

Perceptual Quality Assessment of Screen Content Images

Huan Yang, Yuming Fang, and Weisi Lin, *Senior Member, IEEE*

Abstract—Research on screen content images (SCIs) becomes important as they are increasingly used in multi-device communication applications. In this paper, we present a study on perceptual quality assessment of distorted SCIs subjectively and objectively. We construct a large-scale screen image quality assessment database (SIQAD) consisting of 20 source and 980 distorted SCIs. In order to get the subjective quality scores and investigate, which part (text or picture) contributes more to the overall visual quality, the single stimulus methodology with 11 point numerical scale is employed to obtain three kinds of subjective scores corresponding to the entire, textual, and pictorial regions, respectively. According to the analysis of subjective data, we propose a weighting strategy to account for the correlation among these three kinds of subjective scores. Furthermore, we design an objective metric to measure the visual quality of distorted SCIs by considering the visual difference of textual and pictorial regions. The experimental results demonstrate that the proposed SCI perceptual quality assessment scheme, consisting of the objective metric and the weighting strategy, can achieve better performance than 11 state-of-the-art IQA methods. To the best of our knowledge, the SIQAD is the first large-scale database published for quality evaluation of SCIs, and this research is the first attempt to explore the perceptual quality assessment of distorted SCIs.

Index Terms—Screen content image, quality assessment, subjective quality assessment, objective quality assessment.

I. INTRODUCTION

SCREEN Content Images (SCIs), which include texts, graphics and pictures together, have been increasingly involved in multi-client communication systems, such as virtual screen sharing [1], information sharing between computer and smart phones [2], cloud computing and gaming [3], remote education, product advertising, etc. In these systems, visual content (e.g., web pages, emails, slide files and computer screens) is typically rendered in the form of SCIs, and then transmitted between different digital devices (computers, tablets or smart phones). For fast sharing among different devices, it is important to acquire, compress,

store or transmit SCIs efficiently. Numerous solutions have been proposed to process SCIs, including segmentation and compression of SCIs [4]–[9]. Lately, MPEG/VCEG called for proposals to efficiently compress screen content image/videos as an extension of the HEVC standard, and many proposals have been reported to address this need [10].

When SCIs are processed, various distortions may be involved, such as blurring, contrast change and compression artifacts. For example, when we capture SCIs by smart phones, blurring appears on images along with hand-shake or out-of-focus of camera. Different settings of brightness or contrast of screens would result in the contrast change of captured SCIs. Compression artifacts (e.g., blocking and quantization noises) commonly appear on encoded SCIs. *Peak Signal-to-Noise Ratio* (PSNR) may be adopted in the aforementioned proposals to evaluate the visual quality of processed SCIs. However, it is known that PSNR is not consistent with human visual perception [11], [12]. *Quality of Experience* (QoE) has been investigated to evaluate users' viewing experience on webpages, which is called Web QoE [13]. Unfortunately, the current Web QoE mainly focuses on *Quality of Service* (QoS) metrics, e.g., loss ratio, rendering and round-trip time, rather than taking differences of human perception for pictures and texts into account [14], [15]. In these cases, the predicted QoS values would be constant if overall loss ratio is determined. However, different loss ratios to pictorial and textual parts may lead to quite different QoE. Therefore, perceptual quality assessment of SCIs is much desired for various applications. Although many IQA methods have been proposed for quality assessment of natural images [16], whether these IQA methods can be applicable to SCIs is still an open question. Hence, it is meaningful to investigate both subjective and objective metrics for the quality evaluation of SCIs.

In this work, we aim to carry out the first in-depth study on perceptual quality assessment of SCIs from both subjective and objective aspects. A large-scale *Screen Image Quality Assessment Database* (SIQAD) is built for the subjective test, in which three subjective quality scores are obtained respectively for the entire, textual and pictorial regions of each test image. The discrete 11 scale *Single Stimulus* (SS) method is adopted to carry out the subjective test. According to the analysis of subjective data, we propose a new scheme, *SCI Perceptual Quality Assessment* (SPQA), to objectively evaluate the visual quality of distorted SCIs. The SPQA consists of an objective metric and a weighting strategy. The objective metric is designed to separately evaluate the visual quality of textual and pictorial regions. In particular, a new

Manuscript received November 4, 2014; revised March 10, 2015 and May 25, 2015; accepted July 28, 2015. Date of publication August 5, 2015; date of current version August 18, 2015. This work was supported in part by the National Science Foundation of China, and in part by the Social Responsibility Foundation for Returned Overseas Chinese Scholars, State Education Ministry, China. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick Le Callet. (Corresponding author: Yuming Fang.)

