ESIM: Edge Similarity for Screen Content Image Quality Assessment

Zhangkai Ni, Lin Ma, *Member, IEEE*, Huanqiang Zeng, *Member, IEEE*, Jing Chen,
Canhui Cai, *Senior Member, IEEE*, and Kai-Kuang Ma, *Fellow, IEEE*

Abstract—In this paper, an accurate full-reference *image quality assessment* (IQA) model developed for assessing *screen content images* (SCIs), called the *edge similarity* (ESIM), is proposed. It is inspired by the fact that the *human visual system* (HVS) is highly sensitive to edges that are often encountered in SCIs; therefore, essential edge features are extracted and exploited for conducting IQA for the SCIs. The key novelty of the proposed ESIM lies in the extraction and use of three salient edge features—i.e., *edge contrast*, *edge width*, and *edge direction*. The first two attributes are simultaneously generated from the input SCI based on a parametric edge model, while the last one is derived directly from the input SCI. The extraction of these three features will be performed for the reference SCI and the distorted SCI, individually. The degree of similarity measured for each above-mentioned edge attribute is then computed independently, followed by combining them together using our proposed *edge-width pooling* strategy to generate the final ESIM score. To conduct the performance evaluation of our proposed ESIM model, a new and the largest SCI database (denoted as SCID) is established in our work and made to the public for download. Our database contains 1800 distorted SCIs that are generated from 40 reference SCIs. For each SCI, nine distortion types are investigated, and five degradation levels are produced for each distortion type. Extensive simulation results have clearly shown that the proposed ESIM model is more consistent with the perception of the HVS on the evaluation of distorted SCIs than the multiple state-of-the-art IQA methods.

Index Terms—Image quality assessment (IQA), screen content images (SCIs), edge modeling, edge direction.

I. INTRODUCTION

IN THE era of multimedia communications, mobile and cloud computing, and the Internet of Things, the contents of digital images are no longer just limited to natural scenes. In fact, the contents of digital images nowadays can have a mixture of sources, such as natural scene, computer-generated graphics, texts, charts, maps, user's hand-writing and -drawing, and even some special symbols or patterns (e.g., logo, bar code, QR code) imposed and rendered by an electronic device or a photo editing software. Such kind of images is denoted as the *screen content images* (SCIs) [1], [2], and they are frequently encountered in various multimedia applications and services, such as online news and advertisement, online education, electronic brochures, remote computing, cloud gaming, to name a few [3]. It has been observed that the SCIs can yield fairly different image characteristics compared to that of natural images [4]. Refer to Fig. 1 for some SCI examples. One can see that the image contents of SCIs tend to have radical changes due to sharp region transitions and texts; all these lead to abundant of edges [5].

One critical issue associated with the SCIs is: how to conduct objective *image quality assessment* (IQA) for this kind of images? The *objectiveness* here means that the measurements yielded from the developed IQA model should be consistent with the judgment made by the *human visual system* (HVS). Since the majority of existing IQA models are developed for natural images, they cannot be exploited to accurately evaluate the SCI perceptual quality. For that, a new IQA model for the evaluation of SCIs is proposed in this paper, called the *edge similarity* (ESIM). Note that, besides being used for perceptual quality assessment, an IQA model for SCIs can be also exploited as an effective *performance index* to guide the development of various SCI-based image processing algorithms (e.g., coding, interpolation, super-resolution, enhancement, and so on). In what follows, related works of existing IQA models will be succinctly discussed.

A comprehensive overview of the current IQA methods can be found in [6]. On the evaluation of natural images, the simplest and commonly-used IQA models are the *peak signal-to-noise ratio* (PSNR) and the *mean square error* (MSE). However, it is well-recognized that these two models often lead to inconsistent measurements, compared with the judgments made by the HVS, since the PSNR and MSE only consider the differences of the pixel intensities [7], [8]. To address this problem, much effort has been devoted to incorporate the properties of the HVS into the design of IQA model.
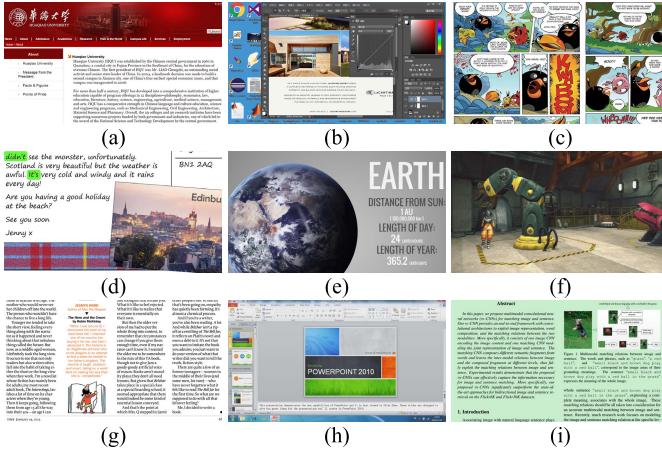
Fig. 1. Nine *screen content images* (SCIs) selected from our newly established SCI database (containing nearly 2,000 images) for demonstrations.

A milestone work in line with this goal is the *structural similarity* (SSIM) [8]. It considers the degradations incurred on image's structure, instead of pixel differences, based on the fact that the HVS is highly sensitive to the image's structural information.

Further consider that the HVS is highly sensitive to image contents that have strong edges and contours, the edge information extracted from the image can be utilized for the development of the IQA models (e.g., [9]–[14]). For example, Zhang *et al.* [12] proposed an IQA model by measuring the degree of similarity on *edge strength* (ES). Inspired by SSIM, Liu *et al.* [13] used the *gradient similarity* to measure the changes in contrast and structure to perform image quality assessment. Xue *et al.* [14] further utilized the *gradient magnitude similarity* (GMS) to capture the image local quality and then exploit the standard deviation computed on the GMS map as the quality index. Besides, various perceptual features, e.g., visual saliency [15], information fidelity [16], [17], biologically inspired features [18], [19] and various computational strategies, such as, multi-scale analysis [20], multiple kernel learning [21], support vector regression [22], extreme learning machine [23], deep learning [24]–[28], and so on, are employed in the design of the IQA models.

Among the existing literatures on the topic of the IQA models, most of them were proposed for evaluating the perceptual quality of the natural images. Obviously, these models are not suitable for conducting the quality evaluation of the SCIs, as the image structures and statistical properties of the SCIs are normally quite different from that of the natural images. For conducting the perceptual quality assessment of the SCIs, Wang *et al.* [29] proposed an IQA model by incorporating visual field adaptation and information content weighting. Yang *et al.* [1] considered the visual difference of the textual and pictorial regions. Gu *et al.* [30] proposed to weight the classical SSIM with a structural degradation measurement. In this paper, a novel and accurate full-reference IQA model for conducting objective evaluations of SCIs is proposed, called the *edge similarity* (ESIM). The key novelty of the proposed ESIM lies in the extraction and use of edge

information, since a typical SCI contains abundant of edge information, and the HVS is highly sensitive to such type of information. For that, three salient edge attributes—i.e., *edge contrast*, *edge width*, and *edge direction*, will be extracted from the reference SCI and the distorted SCI, individually. The first two attributes are simultaneously generated from a parametric edge model [31], while the last one is directly derived from each SCI. The *degree of similarity* of each above-mentioned edge attribute is then measured independently, and the obtained three similarity maps are then combined using our proposed *edge-width pooling* strategy to generate the final ESIM score.

Another significant contribution presented in this paper is our newly established SCI database (denoted as SCID)[1]. Our established SCID is much larger than the only available SCI database, SIQAD [1]. It contains 40 reference SCIs and their generated 1,800 distorted SCIs. Nine different types of distortions are considered with 5 degradation levels generated for each distortion type. This SCID database can be served as the 'ground truth' to quantitatively assess how accurate of the proposed IQA model compared with that of existing state-of-the-art models on the evaluation of SCIs.

Compared with our previous work [32], the proposed ESIM model incorporates an additional edge attribute—*edge direction* for conducting SCI quality assessment. Secondly, the complementary behaviors of the three edge attributes (i.e., the edge width, the edge contrast, and the edge direction) for SCI quality assessment are investigated and demonstrated. Thirdly, a new SCI database is established and extensive simulation are conducted using our established SCI database (i.e., SCID) and SIQAD [1]. Simulation results demonstrate that our proposed model outperforms the existing state-of-the-art IQA models.

The remaining of this paper is organized as follows. In Section II, the proposed SCI quality assessment model using edge similarity is presented in detail. In Section III, a newly established SCI database is presented. In Section IV, extensive performance evaluation of the proposed IQA model and other state-of-the-art IQA models based on public database and our established database are performed and compared. Finally, Section V draws the conclusion.

II. ESIM: PROPOSED EDGE-SIMILARITY-BASED IQA MODEL FOR SCI QUALITY ASSESSMENT

A. Overview

The proposed IQA model, *edge similarity* (ESIM), is used for conducting an objective image quality evaluation of a given *distorted* SCI, with respect to its *reference* SCI. The flowchart of the proposed ESIM algorithm is shown in Fig. 2, which consists of three processing stages. Each stage will be detailed in the following subsections, respectively. Before that, we would like to highlight the main points of each stage as follows.

---

[1] SCID, [Online]. Available: http://smartviplab.org/pubilcations/SCID.html
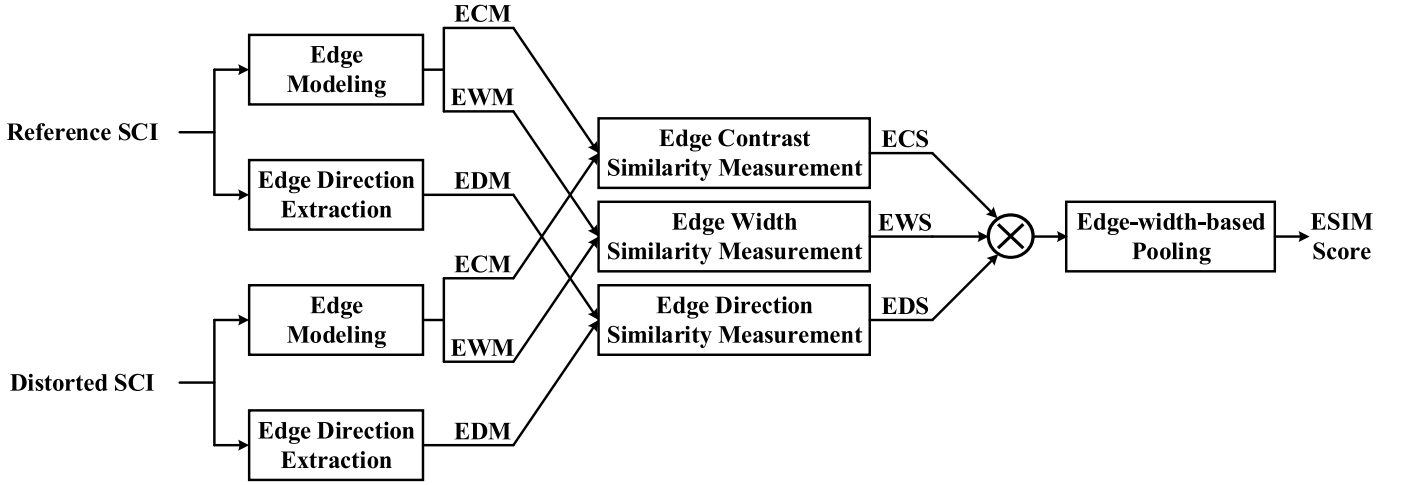
Fig. 2. The framework of the proposed *edge similarity* (ESIM) score computation algorithm. The acronyms, ECM, EWM and EDM, as shown in Stage 1 denote the *edge contrast map*, the *edge width map*, and the *edge direction map*, respectively. In Stage 2, the acronyms ECS, EWS, and EDS are the abbreviations for the *edge contrast similarity*, the *edge width similarity*, and the *edge direction similarity* (EDS), respectively.

