing. He received the Award of the National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012.

Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree from The University of Texas at Austin in 2001. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image and video processing and coding, visual quality assessment and optimization, computational vision and pattern analysis, multimedia communications, and biomedical signal processing. He has over 100 publications in these fields with over 30 000 citations (Google Scholar). He is a fellow of the Canadian Academy of Engineering. He served as a member of the IEEE Multimedia Signal Processing Technical Committee from 2013 to 2015. He was a recipient of the 2016 IEEE Signal Processing Society Sustained Impact Paper Award, the 2015 Primetime Engineering Emmy Award, the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Magazine Best Paper Award, and the 2009 IEEE Signal Processing Society Best Paper Award. He has been a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2015, and an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY since 2016 and the *Pattern Recognition* since 2006. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2009 to 2014 and the IEEE SIGNAL PROCESSING LETTERS from 2006 to 2010, and a Guest Editor of the *IEEE Journal of Selected Topics in Signal Processing* from 2013 to 2014 and from 2007 to 2009.



Weisi Lin (F'16) received the Ph.D. degree from Kings College London. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with over 340 refereed papers published in international journals and conferences. He is a fellow of Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He has been on the Editorial Board of the IEEE TIP, T-MM from 2011 to 2013, SPL, and JVCI. He has been elected as an APSIPA Distinguished Lecturer from 2012 to 2013. He served as a Technical-Program Chair of the Pacific-Rim Conference on Multimedia 2012, the IEEE International Conference on Multimedia and Expo 2013, and the International Workshop on Quality of Multimedia Experience 2014.