H. Yang and W. Lin are with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: hyang3@e.ntu.edu.sg; wslin@ntu.edu.sg).

Y. Fang is with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: fa0001ng@e.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.2015.2465145

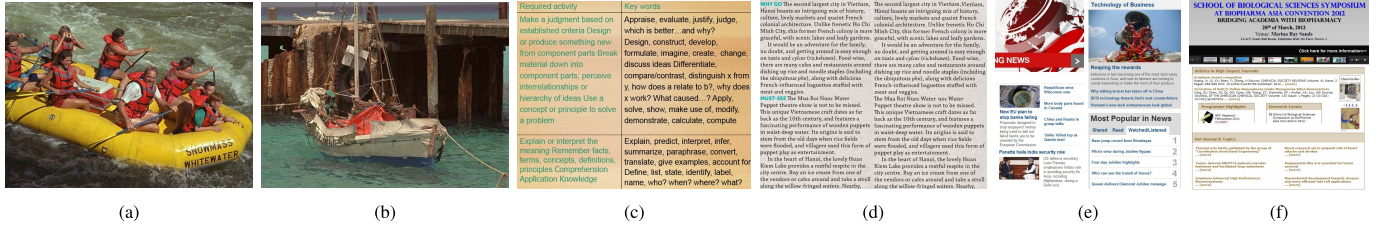


Fig. 1. Examples of natural images, textual images and screen content images. (a) image1. (b) image2. (c) text image1. (d) text image2. (e) screen image1. (f) screen image2.

scheme is designed to adaptively adjust the effect of luminance and sharpness variations in SCIs to human visual perception. The weighting strategy is designed to combine the predicted quality scores of textual and pictorial regions to obtain the overall quality scores of tested SCIs. Compared with 11 state-of-the-art IQA metrics, the proposed SPQA scheme achieves much higher consistency with human visual perception when judging the quality of distorted SCIs.

II. RELATED WORK

Natural Image Quality Assessment (NIQA) has been studied tremendously during the last decades [16], [17]. Several image quality assessment databases [18]–[22] have been constructed by adopting subjective testing strategies [23]. Based upon these databases, various *Full Reference* (FR) IQA methods [21], [24]–[28], such as SSIM, VIF, FSIM, MAD, GSI and GMSD, have been proposed to objectively assess the quality of distorted natural images. Besides, many *Reduced Reference* (RR) IQA [29] and *No Reference* (NR) IQA metrics [30] are also reported.

Document Image Quality Assessment (DIQA) has also attracted attention in the research community recently due to the increasing requirements of digitization of historical or other typewritten documents [31]. Many document image databases [32]–[35] are released, based on which various DIQA methods have been proposed [36]–[38]. The document images in these databases mainly consist of gray-scale or binary texts, without pictures. Most of these document images suffer from degradations related to the environment, e.g., paper aging, stains, carbon copy effect and reader annotations. Almost all the DIQA methods are designed in no-reference manner and implemented at the character (or string) level. The effectiveness of the DIQA methods is finally evaluated by the *Optical Character Recognition* (OCR) accuracy calculated by the OCR software rather than human visual judgement.

The topic of *Screen Image Quality Assessment* (SIQA) remains relatively un-explored. Obviously, the DIQA methods cannot be adopted to evaluate the visual quality of SCIs directly, since SCIs include pictorial regions besides textual regions and do not have the aforementioned environment-related degradations. The NIQA metrics cannot be directly applied to evaluate the quality of distorted SCIs either, since the statistical features of SCIs are different from those of natural images [4], [39], especially for the textual regions. We provide some natural, text and screen image examples in Fig. 1. The statistical differences of natural and screen