In the first stage, a parametric edge model [31] is used to extract two salient edge attributes, *edge contrast* and *edge width*, and this process will be applied to the distorted SCI and the reference SCI, respectively. This modeling process will be conducted at each pixel location, individually and independently. As a result, the extracted edge contrast and edge width information are expressed in terms of *maps*—i.e., the *edge contrast map* (ECM) and the *edge width map* (EWM), respectively. It is important to note that these maps have the same size as that of the input image. In addition, the *edge direction*, which is another salient edge feature, is considered and incorporated into our proposed IQA model. The *edge direction map* (EDM) will be generated directly from each SCI via our proposed edge direction computation method.

In the second stage, the computed edge feature maps, one from the reference SCI and the other from the distorted SCI, will be compared to yield their edge similarity measurement. For example, the two ECMs, respectively obtained from the distorted SCI and the reference SCI, will be compared to arrive at the *edge contrast similarity* (ECS) map. Likewise, the *edge width similarity* (EWS) map and *edge direction similarity* (EDS) map will be generated based on the corresponding pair of EWMs and EDMs, respectively. The three generated similarity measurement maps will be combined to yield *one* measurement map, which is used as the input of the third, and the last, stage to compute the final ESIM score using our proposed edge-width-based pooling process.

B. Stage 1: Computation of Three Edge-Feature Maps

1) Edge Contrast Map (ECM) and Edge Width Map (EWM): Since the HVS is highly sensitive to edges, a parametric edge model [31] is employed to model each input SCI for extracting its edge information; to be specific, edge *contrast* and edge *width*.

To model an *ideal* step edge, centered at the location $x = x_0$, this can be mathematically expressed as

$$u(x; b, c, x_0) = c \cdot U(x - x_0) + b, \qquad (1)$$
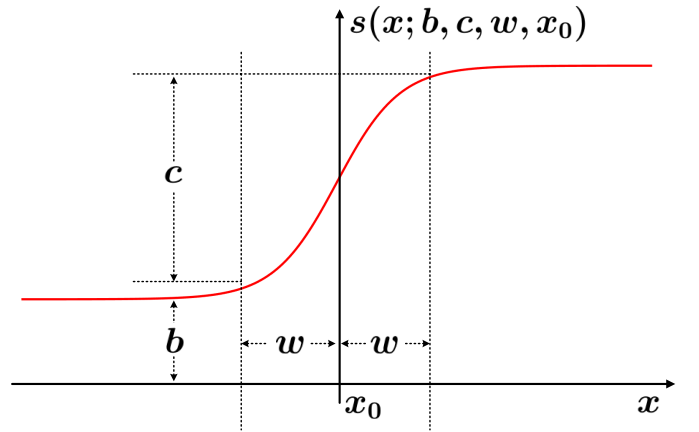
Fig. 3. An illustration of a *practical* edge model $s(x; b, c, w, x_0)$ with an edge incurred at the location $x = x_0$, where c is the contrast parameter, w is the width parameter, and b is the basis parameter [31].

where $U(\cdot)$ denotes the unit-step function, c represents the edge contrast, and b is the basis (i.e., luminance intensity at the lower side of the edge contrast). Generally speaking, a typical edge has a smooth transition, rather than a unit-step like sharp edge. Therefore, a *practical* edge model, $s(x; b, c, w, x_0)$, as illustrated in Fig. 3 can be obtained by convolving the *ideal* edge model (i.e., the unit-step edge as expressed in (1)) with the Gaussian kernel $g(x; w)$, where w is the standard deviation that is used to model the edge *width*; that is,

$$
\begin{aligned}
s(x; b, c, w, x_0) &= u(x; b, c, x_0) \otimes g(x; w) \\
&= u(x; b, c, x_0) \otimes \frac{1}{\sqrt{2\pi w^2}} \exp\left(\frac{-x^2}{2w^2}\right) \\
&= b + \frac{c}{2}\left[1 + \mathrm{erf}\left(\frac{x - x_0}{\sqrt{2}w}\right)\right], \qquad (2)
\end{aligned}
$$

where the symbol "$\otimes$" denotes the convolution operation, $\mathrm{erf}(\cdot)$ is the error function. Note that any edge can be modelled by this parametric edge model as shown in Fig. 3 via a

three-parameter set—i.e., the contrast c, the width w, and the basis b. The parameter c determines the edge contrast (i.e., the strength of the edge); the larger the c, the stronger the edge contrast. The parameter w governs the edge width; the smaller the value, the sharper the edge transition. (That is, the ideal unit-step edge has a *zero* edge width.) Lastly, the parameter b conveys whether the above-mentioned edge occurs in a brighter region (i.e., a larger b) or in a darker region (i.e., a smaller b). Now, the key issue is how to determine these three parameters, b, c, and w, for each edge model centered at a local position, $x = x_0$.

Intuitively, these parameters can be derived by using the local pixel information near the edge center $x = x_0$. For that, the edge center should be detected and localized firstly. It can be realized by simply taking the local maxima of the first derivative of the Gaussian smoothed step edge $s(x; b, c, w, x_0)$. To prevent such derivation being conducted on those pixel locations involving noise, it is effective to convolve the $s(x; b, c, w, x_0)$ with a Gaussian kernel first before taking the derivatives. By exploiting the well-known fact that $(f_1(x) \otimes f_2(x))' = f_1(x) \otimes f_2'(x)$, where the symbol "$\otimes$" denotes the convolution operation of two given functions $f_1(x)$ and $f_2(x)$, the above-mentioned two operations (i.e., convolution and then derivation) can be realized by performing the convolution of the $s(x; b, c, w, x_0)$ with the *derivative* of the Gaussian filter $g'(x; \sigma_d)$ and arrives at [31]

$$d(x; c, w, \sigma_d, x_0) = \frac{c}{\sqrt{2\pi(w^2 + \sigma_d^2)}} \exp\left[\frac{-(x - x_0)^2}{2(w^2 + \sigma_d^2)}\right]. \quad (3)$$

Since we have three parameters required to be solved, thus we need to have three equations to solve them. For that, we shall sample $d(x; c, w, \sigma_d, x_0)$ at three locations $x = 0, a$, and $-a$, where a is the sampling distance, and it can be chosen freely (for that, $a = 1$ is used in this paper). Based on the above-mentioned three locations, denote $d_1 = d(0; c, w, \delta_d, x_0)$, $d_2 = d(a; c, w, \delta_d, x_0)$, and $d_3 = d(-a; c, w, \delta_d, x_0)$, for $x = 0$, $x = a$, and $x = -a$, respectively. The three parameters of the edge model can be derived as [31]:

$$c = d_1 \cdot \sqrt{2\pi a^2 / ln(l_1)} \cdot l_2^{\frac{1}{4a}}, \quad (4)$$

$$w = \sqrt{a^2 / ln(l_1) - \sigma_d^2}, \quad (5)$$

$$b = s(x_0) - \frac{c}{2}, \quad (6)$$

where $l_1 = d_1^2 / (d_2 d_3)$ and $l_2 = d_2 / d_3$.

Resulted from the above-described edge modeling process, which is performed at each pixel location, all the computed values of parameters c and w can be separately grouped and formed as a 'map' for each parameter, respectively. Thus, the established *edge contrast map* (ECM) and *edge width map* (EWM) are two extracted salient edge information, and this process will be applied to the distorted SCI and the reference SCI individually. These edge-feature maps will be served as the inputs to the subsequent Stage 2 for conducting edge similarity measurement (refer to Fig. 2).

To demonstrate the effectiveness of the edge model [31] on the extraction of salient edge attributes from the SCIs, a test SCI was selected from the SIQAD database [1] as shown in Fig. 4 (a). This image is served as the *reference* SCI for the quality assessment against each *distorted* SCI as shown in Fig. 4 (e) and (i); all are locating in the first column in Fig. 4. Note that Fig. 4 (e) contains motion blur, while Fig. 4 (i) is a JPEG compressed version of Fig. 4 (a). The computed ECMs and EWMs based on these three SCIs are presented in columns 2 and 3 of Fig. 4, respectively. Comparing Fig. 4 (b) and (f) for example, it can be easily observed that a significant amount of edge information and texture information got lost due to motion blur, as expected. This shows that the edge modeling process reflects the edge information quite well.

2) Edge Direction Map (EDM): It has been noticed that *edge direction* is another salient edge attribute that can be beneficial to the SCI quality assessment. This is motivated by the fact that the visual cortical neurons are highly sensitive to edge direction [33]–[35]. Similar to ECM and EWM, the *edge direction map* (EDM) needs to be computed, which is described as follows.

Let $I(x, y)$ denote the luminance component of the input SCI under consideration. Define G_H and G_V are the directional derivatives computed along the horizontal and vertical directions, respectively, as follows:

$$G_H = I(x + 1, y) - I(x, y)$$
$$G_V = I(x, y + 1) - I(x, y). \quad (7)$$

The local gradient magnitude measured at the location (x, y), which is denoted as $G(x, y)$, is defined as

$$G(x, y) = |G_H| + |G_V|. \quad (8)$$

Note that the l_1 norm is used here, rather than l_2 norm, since the total amount of gradient changes (regardless the involved directions) is an effective impetus to the HVS and $|G_H| + |G_V| \geq \sqrt{G_H^2 + G_V^2}$. It turns out that this leads to better performance according to our extensive simulation results. Based on (8), the obtained edge gradient map will be convolved with the kernel $\mathscr{L}_l$ at the lth direction, for $l \in \{0, \ldots, 11\}$; that is,

$$EDM_l(x, y) = \mathscr{L}_l \otimes G(x, y), \quad (9)$$

where the symbol "$\otimes$" denotes the convolution operator. Note that the convolution kernel $\mathscr{L}_l$ is obtained by rotating the convolution kernel $\mathscr{L}_0$ with an angle of $l \times \frac{\pi}{12}$. Here, $\mathscr{L}_0$ is a matrix with a size of 27×27, and it is exploited as a convolution kernel operating along the horizontal direction. In this 27×27 matrix, all the entries in the middle row of $\mathscr{L}_0$ have a constant value 1, and the value of 0 for the remaining entries.

Lastly, the final $EDM(x, y)$ will be generated based on the twelve $EDM_l(x, y)$ by simply identifying which map (associating with a fixed direction) has yielded the largest value (or response) at each pixel location (x, y); that is,
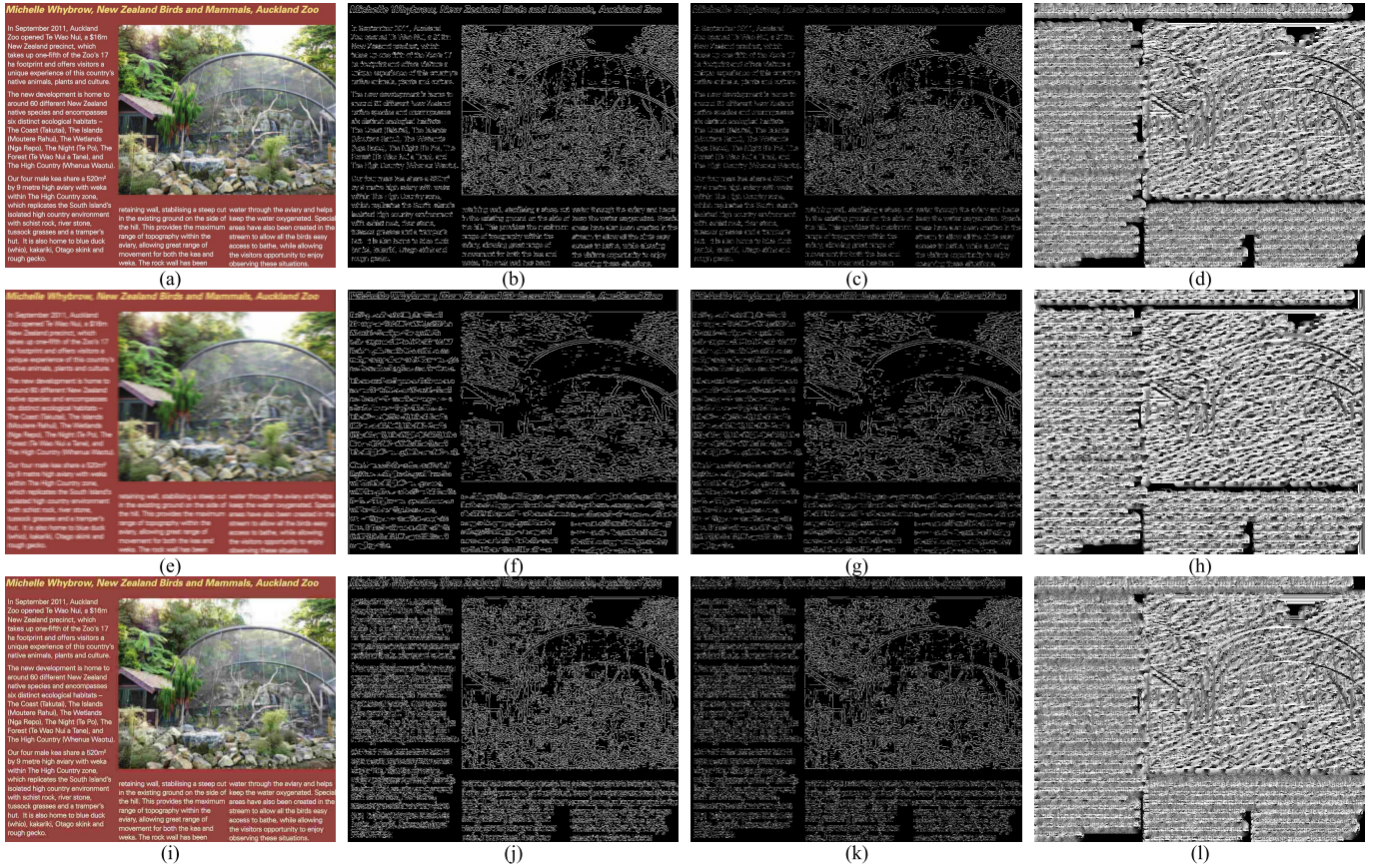
$$EDM(x, y) = n \times \frac{\pi}{12}, \quad (10)$$

Fig. 4. An example of *edge contrast map* (ECM), *edge width map* (EWM) and *edge direction map* (EDM) of an SCI and its two distorted versions. [Column 1]: (a) an original or reference SCI; (e) a distorted SCI resulted from motion blur; (i) a distorted SCI resulted from JPEG compression. [Columns 2 to 4]: their corresponding ECM, EWM, and EDM, respectively.

where

$$n = \arg\max_l \{EDM_l(x, y)\}, l \in \{0, \ldots, 11\}.$$

It is important to remark that the commonly-practiced approach for finding the angle is based on arctan $(\frac{G_V}{G_H})$, which is quite simple and straightforward. However, compared with arctan $(\frac{G_V}{G_H})$, our proposed approach adopts a larger neighborhood of the current pixel, and thus can be more accurate and stable to depict the edge direction. Therefore, our proposed approach is employed to explore the important edge attribute (i.e., edge direction) for SCI perceptual quality assessment. For a demonstration of EDM, refer to the last column in Fig. 4.

C. Stage 2: Computation of Three Edge Similarity Measurements

This subsection is about the second stage in Fig. 2. For ease of notation and paper development, denote the reference SCI as r and its distorted version as d; both are added on the subscript. The computed similarity measurements for the concerned three edge attributes are called the *edge contrast similarity* (ECS), the *edge width similarity* (EWS), and the *edge direction similarity* (EDS), respectively. These measurements are computed by following the same practice as proposed in [8] on the computation of the degree of similarity,

which is also commonly adopted in other IQA models (e.g., [10], [13], [14], [36]). In summary, the above-mentioned three similarity measurements can be computed according to the following format:

$$s_\psi(\psi_r, \psi_d, T_\psi) = \frac{2\psi_r \psi_d + T_\psi}{\psi_r^2 + \psi_d^2 + T_\psi}, \qquad (11)$$

where the symbol ψ can be ECM, EWM, or EDM, and the subscript of ψ_r and ψ_d indicates whether the involved SCI is the reference (i.e., r) or the distorted (i.e., d) SCI. T_ψ is a positive constant that is used to prevent from numerical fluctuation. Finally, the obtained $s_\psi(\cdot)$ is the final similarity measurement map, and it could be the ECS, EWS, or EDS. Note that each computed value of $s_\psi(\psi_r, \psi_d, T_\psi)$ should be falling in the range of (0, 1].

By following the same practice as suggested in [8], the *total edge similarity measurement*, denoted as $S(x, y)$, can be computed by simply multiplying these three maps together as (refer to Fig. 2)

$$S(x, y) = [ECS(x, y)]^\alpha \cdot [EWS(x, y)]^\beta \cdot [EDS(x, y)]^\gamma, \qquad (12)$$

where α, β, and γ are three positive constants that can be used to adjust the relative importance of $ECS(x, y)$, $EWS(x, y)$

and $EDS(x, y)$. By simply treating these three measurements equally important, $\alpha = \beta = \gamma = 1$ is set in our work.

D. Stage 3: Edge-Width-Based Pooling

This subsection is about the third, and the last, processing stage as outlined in Fig. 2 that will yield the final *edge similarity* (ESIM) score. Instead of simply taking an average over the combined edge attributes map $S(x, y)$, intuitively a proper weighting process should be considered, as all image pixels are not equally perceived by the HVS. For that, we have investigated several different edge attributes for weighting. It turns out that the edge *width* information is the most effective data field to be incorporated for weighting. This is probably due to the fact that the HVS perception is very sensitive to edge structure, which is intimately related to edge width [2].

To fulfill the above-mentioned goal, a simple and yet effective weighting factors can be designed based on the existing data field, EWM, as follows. If the pixel location (x, y) of either the original or distorted SCI has incurred a large edge width w (be it in the reference SCI or the distorted SCI), it means that HVS will be quite sensitive to this position. Therefore, denote EWM_r and EWM_d be the edge width maps of SCIs r and d, respectively. The maximum edge width value based on these two maps can be utilized as the weighting factor; that is,

$$W(x, y) = max(EWM_r, EWM_d). \tag{13}$$

Therefore, the IQA score, which is termed as the *edge similarity* (ESIM) score, can be computed 'pooled' through the weighted average as

$$ESIM = \frac{\sum\limits_{(x,y)} W(x, y) \cdot S(x, y)}{\sum\limits_{(x,y)} W(x, y)}, \tag{14}$$

where the weighted average is carried out in all pixel locations (x, y).

III. Proposed Screen Content Image Database (SCID)

A. Methodology on Building Our SCI Database

Our developed SCI database (denoted as *SCID*) contains 40 *reference* SCIs and 1,800 *distorted* versions rendered from these reference SCIs. The 40 reference SCIs were thoughtfully identified from the Internet, and they cover a wide variety of image contents, including texts, graphics, symbols, patterns, and natural images. The sources of these SCIs could come from the web pages, PDF files, power-point slides, comics, digital magazines (via screen snapshots), and so on. All the selected reference SCIs are cropped to a fixed size in 1280×720. For demonstration, few reference SCIs from our database are shown in Fig. 1.

Since various types of distortions could be inevitably introduced on SCIs during the acquisition, processing, compression, transmission, and display stages, our SCID database includes 9 types of distortions that are often encountered

in practical applications. For each distortion type, 5 levels of degradations (ranging from imperceptible level to highly-annoying one) are generated and included in our database. As a result, we have produced 1,800 distorted SCIs for our database. These 9 types of distortions include: noisy SCIs with three types of noise considered—i.e., the Gaussian noise (GN), the Gaussian blur (GB), and the motion blur (MB), as they are often encountered on the transmission stages.