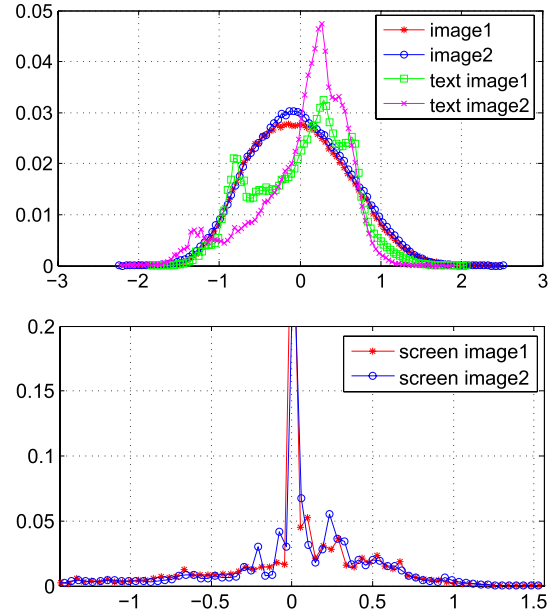


Fig. 2. Distribution of values in the naturalness maps of the example images.

images can be measured in terms of naturalness and activity level. The *naturalness* value of an image pixel $I(i, j)$ can be calculated as follows [40]:

$$N'(i, j) = \frac{I(i, j) - u(i, j)}{\sigma(i, j) + 1} \quad (1)$$

where $i \in \{1, 2, \dots, m\}$ and $j \in \{1, 2, \dots, n\}$ denote spatial indices; m and n represent the image dimension; the local mean $u(i, j)$ and deviation $\sigma(i, j)$ are computed as follows.

$$u(i, j) = \sum_{k=-K}^K \sum_{l=-L}^L \omega_{k,l} I(i+k, j+l) \quad (2)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L \omega_{k,l} [I(i+k, j+l) - u(i, j)]^2} \quad (3)$$

where ω is a 2D circularly-symmetric Gaussian weighting function with $K = L = 3$. We compute the distribution of coefficients $N'(i, j)$. The distributions of naturalness values of the example images are shown in Fig. 2. It can be observed that the coefficients of natural images follow a Gaussian distribution. In other words, the naturalness of a natural image is high, as demonstrated in [40], while for textual or screen images, the distributions vary greatly. For textual images (e.g., (c) and (d) in Fig. 1), the distribution