Since the HVS is sensitive to image contrast, defined as the pixel-intensity difference yielded between the darkest and the brightest areas within the image, the contrast change (CC) is introduced as another type of distortion considered in our SCID database. Furthermore, color saturation change (CSC) and color quantization with dithering (CQD) are also important to include. The CC, CSC, and CQD are quite likely resulted from color rendering and screen sharing among different display devices with different brightness and contrast.

Besides, distortion caused by compression is also often encountered on image. For that, we consider the JPEG and JPEG2000 (J2K) image compression standards, as they are often used to encode still color images. Moreover, the latest video compression standard, *high efficiency video coding* (HEVC) [37], [38], is also considered, since the HEVC has recently added the *screen content coding* (SCC) as its extension, denoted as HEVC-SCC. Thus, the distorted still image frames resulted by applying the HEVC-SCC are also included in our database.

B. Methodology on Conducting Subjective Test

As specified in ITU-R BT.500-13 [39], subjective test procedures aim at assessing the quality of television pictures via two approaches: *single stimulus* and *double stimulus*. The single stimulus approach only presents a single image to the assessors and seeks their opinion scores. Obviously, the single stimulus approach is straightforward and simple on performing subjective test [40]. On the other hand, the double stimulus approach seeks the assessor's subjective evaluation on the assessment of the quality difference yielded between the reference image and the distorted image. Compare to the single stimulus method, the *opinion scores* resulted from double stimulus method are more reliable [39], [41]. For that, the *double stimulus impairment scale* (DSIS) method is employed in our work using five discrete scales (from 1 to 5) for each specific type of distortion under consideration. Note that both the reference SCI and the distorted SCI are presented to the assessors for 10 seconds each with a mid-grey presentation inserted in between. Note that the higher the score, the better the perceptual quality is deemed. The desktop PCs were used as the evaluation test beds. Each is equipped with a 23-inch LED monitor (with a resolution of 1920×1080), 8 GB RAM, and 64-bit Windows operating system. The evaluation process is conducted indoors, under a normal lighting condition.

It is essential and informative to describe how these evaluation tests were carried out in order to judge how objective and useful of our established database and the obtained evaluation results. For that, several training sessions were held to instruct
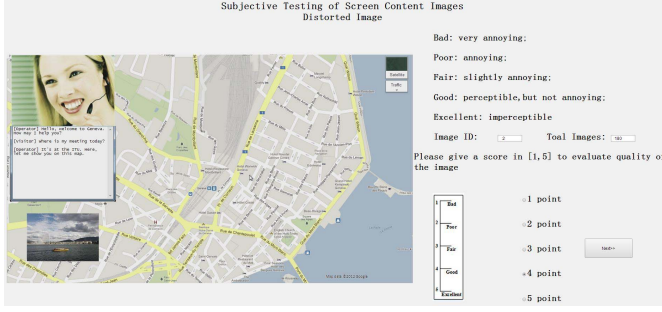
Fig. 5. A screenshot of a graphical user interface for conducting subjective evaluation.

all the assessors about the rules and procedures on how to conduct the subjective evaluation of the distorted SCIs. Each distorted SCI presented to the assessor will be judged as one of five opinion levels (together with their corresponding definitions) as shown in Fig. 5. All these information are located at the bottom right of the window. There are 186 assessors, including 104 males and 82 females. All assessors are students (with ages ranging from 18 to 27), and they do not have any background or experience in image processing so as to avoid bias as much as possible. According to [39], the time spent on one evaluation session should not exceed 30 minutes to avoid fatigue that might affect scores' accuracy. Therefore, the entire 1,800 distorted SCIs were randomly and non-overlappingly divided into 10 sessions (i.e., 180 images per session) for conducting subjective test. In each session, no two consecutive SCIs to be evaluated in sequence are the same, nor their distortion types and levels of distortion. Each assessor was randomly assigned with two sessions of SCIs on average (with three sessions at most) for conducting their subjective evaluation, and each SCI is evaluated by at least 40 assessors. Moreover, the assessors had been instructed to take a break after 30 minutes of evaluation work to avoid fatigue that might lead to unreliable scores. Finally, the subjective ratings of 430 evaluation sessions are obtained in total.

C. MOS Computation and Its Reliability Analysis

As above-mentioned, each distorted SCI in our built SCID database was subjectively evaluated to obtain the corresponding opinion (raw) scores. These scores are required to be further processed in order to generate the final *mean opinion score* (MOS). The obtained MOS for each distorted SCI can thus be used as the 'ground truth' and stored in our database for evaluating the performance of the proposed IQA model as well as other classical and state-of-the-art models for comparison. To make sure that the obtained ground truth is as reliable as possible, the rules outlined in [39] have been strictly followed to determine the MOS for each distorted SCI in our subjective evaluation process.

Although each assessor might have a different interpretation about the image quality of the SCIs under evaluation, the majority of assessors should have reached the same or similar conclusions on the perceptual quality judgment on each distorted SCI. Therefore, any outlier existing in the raw subjective scores needs to be identified

and discarded according to the checking procedures described in [39] before computing the final MOS for each distorted SCI. Let S_{ijk} be the opinion (raw) score recorded by the assessor i on evaluating a distorted SCI j (where $j = \{1, \ldots, 180\}$) from the session k (where $k = \{1, \ldots, 10\}$). To identify outliers, an intuitive approach, which is also suggested in [39], is to compare all the assessors' raw scores that had been recorded for the evaluation of the same session k. For each SCI in this session, the majority of assessors should have the same or close scores so as to differentiate which are the outliers to be discarded. Further note that each session k could have different numbers of assessors involved; for that, let N_k be the number of assessors who have evaluated the session-k's SCIs. First of all, it is essential to 'normalize' the recorded raw scores to yield the so-called Z *scores* [39] for each SCI from the session k and across these N_k's assessors [42]; i.e.,

$$Z_{ijk} = \frac{S_{ijk} - \mu_{jk}}{\sigma_{jk}}, \qquad (15)$$

where

$$\mu_{jk} = \frac{1}{N_k} \sum_{i=1}^{N_k} S_{ijk},$$

$$\sigma_{jk} = \sqrt{\frac{1}{N_k - 1} \sum_{i=1}^{N_k} (S_{ijk} - \mu_{jk})^2}.$$

After obtaining the Z scores for all N_k's assessors, the *kurtosis* (denoted as β_{ik}) of the raw-scores of assessor i for the k-th session will be computed as the yardstick to conduct the rejection procedures as specified in [39] so as to identify whether this entire evaluation session (with 180 scores) is deemed as an un-trustable 'outlier' and to be discarded. The kurtosis β_{ik} is measured according to

$$\beta_{ik} = \frac{m_{ik,4}}{(m_{ik,2})^2}, \qquad (16)$$

where

$$m_{ik,t} = \frac{\sum_{j=1}^{J_k} (S_{ijk} - \gamma_{ik})^t}{J_k}, \quad \text{for } t = 2, 4;$$

$$\gamma_{ik} = \frac{1}{J_k} \sum_{j=1}^{J_k} S_{ijk},$$

where J_k is the total number of distorted SCIs in the k-th session, i.e., $J_k = 180$ presented in this work.

When the computed kurtosis value β_{ik} falls between 2 and 4, the distribution of the raw scores are considered to be normally distributed [39]. Furthermore, each SCI's raw score in this evaluation session needs to be checked against the mean value and standard variation over N_k's scores, followed by some thresholding-based checking steps. The details of these steps are summarized in the Algorithm 1 [39] as shown below. By performing the above-mentioned outlier rejection procedures for each session (from session index 1 to 10), 54 out of 430 evaluation sessions were discarded.

After outlier rejections, the Z scores of each remaining session should be normally distributed with zero mean and unit standard deviation ($\sigma = 1$) [43]. This means that 99% of

Algorithm 1 Outlier Rejection Procedures

---
if $2 \leq \beta_{ik} \leq 4$ (normally distributed) **then**
    if $S_{ijk} \geq \mu_{jk} + 2\sigma_{jk}$ **then**
        $P_{ik} = P_{ik} + 1;$
    end if
    if $S_{ijk} \leq \mu_{jk} - 2\sigma_{jk}$ **then**
        $Q_{ik} = Q_{ik} + 1;$
    end if
else
    if $S_{ijk} \geq \mu_{jk} + \sqrt{20}\sigma_{jk}$ **then**
        $P_{ik} = P_{ik} + 1;$
    end if
    if $S_{ijk} \leq \mu_{jk} - \sqrt{20}\sigma_{jk}$ **then**
        $Q_{ik} = Q_{ik} + 1;$
    end if
end if
if $(P_{ik} + Q_{ik})/J_k > 0.05$ and $|(P_{ik} - Q_{ik})/(P_{ik} + Q_{ik})| < 0.3$ **then**
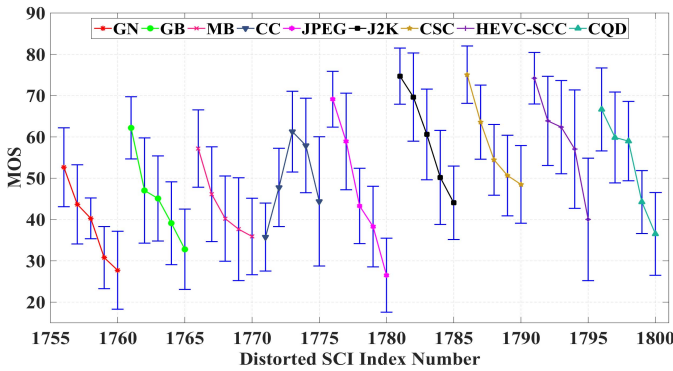   Reject the assessor i.
end if

Fig. 6. The MOS values and their corresponding confidence intervals (i.e., standard deviation as indicated by the blue-color vertical bars). These scores were individually measured from 45 *distorted* SCIs (i.e., 9 distortion types, with 5 level of degradations each), and these images were generated from the *reference* SCI as shown in Fig. 1 (h).

the scores will lie in the range of $\pm 3\sigma$; i.e., $[-3, +3]$. To avoid negative value in our context, all the values are shifted upward by adding 3, followed by a linear rescaling over the range of [0, 100]. Therefore, the *re-scaled Z* scores can be obtained as

$$\widetilde{Z}_{ijk} = \frac{100(Z_{ijk} + 3)}{6}. \tag{17}$$

Finally, for each distorted SCI j from the SCI session k, its MOS value (denoted as MOS_{jk}) is computed as the mean of the re-scaled Z scores (i.e., $\widetilde{Z}_{ijk}$) from (17):

$$\text{MOS}_{jk} = \frac{1}{M_k} \sum_{i=1}^{M_k} \widetilde{Z}_{ijk}, \tag{18}$$

where M_k is the number of remaining evaluation sessions after discarding the identified outliers (i.e., $M_k \leq N_k$). As a result, the standard deviation of $\widetilde{Z}_{ijk}$ can be computed as

$$\widetilde{\sigma}_{jk} = \sqrt{\frac{1}{M_k - 1} \sum_{i=1}^{M_k} (\widetilde{Z}_{ijk} - MOS_{jk})^2}. \tag{19}$$

The MOS value of each distorted SCI j is included in our database, which is then served as the 'ground truth' to evaluate the performance of the IQA models.

To demonstrate, 45 distorted SCIs were generated from Fig. 1 (h), which is chosen as the reference SCI. These 45 images cover all 9 distortion types and with 5 levels of degradations for each distortion type considered. The corresponding obtained MOS are shown in Fig. 6, where the error bar indicates the standard deviation computed from (19). One can see that the distorted SCIs with different distortions types and levels present different MOS values and similar standard deviation values. It means that the assessors have reached an agreement on the SCI perceptual quality. Note that similar observations can also be found for other distorted SCIs.

Moreover, Fig. 7 (a) illustrates the MOS values of all 1,800 distorted SCIs stored in our SCID database. For better visualization, in this plot, 1,800 symbols (i.e., performance points) represents 1,800 distorted SCI and 9 types of symbols correspond to 9 types of distortion (i.e., GN, GB, MB, CC, JPEG, J2K, CSC, HEVC-SCC, and CQD) in our SCID database. Fig. 7 (b) summarized the results in terms of histogram as a complementary presentation. One can see that the distribution of MOS values is, generally speaking, quite even at different quality degradation levels. This indicates that the constructed SCID obeys the rules that the perceptual quality in a database should span the entire range of visual quality from severally annoying to imperceptible with a good separation [43]. In summary, it is believed that the computed MOS based on those assessors' scores (i.e., after outlier rejection) are reliable and can be further employed for the evaluation of the IQA models.

IV. EXPERIMENTAL RESULT AND ANALYSIS

A. Database and Evaluation Criteria

In this section, the performances resulted from the proposed ESIM model and other state-of-the-art models are compared based on the SIQAD database [1] and our established database SCID as described in Section III. Note that the SIQAD database was also established for evaluating the perceptual quality of the SCIs, and it contains 20 reference SCIs and their 980 distorted SCIs, involving 7 distortion types as listed in the second column in Table I—i.e., Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), JPEG compression, JPEG2000 (J2K) compression, and layer-segmentation-based coding (LSC). Furthermore, 7 different levels of degradations have been created for each distortion type. Our proposed SCID database has 9 distortion types as listed in the second column in Table II, and 5 levels of degradations have been created for each distortion type.

It is important to note that the dynamic range of the IQA scores (denoted as s_i) yielded by using a specific model would be different from that of exploiting another model. Therefore, it is necessary to map the dynamic range of the scores produced from each model onto a common scale so that the mapped scores (denoted as Q_i) can be meaningfully compared, with respect to the ground-truth scores that have been stored in the database. For that, the mapping process
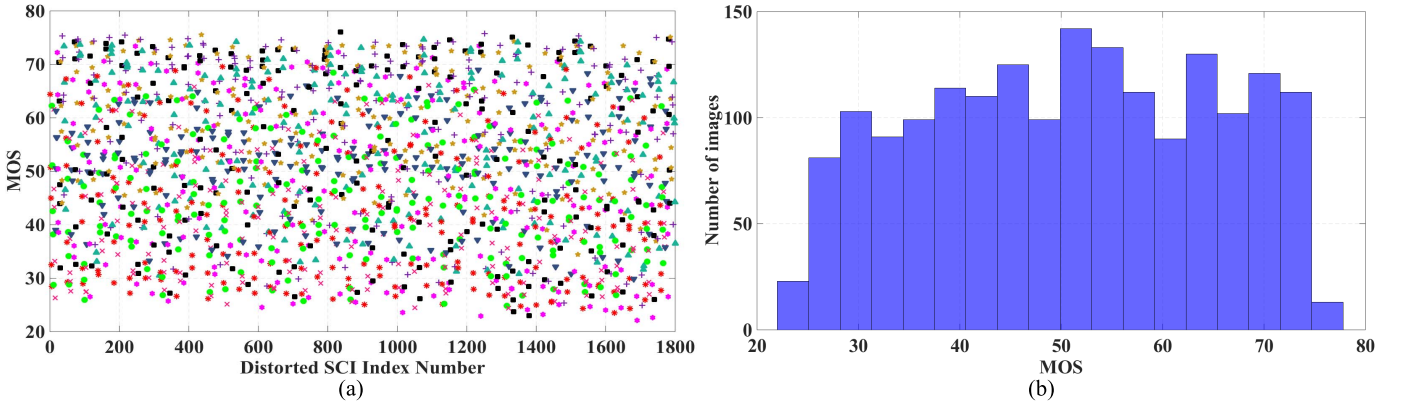
Fig. 7. The profiles of the obtained MOS (i.e., the 'ground truth') in our established SCID: (a) Scatter plot of the entire 1,800 distorted SCIs and their MOS values; (b) Histogram of the MOS values of the entire 1,800 distorted SCIs.

TABLE I

PERFORMANCE COMPARISONS OF DIFFERENT IQA MODELS EXPERIMENTED ON THE *SIQAD* DATABASE

Criteria	Distortions	PSNR	SSIM [8]	MSSIM [20]	IWSSIM [46]	VIF [16]	IFC [17]	MAD [47]	FSIM [36]	ES [12]	GSIM [13]	GMSD [14]	VSI [15]	SCQI [48]	SIQM [30]	SPQA [1]	SQI [29]	EMSQA [32]	ESIM
PLCC	GN	**0.9053**	0.8806	0.8783	0.8804	**0.9011**	0.8791	0.8852	0.7428	0.8167	0.8448	**0.8956**	0.8762	0.8807	0.892	0.8921	0.8829	0.8889	0.8891
	GB	0.8603	0.9014	0.8984	0.9079	0.9102	0.9061	0.9120	0.7206	0.8653	0.8831	0.9094	0.8502	0.8535	0.912	0.9058	**0.9202**	0.9156	**0.9234**
	MB	0.7044	0.8060	0.8240	0.8414	0.8490	0.6782	0.8361	0.6874	0.7873	0.7711	0.8436	0.6620	0.6949	0.845	0.8315	**0.8789**	0.8753	**0.8886**
	CC	0.7401	0.7435	**0.8371**	**0.8404**	0.7076	0.6870	0.3933	0.7507	0.8005	**0.8077**	0.7827	0.7723	0.7119	0.790	0.7992	0.7724	0.7688	0.7641
	JPEG	0.7545	0.7487	0.7756	**0.7998**	0.7986	0.7606	0.7662	0.5566	0.6065	0.6778	0.7746	0.7124	0.6782	0.771	0.7696	**0.8218**	0.7904	**0.7999**
	J2K	0.7893	0.7749	0.7951	0.8040	0.8205	0.7963	**0.8344**	0.6675	0.6717	0.7242	**0.8509**	0.7479	0.7225	0.794	0.8252	**0.8271**	0.7850	0.7888
	LSC	0.7805	0.7307	0.7729	0.8155	**0.8385**	0.7679	0.8184	0.5964	0.6569	0.7218	**0.8559**	0.7454	0.7418	0.720	0.7958	**0.8310**	0.7747	0.7915
	Overall	0.5869	0.7561	0.6161	0.6527	0.8189	0.6395	0.6191	0.5389	0.5273	0.5663	0.7387	0.5543	0.6029	0.852	0.8584	**0.8644**	0.8648	**0.8788**
SROCC	GN	0.8790	0.8694	0.8679	0.8743	**0.8888**	0.8717	0.8721	0.7373	0.8295	0.8404	**0.8856**	0.8655	0.8821	0.871	**0.8823**	0.8602	0.8745	0.8757
	GB	0.8573	0.8921	0.8883	0.9060	0.9059	0.9106	0.9087	0.7286	0.8922	0.8796	0.9119	0.8495	0.8463	0.910	0.9017	**0.9244**	0.9154	**0.9239**
	MB	0.7130	0.8041	0.8238	0.8421	0.8492	0.6737	0.8357	0.6641	0.8083	0.7753	0.8441	0.7658	0.7604	0.840	0.8255	**0.8810**	0.8788	**0.8938**
	CC	0.6828	0.6405	**0.7506**	**0.7563**	0.6433	0.6396	0.3907	**0.7175**	0.6833	0.7148	0.6378	0.6495	0.5780	0.705	0.6154	0.6677	0.6306	0.6108
	JPEG	0.7569	0.7576	0.7787	**0.7978**	0.7924	0.7636	0.7674	0.5879	0.6543	0.6796	0.7712	0.7196	0.7080	0.775	0.7673	**0.8189**	0.7871	**0.7989**
	J2K	0.7746	0.7603	0.7855	0.7998	0.8131	0.7980	**0.8382**	0.6363	0.6824	0.7125	**0.8436**	0.7299	0.7231	0.777	0.8152	**0.8169**	0.7762	0.7827
	LSC	0.7930	0.7371	0.7711	0.8214	**0.8463**	0.7713	0.8154	0.5979	0.6899	0.7145	**0.8592**	0.7419	0.7588	0.725	0.8003	**0.8432**	0.7798	0.7958
	Overall	0.5604	0.7566	0.6115	0.6545	0.8065	0.6011	0.6067	0.5279	0.5637	0.5551	0.7305	0.5381	0.6113	0.845	0.8416	**0.8548**	0.8504	**0.8632**
RMSE	GN	**6.3372**	7.0679	7.1309	7.0744	**6.4673**	7.1096	6.9391	9.9860	8.6084	7.9811	**6.6354**	7.1890	7.0651	7.0165	6.7394	-	6.832	6.8272
	GB	7.7376	6.5701	6.6638	6.3619	6.2859	6.4193	6.2269	10.5230	7.6073	7.1210	6.3111	7.9900	7.9092	**5.8367**	6.4301	-	**6.071**	**5.8270**
	MB	9.2287	7.6967	7.3675	7.0600	6.8704	9.5544	7.1322	9.4432	8.0165	8.2788	6.9816	9.7450	9.3502	**6.0869**	7.2223	-	**6.288**	**5.9639**
	CC	8.4591	8.4116	**6.8818**	**6.8184**	8.8876	9.1407	11.5652	8.3190	7.5395	**7.4160**	7.8294	7.9900	8.8342	8.1079	7.6184	-	8.044	8.1141
	JPEG	6.1665	6.2295	5.9311	**5.6404**	5.6551	6.1004	6.0380	7.8072	7.4712	6.9085	5.9427	6.5950	6.9057	**5.6548**	6.0000	-	5.756	**5.6401**
	J2K	6.3819	6.5691	6.3040	6.1804	5.9412	6.2875	**5.7276**	7.7404	7.6998	7.1675	**4.5491**	6.8990	7.1859	6.0820	**5.8706**	-	6.439	6.3877
	LSC	5.3336	5.8253	5.4141	4.9379	**4.6497**	5.4657	**4.9025**	6.8486	6.4327	5.9046	**4.4121**	5.6880	5.7226	5.3576	5.1664	-	5.395	5.2150
	Overall	11.5898	9.3676	11.2744	10.8444	8.1969	11.0048	11.2409	12.0583	12.1634	11.798	9.6484	11.915	11.4206	7.4936	7.3421	**7.1982**	**7.186**	**6.8310**

suggested in the *Video Quality Experts Group* (VQEG) HDTV test [44], [45] is exploited. That is,

$$Q_i = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + \exp\left[\beta_2(s_i - \beta_3)\right]} \right] + \beta_4 s_i + \beta_5, \quad (20)$$

where s_i is the perceptual quality score of the i-th distorted image computed from an IQA model and Q_i is the corresponding mapped score. Parameters β_1, β_2, β_3, β_4, and β_5 are to be fitted by minimizing the sum of squared errors yielded between the mapped objective score Q_i and the ground truth value MOS_i (for details, see [44] and [45]).