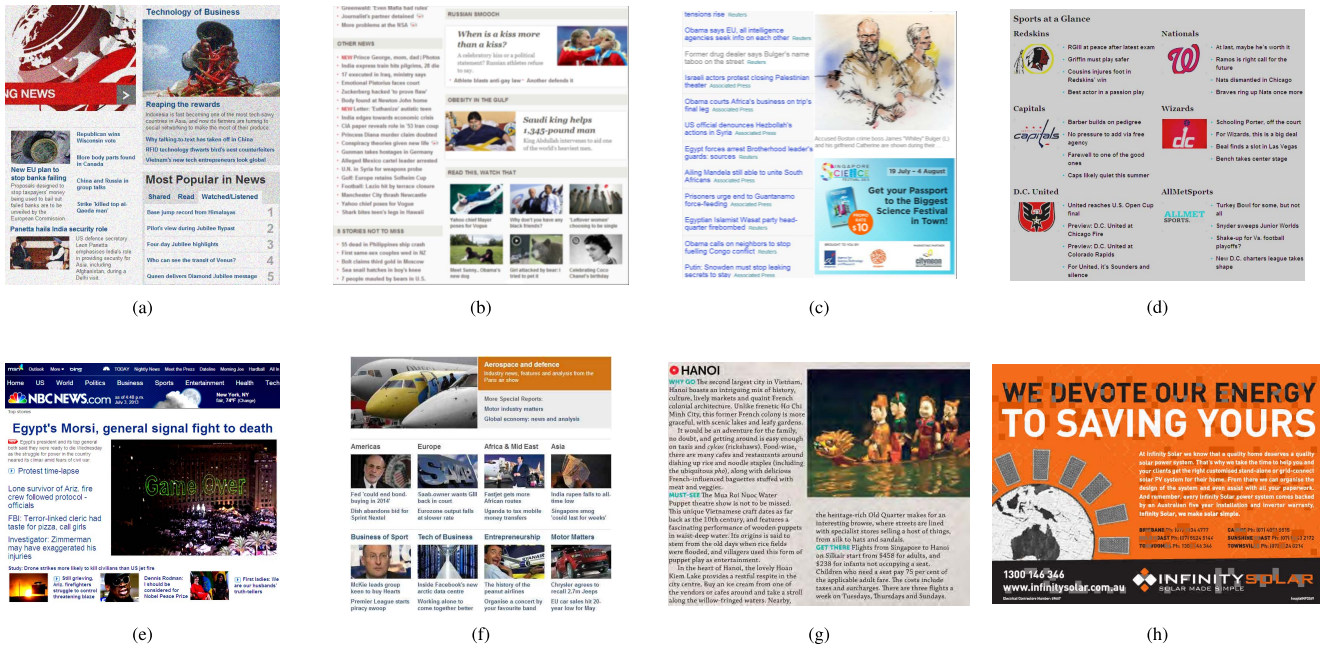


Fig. 3. Some image examples with different distortion types in *SIQAD* (refer to the images at the original resolution for better visual comparison). (a) SCI with Gaussian noise. (b) SCI with Gaussian blur. (c) SCI with motion blur. (d) SCI with contrast change. (e) SCI with contrast change. (f) SCI encoded by JPEG. (g) SCI encoded by JPEG2000. (h) SCI encoded by LSC.

curve fluctuates greatly; for screen images (e.g., (e) and (f) in Fig. 1), a sharp pimpling appears with the rest parts still waving. We utilize the *Block Activity Measure* (BAM) reported in [41] for the activity analysis. The image activity reflects the degree of pixel variations in local image regions. It has been demonstrated that the activity values of textual blocks are larger than those from the pictorial blocks, which confirms that a textual image has sharper and more intensive variation among neighboring pixel values than a natural image.

Hence, the NIQA metrics may not be applicable to evaluate the quality of distorted SCIs due to the statistical differences between natural and textual images. In this paper, we firstly study the subjective quality of distorted SCIs, and then further investigate the applicability of several state-of-the-art NIQA methods to distorted SCIs. Finally, a specific metric is proposed to objectively evaluate the visual quality of SCIs based on the in-depth analysis of the subjective data for SCIs.

III. SIQAD: SCI QUALITY ASSESSMENT DATABASE

To investigate quality evaluation of SCIs, we construct a large-scale screen image database (i.e., *SIQAD*) with seven distortion types, each with seven degradation levels. Totally, 20 reference and 980 distorted SCIs are included in the *SIQAD*. Subjective evaluation of these SCIs is conducted to obtain the subjective quality scores. All the SCIs and the corresponding subjective scores are now available [42], [43].

A. Construction of the *SIQAD*

In total, twenty SCIs are collected from webpages, slides, PDF files and digital magazines through screen snapshot. The reference SCIs are cropped from these twenty images to proper

sizes (the dimension scale is from about 600 to 900 pixels) for natively displaying on computer screens during the subjective test. The reference SCIs are selected with various layout styles, including different percentages, positions and ways of textual/pictorial region combination. The percentage of textual regions in the reference SCIs varies from 35% to 60%. Meanwhile, pictorial or textual regions are also diverse in visual content. Two examples of the reference SCIs are given in Fig. 1 (e) and (f), and some distorted SCIs with different distortion types are given in Fig. 3.

Seven distortion types which usually appear on SCIs are applied to generate distorted images. *Gaussian Noise* (GN) is often involved in image acquisition and included in most existing image quality databases [18], [19]. *Gaussian Blur* (GB) and *Motion Blur* (MB) are also considered due to their commonly existing in practical applications. For example, when SCIs are captured by digital cameras, hand-shaking, out-of-focus or object moving would bring blur into images. *Contrast Change* (CC) is also an important factor affecting peculiarities of the HVS. Different settings of brightness and contrast of screens will result in different visual experiences of viewers. As compression is widely used in most SCI-based applications, three commonly used compression algorithms are utilized to encode the reference SCIs: JPEG, JPEG2000 and *Layer Segmentation based Coding* (LSC) [7]. The JPEG and JPEG2000 are two widely used methods in image compression, and have been introduced into many quality assessment databases. We include LSC as another codec due to its efficient compression of SCIs. The LSC firstly separates SCIs into textual and pictorial blocks with a segmentation index map in which textual blocks are marked by one and pictorial blocks by zero. The textual layer is encoded by using the

Basic Colors and Index Map (BCIM) method [7] while the pictorial layer is encoded by the JPEG algorithm. Specifically, in order to investigate the effect of misclassification to visual quality, we artificially adjust the segmentation index map and randomly misclassify some textual blocks to pictorial ones with different misclassification ratios. Since the JPEG cannot effectively encode the misclassified textual regions, misclassification artifacts will appear on the compressed SCIs, as illustrated in Fig. 3 (h).