After mapping, three performance evaluation criteria as suggested in [44] are computed for various IQA models, respectively. They are the *Pearson linear correlation coefficient* (PLCC) for predicting *accuracy*, the *Spearman rank order correlations coefficient* (SROCC) for predicting *monotonicity*, and the *root mean square prediction error* (RMSE) for predicting *consistency*, as follows.

$$PLCC = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(MOS_i - \overline{MOS})}{\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2(MOS_i - \overline{MOS})^2}},$$

$$SROCC = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)},$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_i - MOS_i)^2}, \quad (21)$$

TABLE II

PERFORMANCE COMPARISONS OF DIFFERENT IQA MODELS EXPERIMENTED ON OUR ESTABLISHED *SCID* DATABASE

Criteria	Distortions	PSNR	SSIM [8]	MSSIM [20]	IWSSIM [46]	VIF [16]	IFC [17]	MAD [47]	FSIM [36]	ES [12]	GSIM [13]	GMSD [14]	VSI [15]	SCQI [48]	SIQM [30]	EMSQA [32]	ESIM
PLCC	GN	0.9530	0.9354	0.9448	0.9431	**0.9699**	0.8897	0.9315	0.9516	0.8718	0.9170	0.9273	**0.9556**	0.9319	0.9269	0.9521	**0.9563**
	GB	0.7772	0.8711	0.8991	**0.9174**	**0.8999**	0.8406	0.8559	0.8493	0.7555	0.8449	0.7348	0.8307	0.8244	**0.9266**	0.8380	0.8700
	MB	0.7615	0.8794	**0.9018**	**0.9055**	0.8421	0.3372	0.8362	0.8523	0.7499	0.8383	0.7954	0.8177	0.8147	**0.9152**	0.8451	0.8824
	CC	0.7435	0.6903	**0.8881**	**0.8989**	0.8092	0.1198	0.4987	**0.8947**	0.8653	0.8675	0.8041	0.8093	0.8353	0.7821	0.7975	0.7908
	JPEG	0.8393	0.8581	0.9227	0.9308	**0.9418**	0.8762	0.9251	**0.9419**	0.8983	0.9373	0.9351	0.9140	0.9036	0.9226	0.9312	**0.9421**
	J2K	0.9176	0.8586	0.9185	0.9195	**0.9489**	0.8570	0.9381	**0.9607**	0.9139	0.9441	0.9422	0.9441	0.9312	0.9076	0.9408	**0.9457**
	CSC	0.0622	0.0890	0.0977	0.0527	0.0898	0.0764	**0.1296**	0.0966	0.0586	0.0560	0.0952	**0.9119**	**0.8393**	0.0683	0.0626	0.0694
	HEVC-SCC	0.7991	0.7914	0.8635	0.8883	0.8656	0.7918	0.8953	**0.9228**	0.8650	0.8835	0.9043	0.9035	0.8708	0.8316	**0.9101**	**0.9108**
	CQD	**0.9210**	0.7810	0.8668	0.8930	0.9085	0.7655	0.9014	**0.9202**	0.8467	0.8974	**0.9177**	0.8873	0.8823	0.8385	0.9056	0.9005
	Overall	0.7622	0.7343	0.7579	0.7877	0.8179	0.6285	0.7736	0.7719	0.6755	0.7042	**0.8337**	0.7694	0.7596	0.8304	**0.8521**	**0.8630**
SROCC	GN	0.9424	0.9171	0.9309	0.9305	**0.9616**	0.8877	0.9262	0.9378	0.8854	0.9112	0.9341	0.9455	**0.9556**	0.9133	0.9423	**0.9460**
	GB	0.7702	0.8698	0.8949	**0.9165**	**0.8954**	0.8351	0.8603	0.8476	0.8311	0.8420	0.7931	0.8221	0.8638	**0.9232**	0.8381	0.8699
	MB	0.7375	0.8588	**0.8890**	0.8918	0.8259	0.4477	0.8296	0.8370	0.8240	0.8194	0.8148	0.8013	0.8587	**0.9006**	0.8596	0.8608
	CC	0.7265	0.6564	**0.8368**	**0.8475**	0.6115	0.1198	0.4784	**0.8473**	0.8177	0.8304	0.5672	0.8158	0.7465	0.7435	0.6511	0.6182
	JPEG	0.8321	0.8490	0.9219	0.9275	0.9349	0.8770	0.9242	**0.9403**	0.9296	**0.9366**	0.9344	0.9142	0.9171	0.9158	0.9206	**0.9455**
	J2K	0.9074	0.8439	0.9097	0.9067	**0.9369**	0.8457	0.9330	**0.9484**	0.9195	0.9349	0.9279	0.9307	0.9270	0.8935	0.9191	**0.9359**
	CSC	0.0908	0.0963	0.1274	0.1336	0.1221	0.0521	**0.1440**	0.1182	0.1400	0.1214	0.1165	**0.9141**	**0.8970**	0.0617	0.0991	0.1037
	HEVC-SCC	0.8074	0.8263	0.8688	0.8867	0.8580	0.7869	0.8771	**0.9098**	0.8663	0.8730	0.8958	0.8929	0.8721	0.8517	**0.9008**	**0.9036**
	CQD	**0.9080**	0.7766	0.8626	0.8846	0.8918	0.7368	0.9024	**0.9077**	0.8634	0.8707	0.9047	0.8820	**0.9099**	0.8301	0.8908	0.8868
	Overall	0.7512	0.7146	0.7407	0.7714	0.7969	0.5799	0.7576	0.7550	0.6993	0.6945	**0.8138**	0.7621	0.7814	0.8086	**0.8390**	**0.8478**
RMSE	GN	3.8093	4.4458	4.1180	4.1780	**3.0629**	5.7380	4.5714	3.8613	6.1581	5.0127	4.7044	**3.7138**	4.5600	4.7185	3.6989	**3.6760**
	GB	6.6633	5.1998	4.6360	**4.2163**	**4.6179**	5.7354	5.4775	5.5903	6.9381	5.6648	7.1821	5.8956	5.9943	**3.9831**	5.7779	5.2213
	MB	7.0843	5.2044	**4.7233**	**4.6376**	5.8960	10.2905	5.9947	5.7180	7.2316	5.9607	6.6249	6.2922	6.3394	**4.4040**	5.2901	5.1431
	CC	5.9867	6.4767	**4.1151**	**3.9218**	5.2594	8.8876	7.7590	**3.9979**	4.4875	4.4524	5.3211	5.2583	4.9217	5.5783	5.4008	5.4790
	JPEG	8.1718	7.7179	5.7955	5.4930	**5.0536**	7.2431	5.7076	**5.0471**	6.6052	5.2369	5.3275	6.0971	6.4390	5.7974	5.4785	**5.0373**
	J2K	6.3222	8.1562	6.2890	6.2555	**5.0207**	8.1986	5.5103	**4.4180**	6.4580	5.2462	5.3283	5.2451	5.8002	6.6779	5.2156	**5.1695**
	CSC	9.8203	9.8003	9.7923	9.8257	9.7996	9.8106	**9.7564**	9.7933	9.8224	9.8239	9.7947	**4.0392**	**5.3503**	9.8394	9.8200	9.8156
	HEVC-SCC	8.4009	8.5037	7.0166	6.3904	6.9657	8.4969	6.1988	**5.3583**	6.9795	6.5176	5.9393	5.9628	6.8407	7.6936	**5.7935**	**5.7446**
	CQD	**4.9814**	7.9855	6.3768	5.7530	5.3440	8.2269	5.5354	**5.0054**	6.8043	5.6406	**5.0796**	5.8964	6.0188	6.9661	5.4226	5.5607
	Overall	9.1682	9.6133	9.2400	8.7243	8.1479	11.0157	8.9739	9.0040	10.4433	10.0552	**7.8210**	9.0456	9.2113	7.8910	**7.3478**	**7.1552**

where n denotes the total distorted images involved in the evaluation of the IQA model. Parameters $\overline{MOS}$ and $\overline{Q}$ are the mean values of MOS_i and Q_i, respectively, d_i is the difference between the i-th image's rank in the subjective and objective evaluations. Lastly, it is important to note that for the PLCC and SROCC, the *larger* of their values, the better the performance. For the RMSE, this trend is reversed; that is, the smaller the better.

B. Performance Comparison and Analysis

To demonstrate the superiority, the proposed ESIM model is compared with the classical and several state-of-the-art quality assessment models, including PSNR, SSIM [8], MSSIM [20], IWSSIM [46], VIF [16], IFC [17], MAD [47], FSIM (for gray-scale image) [36], ES [12], GSIM [13], GMSD [14], VSI [15], SCQI [48], SIQM [30], SPQA [1], SQI [29], and EMSQA [32], where the last four IQA models are specifically designed for the evaluation of SCIs, while the rest are all for assessing natural images. Note that the parameter T_ψ of the proposed ESIM model as shown in (11) needs to be set for each edge attribute case; i.e., EWM, ECM, and EDM. For that, a subset of SCIs chosen from SIQAD database, which contains 8 reference SCIs and their 392 distorted versions, are used for the determination of these parameters. Following the same practices as suggested in [13] and [15], those parameter values that lead to higher SROCC will be selected. Through extensive

experiments, the values of T_ψ are empirically determined for EWM, ECM, and EDM as 0.9, 800, and 10, respectively.