For each distortion type, seven degradation levels are set to generate images from low to high degradation levels, which create a broad range of image impairment. The detailed configurations of these algorithms, e.g., the standard deviation for GB, the scale variation of CC, the quality factor for JPEG, the misclassification ratio in LSC, are given in related supporting files in the *SIQAD* [43].

B. Subjective Testing Methodology

Subjective testing methodologies of image quality evaluation have been recommended by *International Telecommunications Union* (ITU) [23], [44], including *Single Stimulus* (SS), double-stimulus and paired comparison. In this study, the SS with an 11 point discrete scale is employed. Given one image displaying on the screen, the human subject is asked to give a score (from 0 to 10: 0 is the worst, and 10 is the best) on the image quality based on her/his visual perception. This methodology is chosen because the viewing experience of subjects is close to that in practice, where there is no access to the reference images [45]. The subjective test is performed using two identical desktops with 16 GB RAM and 64-bit Windows operating system. The desktops with calibrated 24-inch LED monitors (Dell P2412H) are placed in the laboratory with normal indoor lighting. Viewing conditions are in accordance with ITU Recommendation [23]. All subjects are required to sit at a viewing distance about 2–2.5 times of the screen height. The subjects are all university under-graduate or graduate students with no experience in image processing and quality assessment. The percentage of female subjects is about 40%. They are all with normal or corrected vision, aging from 19 to 38 years old.

Before the start of testing stage, subjects have to go through the training stage in which some examples with representative distortion types and levels are presented. These examples are not included in the testing stage. When judging SCIs, three aspects are mainly considered: content recognizability, clarity and viewing comfort. Content recognizability is used to check whether the content of distorted SCIs can be recognized. Content clarity is used to judge the impairment appearance on the images. Viewing comfort reflects subjects' viewing experience. We explain these three aspects to each subject in the training stage, and emphasize them at the beginning of the testing stage. The graphical user interface is shown in Fig. 4. Users give their judgment by clicking the radio buttons and have to finish all assigned images, otherwise their judgments will not be recorded.

In this study, we would like to not only get the overall quality scores of all distorted SCIs, but also investigate

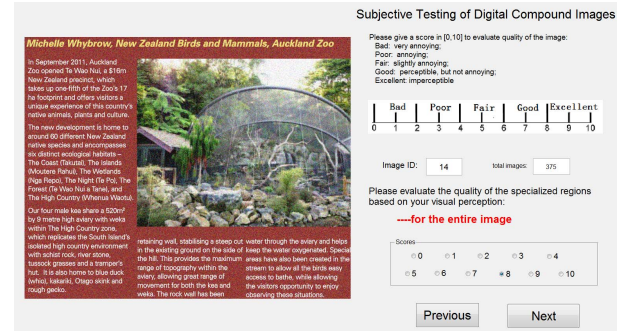


Fig. 4. Graphical user interface in the subjective test. The red tooltip will change if subjects need to judge different regions.

which part (text or picture) contributes more to the overall visual quality. Hence, subjects are required to give three scores to each test image, corresponding to overall, textual and pictorial regions, respectively. In this subjective test, all the reference images are included and tested. We generate a random permutation of 1,000 images (20 reference and 980 distorted SCIs) for each round, and make sure that every two consecutive images are not generated from the same reference image. We then split each permutation into 8 batches and assign one batch of 125 images to one subject at a time. Each of these 125 images is shown three times (not consecutively), and subjects give scores to one specific region at each time (reminded by the red tooltip on the user interface). After finishing the judgment of one region, subjects would take five minutes' break. It takes about one hour for each subject to finish all the judgements in one batch. In the experiment, one subject can finish the evaluation of several batches (e.g., 2-4 batches) at different time. Totally, 96 subjects take part in the study, and each image is evaluated by at least 30 subjects.

C. Analysis of Subjective Scores

When processing the raw subjective scores, outliers are firstly detected and rejected according to the method [18]. Totally, six subjects are rejected, and we delete all the subjective data reported by the rejected subjects. After outlier rejection, we follow the data processing steps utilized in [18], and transform the raw data to Z scores. Since we separate all the 1000 images into 8 sessions, scale realignment is then conducted to compensate the scale difference in different sessions [46], [47], as done in the LIVE database [18]. A set of 80 images (ten images were chosen from each session) is selected, including all distortion types at different distortion levels, and these images are re-evaluated by subjects (all the 80 images are evaluated by each subject in one round). A linear mapping function is also learned to convert Z scores to *Difference Mean Opinion Score* (DMOS) values. Finally, we normalize the DMOS values to a commonly used scale (i.e., 0-100). We repeat this procedure to the three groups of subjective scores for entire, textual and pictorial regions, respectively.

Generally, the quality scales of the distorted SCIs in the database should exhibit good separation of perceptual quality

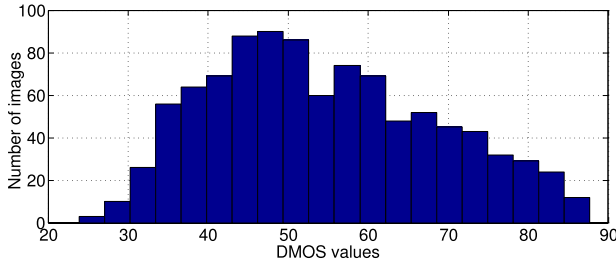


Fig. 5. Histogram of DMOS values of images in the SIQAD.

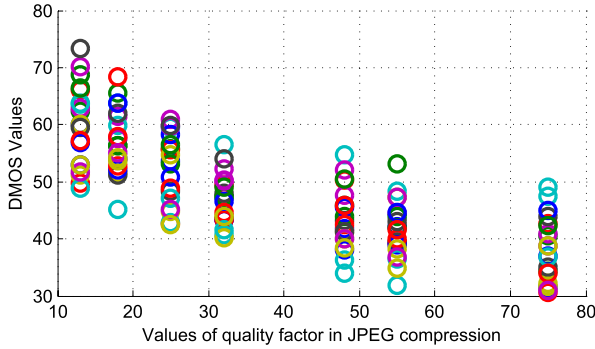


Fig. 6. Distribution of DMOS values of JPEG compressed SCIs.

and span the entire range of visual quality (from distortion imperceptible to severely annoying) [48]. Fig. 5 shows the histogram of the DMOS values (0:100) of all distorted images in the database. It can be observed that the DMOS values of images range from low to high, and have a good spread at different levels. Besides, the diversity of images in the constructed database can be well reflected from the distribution of DMOS values. In Fig. 6, perceptual qualities of all distorted images derived from JPEG compression at different levels are displayed. We can see that, at each compression level, the twenty distorted images derived from the twenty reference images have different perceptual quality scores.

We examine the consistency of all subjects' judgements of each image. According to [23] and [45], the consistency can be measured by the confidence interval derived from the number and standard deviation of scores for each image. With a probability of 95% confidence level, the difference between the computed DMOS value and the "true" quality value is smaller than the 95% confidence interval [23]. The mean values of confidence intervals according to the three regions (i.e., overall, textual and pictorial regions) are 3.00, 3.07 and 2.94, respectively. The distribution of confidence intervals related to the overall DMOS values is shown in Fig. 7. The confidence intervals for the textual and pictorial DMOS values have similar distribution with that of the overall DMOS values, and are also concentrate on small values, varying from about 0.5 to 7. In Fig. 8, two examples of DMOS distributions with 95% confidence interval are shown, which demonstrate the reliability of the subjective scores for approximating the visual quality of distorted images.

We also check the consistency of the subjects' judgements on the basis of SOS (Standard deviation of Opinion Scores)

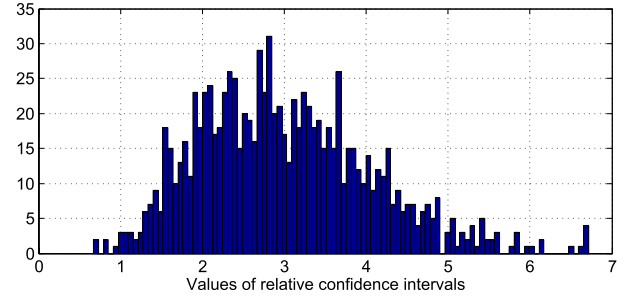


Fig. 7. Histogram of relative confidence intervals related to the overall DMOS values. The quality scale for all images is (0,100). Note that smaller values indicate higher reliability.