Tables I and II document the computed quality assessment measurements using these methods for the SCIs from the SIQAD database and our SCID database, respectively. The best three performance figures of each measurement criterion (i.e., PLCC, SROCC, and RMSE) are highlighted in boldface, and the best one is further highlighted in red color for ease of comparison. Note that the program codes of all the above-mentioned models are downloaded from their original sources, except for the two SCI ones (i.e., SPQA and SQI). Therefore, the results of SPQA and SQI experimented on our established SCID database are not available for being included in Table II. Moreover, the performance comparisons conducted on each distortion type are also provided, except for the SQI for the same reason that does not provide the RMSE result on each distortion.

From the experimental results documented in Tables I and II, one can see that the proposed ESIM model outperforms all the state-of-the-art models under comparison, as it is able to achieve the highest correlation or consistency with the subjective quality ratings judged by the HVS. Note that the proposed ESIM model also significantly outperforms other state-of-the-art *edge*-based IQA models; i.e., ES, GSIM, and GMSD. This reveals that the proposed ESIM model that exploits three edge attributes (i.e., edge width,
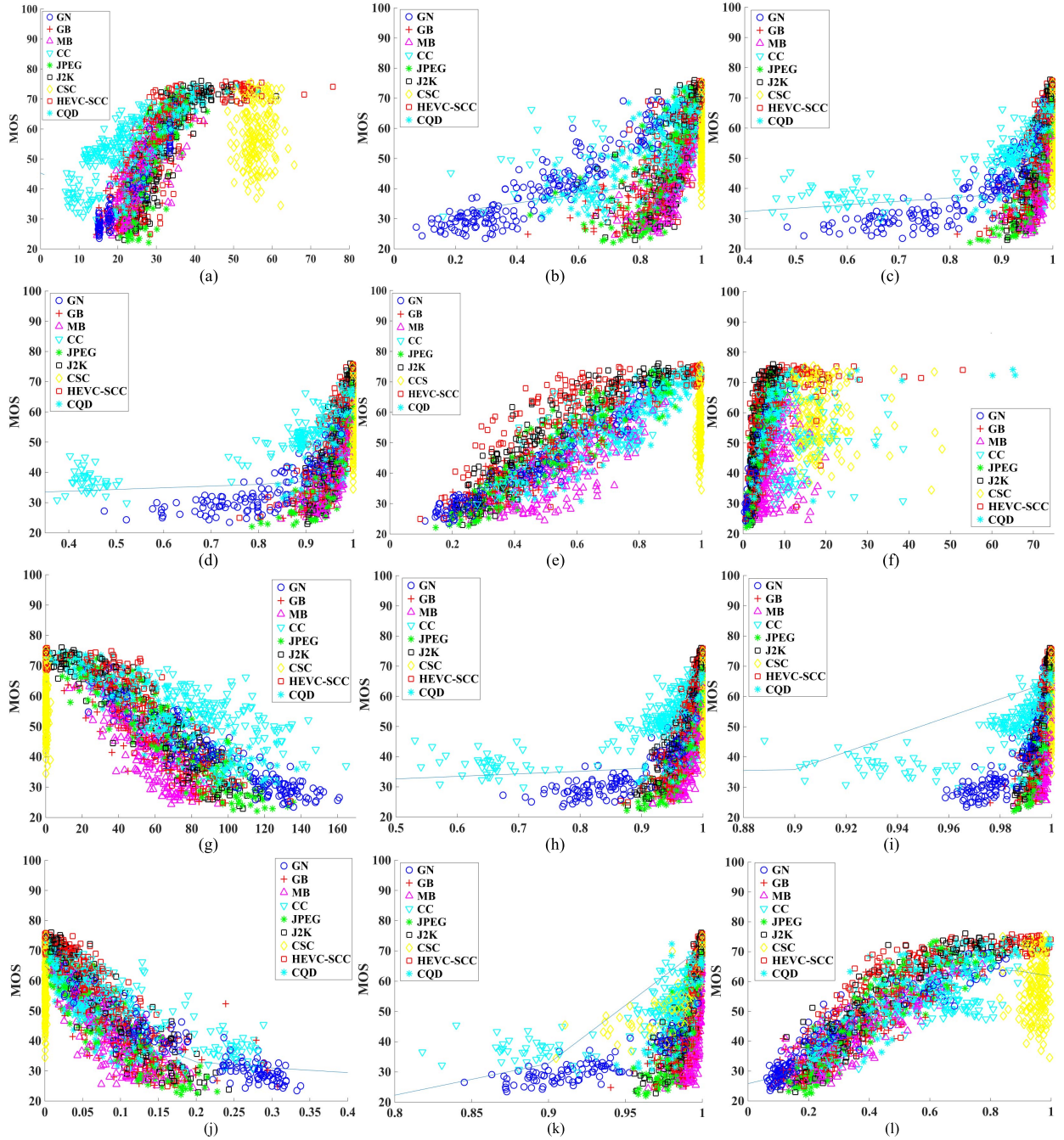
Fig. 8. Scatter plots of the MOS versus the IQA metrics using our established SCID database: (a) PSNR; (b) SSIM; (c) MSSIM; (d) IWSSIM; (e) VIF; (f) IFC; (g) MAD; (h) FSIM; (i) GSIM; (j) GMSD; (k) VSI; and (l) Proposed ESIM, respectively. The blue-color line as shown in each sub-plot is obtained by using a curve fitting procedure according to (20). One can see that the data points on the scatter plot of the proposed ESIM are much closer to this curve (thus, a better fitting to the curve).

edge contrast, and edge direction) is much more effective to explore the edge information for conducting SCI perceptual quality assessment. Another interesting observation is that all the models specifically designed for the SCI quality assessment (i.e., SIQM, SPQA, SQI, EMSQA, and our proposed ESIM) can achieve better performance than those IQA models developed for the *natural* images in SIQAD database. This is mainly due to the fact that they consider the special characteristics of the SCI on the design of SCI quality assessment models.

To further visualize the performance yielded by the IQA models under comparison, Fig. 8 shows the scatter plots of the subjective scores against the objective scores as predicted by some representative IQA models (i.e., PSNR, SSIM, MSSIM, IWSSIM, VIF, IFC, MAD, FSIM, GSIM, GMSD, VSI, and the proposed ESIM) based on our SCID database for a demonstration. The curve as shown in each sub-plot of Fig. 8 is obtained by a nonlinear curve fitting procedure according to (20). One can see that our proposed ESIM has a 'tighter' (thus, better) curve fitting, compared with that of other IQA models.

TABLE III

PERFORMANCE RESULTED FROM EACH EDGE ATTRIBUTE
OF THE PROPOSED ESIM MODEL ON THE *SIQAD* AND
OUR ESTABLISHED *SCID* DATABASES

Database	Criteria	EWS	ECS	EDS	ESIM
SIQAD	PLCC	0.8355	0.8528	0.8060	0.8788
	SROCC	0.8190	0.8369	0.7777	0.8632
	RMSE	7.8654	7.4744	8.4731	6.8310
SCID	PLCC	0.8119	0.8504	0.7634	0.8630
	SROCC	0.7877	0.8382	0.7102	0.8478
	RMSE	8.2683	7.4514	9.1482	7.1552

To more comprehensively evaluate an IQA index's ability to predict image quality degradations caused by specific types of distortions, we examine the performance of the proposed ESIM and other state-of-the-art IQA models on each type of distortions, as shown in Tables I and II. For each distortion type, the top three performance figures of each measurement scheme are highlighted in boldface and the best one is further highlighted in red color. One can see that the proposed ESIM yield the top performance on most of distortion types. Moreover, it can be observed that the proposed ESIM model is able to more accurately assess and reflect the degradations caused by Gaussian blur (GB), motion blur (MB), JPEG compression, and JPEG2000 compression particularly. In fact, this is expected, since blurring and compression artifacts will inevitably degrade edges and make significant changes on the extracted edge information, which would be accurately reflected by the edge attributes used in the proposed ESIM model. Despite the encouraging performance results achieved by the proposed ESIM, it has a relatively poor performance on Contrast change (CC) and Color Saturation Change (CSC). This is because CC only affects the intensity of image rather than the image structure while CSC mainly changes the chrominance component rather than the luminance component. Future work will include devising a model to predict the SCI quality by further considering the intensity changes and chrominance component based on the proposed ESIM model.

Lastly, to analyze how much of the contributions coming from each edge attribute in the proposed ESIM model, the performance on the evaluation of the SCIs resulted from each of three edge attributes (i.e., EWS, ECS, and EDS) are investigated using the SIQAD database and our SCID database. This can be accomplished by assigning different values to α, β, and γ in (12). Specifically, for considering edge width only (i.e., EWS), $\beta = 1$ and $\alpha = \gamma = 0$; for the edge contrast only (i.e., ECS), $\alpha = 1$ and $\beta = \gamma = 0$; and, for the edge direction only (i.e., EDS), $\gamma = 1$ and $\alpha = \beta = 0$. The corresponding experimental results are recorded in Table III. One can see that each edge attribute achieves relatively good performances, and the proposed ESIM that jointly explores three edge attributes performs the best. This investigation suggests that the edge width, edge contrast, and edge direction effectively conveys different attributes of edge, and they play a complementary role to jointly capture the edge properties well for delivering accurate quality assessment for the SCIs.

V. CONCLUSIONS

In this paper, a novel *screen content image* (SCI) quality assessment model, called the *edge similarity* (ESIM), is proposed. The novelty of the ESIM lies in the use of three salient edge attributes—i.e., edge contrast, edge width, and edge direction, individually extracted from the reference and distorted SCIs for conducting their respective similarity measurements, followed by pooling them together to produce the final ESIM score. Another key contribution in our work is that a large SCI database (i.e., SCID) is established and served as the 'ground truth' to quantitatively assess how accurate of the IQA scores computed by the proposed ESIM and other existing state-of-the-art models on the perceptual quality assessment of SCIs. Extensive experiments conducted over an existing SCI database and our established SCID database have clearly demonstrated that the proposed ESIM model delivers the highest performance compared to other state-of-the-art IQA models on providing more accurate and consistent assessment in accordance with what the HVS perceives and judges the SCIs.

REFERENCES

[1] 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, Aug. 2015.

[2] S. Wang, L. Ma, Y. Fang, W. Lin, S. Ma, and W. Gao, "Just noticeable difference estimation for screen content images," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3838–3851, May 2016.

[3] 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.

[4] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Learning a blind quality evaluation engine of screen content images," *Neurocomputing*, vol. 196, pp. 140–149, Jul. 2016.