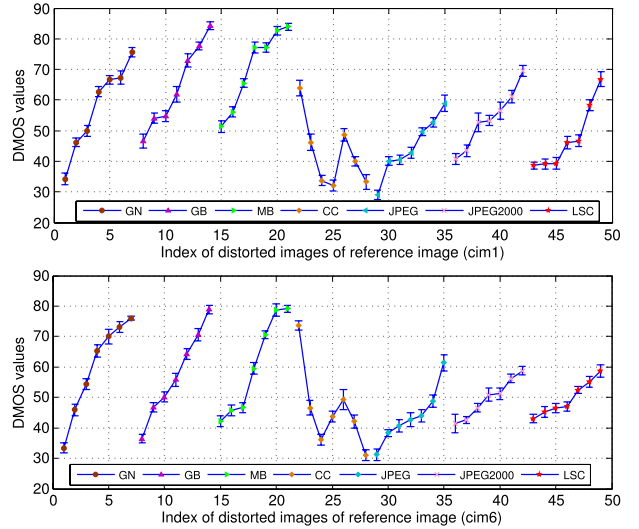


Fig. 8. Distributions of DMOS values of two examples. The error bars indicate the confidence intervals of related scores.

hypothesis [49], [50]. With the SOS hypothesis, the relationship between the SOS values $\{s\}$ and MOS values $\{x\}$ can be estimated by the formula $s(a)^2 = a(-x^2 + 10x)$ in the subjective rating with the discrete 11 point scales, where a is a parameter to estimate the relationship and represents the level of inter-subject agreement, and $a \in [0, 1]$. In Fig. 9, we provide the SOS hypothesis of our experimental results. The minimum difference between $\{s\}$ and the fitted $\{s(a)\}$ is obtained when $a = 0.054$, which indicates that the diversity of subjects' rating is small. The maximum diversity is achieved when a equals to 1, which is also illustrated in this figure.

D. Analysis of Different Regions

In the subjective test, we get three subjective scores for each test image: QE , QT and QP , corresponding to the quality of the entire, textual and pictorial regions, respectively. Based upon the subjective scores, one problem we would like to explore is which part contributes more to the overall visual quality of SCIs, textual or pictorial part? Hence, we analyze the overall correlation of these three quality scores (QE , QT and QP) in terms of *Pearson Linear Correlation Coefficient* (PLCC), *Root Mean Squared Error* (RMSE) and *Spearman rank-order correlation coefficient* (SROCC) [51].

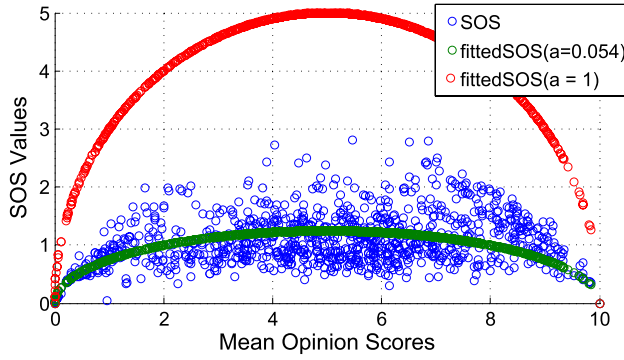


Fig. 9. SOS hypothesis for the subjective rating. Higher value of a indicates larger diversity of subjects' judgment in the subjective test.