[5] Z. Ma, W. Wang, M. Xu, and H. Yu, "Advanced screen content coding using color table and index map," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4399–4412, Oct. 2014.

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

[7] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*. Cambridge, MA, USA: MIT Press, 1993, pp. 207–220.

[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. 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, Aug. 2016.

[10] G. Chen, C. Yang, and S. Xie, "Edge-based structural similarity for image quality assessment," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2006, pp. 14–19.

[11] W. Xue and X. Mou, "An image quality assessment metric based on non-shift edge," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3309–3312.

[12] X. Zhang, X. Feng, W. Wang, and W. Xue, "Edge strength similarity for image quality assessment," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 319–322, Apr. 2013.

[13] 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.

[14] 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.

[15] 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, Aug. 2014.

[16] 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.

[17] 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.

[18] F. Gao and J. Yu, "Biologically inspired image quality assessment," *Signal Process.*, vol. 124, pp. 210–219, Jul. 2016.

[19] C. Deng and D. Tao, "Color image quality assessment with biologically inspired feature and machine learning," *Vis. Commun. Image Process.*, vol. 124, pp. 77440Y-1–77440Y–7, Aug. 2010.

[20] Z. Wang, E. P. Simoncelli, and A. C. Bovil, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Conf. Signals Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.

[21] F. Gao, D. Tao, X. Gao, and X. Li, "Learning to rank for blind image quality assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2275–2290, Oct. 2015.

[22] T. Liu, K. Liu, J. Lin, W. Lin, and C.-C. J. Kuo, "A paraboost method to image quality assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 107–121, Jan. 2017.

[23] S. Wang, C. Deng, W. Lin, G. Huang, and B. Zhao, "NMF-based image quality assessment using extreme learning machine," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 232–243, Jan. 2017.

[24] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.

[25] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognit.*, vol. 59, pp. 176–187, Nov. 2016.

[26] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "Deepsim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, Feb. 2017.

[27] H. Wang, J. Fu, W. Lin, S. Hu, C.-C. J. Kuo, and L. Zuo, "Image quality assessment based on local linear information and distortion-specific compensation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 915–926, Feb. 2017.

[28] S. Bosse, D. Maniry, and K. Müller, T. Wiegand, and W. Samek. (Dec. 2016). "Deep neural networks for no-reference and full-reference image quality assessment." [Online]. Available: https://arxiv.org/abs/1612.01697

[29] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Objective quality assessment and perceptual compression of screen content images," *IEEE Comput. Graph. Appl.*, May 2016, doi: 10.1109/MCG.2016.46.

[30] K. Gu, S. Wang, G. Zhai, S. Ma, and W. Lin, "Screen image quality assessment incorporating structural degradation measurement," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2015, pp. 125–128.

[31] P. J. L. van Beek, "Edge-based image representation and coding," Ph.D. dissertation, Dept. Elect. Eng., Delft Univ. Technol., Delft, The Netherlands, 1995.

[32] Z. Ni, L. Ma, H. Zeng, C. Cai, and K.-K. Ma, "Screen content image quality assessment using edge model," in *Proc. IEEE Int. Conf. Image Process.*, Aug. 2016, pp. 81–85.

[33] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, Mar. 2007.

[34] D. Song and D. Tao, "C1 units for scene classification," in *Proc. 19th Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[35] D. Song and D. Tao, "Biologically inspired feature manifold for scene classification," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 174–184, Jan. 2010.

[36] 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.

[37] H. Zeng, A. Yang, K. N. Ngan, and M. H. Wang, "Perceptual sensitivity-based rate control method for High Efficiency Video Coding," *Multimedia Tools Appl.*, vol. 75, pp. 10383–10396, Oct. 2015.

[38] A. Yang, H. Zeng, J. Chen, J. Zhu, and C. Cai, "Perceptual feature guided rate distortion optimization for High Efficiency Video Coding," *Multidimensional Syst. Signal Process.*, Mar. 2016, pp. 1–18, doi: 10.1007/s11045-016-0395-2.

[39] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-11, International Telecommunications Union, 2012.

[40] M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," *Proc. SPIE Vis. Commun. Image Process.*, pp. 573–582, Jun. 2003.

[41] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, 2013, Art. no. 905685.

[42] A. M. Dijk, J.-B. Martens, and A. B. Waston, "Quality assessment of coded images using numerical category scaling," *Proc. SPIE Adv. Image Video Commun. Storage Technol.*, pp. 90–101, Feb. 1995.

[43] 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.

[44] VQEG. (Aug. 2015). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx

[45] 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.

[46] 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.

[47] 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, 2010.

[48] S.-H. Bae and M. Kim, "A novel image quality assessment with globally and locally consilient visual quality perception," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2392–2406, Apr. 2016.
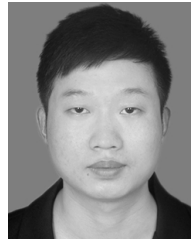
Zhangkai Ni received the B.E. degree in communication engineering from Anhui Normal University, Wuhu, China.

He is currently pursuing the M.E. degree with the School of Information Science and Engineering, Huaqiao University, China. His research interests include visual quality assessment, perceptual signal processing, and computer vision.
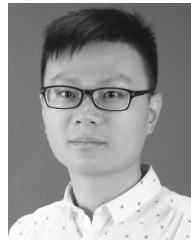
Lin Ma (M'13) received the B.E. and M.E. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, the Ph.D. degree from the Department of Electronic Engineering, The Chinese University of Hong Kong in 2013. He was a Researcher with Huawei Noah's Ark Laboratory, Hong Kong, from 2013 to 2016. He is currently a Senior Research Engineer with the Tencent AI Laboratory, Shenzhen, China. His current research interests lie in the areas of deep learning and multimodal learning, specifically for image and language, image/video understanding, and quality assessment.

Dr. Ma received the Best Paper Award from the Pacific-Rim Conference on Multimedia in 2008. He was a recipient of the Microsoft Research Asia Fellowship in 2011. He was a Finalist to HKIS Young Scientist Award in engineering science in 2012.

Huanqiang Zeng (S'10–M'13) received the B.S. and M.S. degrees from Huaqiao University, Xiamen, China, and the Ph.D. degree from Nanyang Technological University Singapore, Singapore, all in electrical engineering.

He is currently a Professor with the School of Information Science and Engineering, Huaqiao University, Xiamen, China. He was a Post-Doctoral Fellow with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, from 2012 to 2013, and a Research Associate with the Temasek Laboratories, Nanyang Technological University, Singapore, in 2008.

He has authored over 50 papers in well-known international journals and conferences. His research interests are in the areas of visual information processing and analysis, image/video communication, and computer vision. He has been actively serving as an Associate/Guest Editor of multiple international journals, a General Co-Chair of the 2017 IEEE International Symposium on Intelligent Signal Processing and Communication Systems, a Technical Co-Chair of 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, an Area Chair of the 2015 IEEE International Conference on Visual Communications and Image Processing, and a Technical Program Committee Member of multiple international conferences.

Jing Chen received the B.S. and M.S. degrees from Huaqiao University, Xiamen, China, and the Ph.D. degree from Xiamen University, Xiamen, China, all in computer science.

She is currently an Associate Professor with the School of Information Science and Engineering, Huaqiao University. Her current research interests include image processing, video coding, and multiple description coding.

Canhui Cai (M'06–SM'08) received the B.S. degree from Xidian University, Xi'an, China, in 1982, the M.S. degree from Shanghai University, Shanghai, China in 1985, and Ph.D. degree from Tianjin University, Tianjin, China, in 2003, all in electronic engineering.

Since 1984, he has been with the Faculty of Huaqiao University, Quanzhou, China, where he is currently a Professor with the School of Engineering. He was a Visiting Professor with the Delft University of Technology, Delft, The Netherlands, from 1991 to 1992, and a Visiting Professor with the University of California at Santa Barbara, Santa Barbara, CA, USA, from 1999 to 2000. He has authored or co-authored four books, and has authored over 150 papers in journals and conference proceedings. His research areas include video communications, image and video signal processing, and computer vision.

Dr. Cai was a General Co-Chair of the Intelligent Signal Processing and Communication Systems in 2007.

Kai-Kuang Ma (S'80–M'84–SM'95–F'13) received the Ph.D. degree in electrical and computer engineering from North Carolina State University, Raleigh, NC, USA. He was a member of the Technical Staff with the Institute of Microelectronics, from 1992 to 1995, where he was involved in digital video coding and the MPEG standards. From 1984 to 1992, he was with the IBM Corporation, Kingston, NY, USA, and then with Research Triangle Park, NC, USA, where he was involved in various DSP and VLSI advanced product development. He is currently a Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University Singapore, Singapore. He has authored extensively in well-known international journals, conferences, and MPEG standardization meetings. He holds one U.S. patent on fast motion estimation algorithm. His research interests are in the areas of fundamental image/video processing and applied computer vision.

He was serving as the Singapore MPEG Chairman and the Head of Delegation from 1997 to 2001. On the MPEG contributions, two fast motion estimation algorithms (Diamond Search and MVFAST) produced from his research group have been adopted by the MPEG-4 standard, as the reference core technology for fast motion estimation. He is the General Chair of organizing a series of international standard meetings (MPEG and JPEG), JPEG2000, and MPEG-7 workshops held in Singapore in 2001.

He was elected as a Distinguished Lecturer of the IEEE Circuits and Systems Society for 2008-2009. He has served various roles in professional societies, such as the General Co-Chair of APSIPA-2017, ISPACS-2017, ACCV-2016 Workshop, and VCIP-2013, the Technical Program Co-Chair of the ICASSP-2022, ICIP-2004, ISPACS-2007, IIH-MSP-2009, and PSIVT-2010, and the Area Chair of the ACCV-2009 and ACCV-2010. He also had extensive editorship contributions in several international journals, such as the IEEE TRANSACTIONS ON IMAGE PROCESSING as an Associate Editor from 2007 to 2010 and has been a Senior Area Editor since 2015, has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY since 2015, the IEEE SIGNAL PROCESSING LETTERS as an Associate Editor from 2014 to 2016, the IEEE TRANSACTIONS ON COMMUNICATIONS an Editor from 1997 to 2012, the IEEE Transactions on Multimedia as an Associate Editor from 2002 to 2009, the *Journal of Visual Communication and Image Representation* as an Editorial Board Member from 2005 to 2014. He is an Elected Member of three IEEE Technical Committees, such as the Image and Multidimensional Signal Processing Committee, the Multimedia Communications Committee, and the Digital Signal Processing. He was the Chairman of the IEEE Signal Processing, Singapore Chapter from 2000 to 2002.