TABLE I
CORRELATION ANALYSIS OF THE OBTAINED QUALITY SCORES FOR THE
ENTIRE IMAGES, TEXTUAL AND PICTORIAL REGIONS

Distortions	QE and QT			QE and QP		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
GN	0.9424	0.9367	4.9915	0.8958	0.8819	6.6295
GB	0.9268	0.9234	5.7006	0.8889	0.8916	6.9530
MB	0.9042	0.9057	5.5528	0.8513	0.8526	6.8218
CC	0.8332	0.7580	6.9558	0.8405	0.8030	6.8150
JPEG	0.8548	0.8488	4.8765	0.7493	0.7162	6.2226
JPEG2000	0.8474	0.8521	5.5185	0.8058	0.7821	6.1554
LSC	0.7701	0.7755	5.4432	0.6914	0.6923	6.1647
Overall	0.9040	0.8958	6.1204	0.8389	0.8336	7.7899

As such, we can know which part attracts more attention when viewing distorted SCIs. Through in-depth investigation of their correlation, an effective way for integrating textual and pictorial parts can be figured out. Meanwhile, correlations for each distortion type are also calculated to estimate human visual perception to different distortion types. The correlation results are reported in Table I.

To verify the statistical difference between these three sets of subjective scores, we perform the two-way *Analysis of Variance* (ANOVA) [52] with the distortion levels and the three sets of subjective scores (i.e., QE, QT and QP) as factors. Based on the computation of the F-statistic (F) and the degree of freedom (r), the probability (p) that indicates the probability that the null hypothesis can be rejected, where the null hypothesis is that the mean of the compared factors is the same. Generally, p equal to or lower than 0.05 is considered sufficient to suggest that the observed factors are significantly different. The results with $F = 67.66$, $r = 2$ and $p < 0.001$ indicate that there exists statistical difference among the three sets of subjective scores. Besides, a significant effect of distortion levels to the final quality is verified with $F = 187.55$, $r = 48$ and $p < 0.001$. The results ($F = 4.41$, $r = 96$ and $p < 0.001$) indicate that there exists interaction effects between these two quality factors. The final visual quality of SCIs is determined by both the distortion type and the specific region.

From Table I, we can observe that the textual part has higher overall correlation with the entire image than the

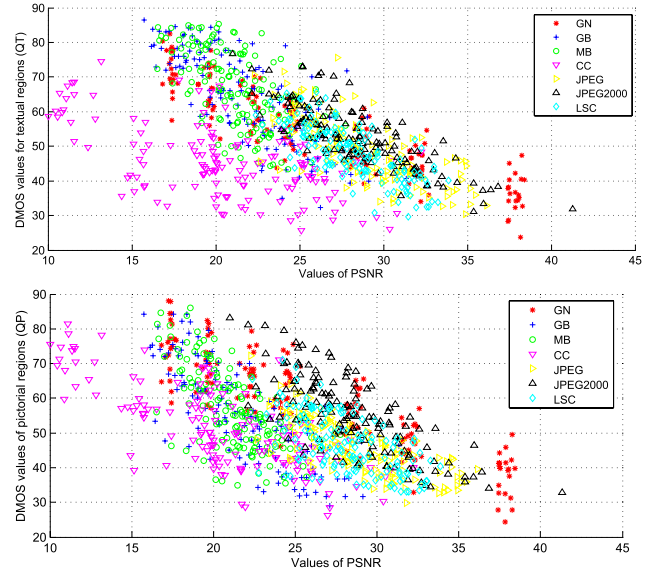


Fig. 10. Distributions of DMOS values of textual and pictorial regions versus PSNR values. PSNR is used here to measure the actual intensity variation.

pictorial part. However, for different distortion types, the results vary to some extent. For example, in the CC case, the contrast variation of pictorial regions affect human vision more compared to that of textual regions. The reason is that, observers prefer to give high scores to texts of high shape integrity and clarity, even though their colors change significantly. For pictorial regions, severe contrast change would result in uncomfortable viewing experiences. Therefore, in this case, pictorial regions contribute more to the quality of the entire image. On the contrary, in the MB case, textual regions attract more attention. The integrity and clarity of texts are easier to be affected by motion blurring. For other distortions, the correlation results also vary from case to case. These phenomenons can also be reflected from the distributions of DMOS values of these two regions. The distributions of textual and pictorial DMOS values are illustrated in Fig. 10. From the upper subfigure, we can see that subjects prefer to give high quality scores (low DMOS values) to contrast changed textual regions, while textual regions impaired by blurring have higher DMOS values. For pictorial regions, the difference between different distortion types is not so obvious. Consequently, it is challenging to have an unified formula to account for the correlation among the three scores. This analysis results can inspire researchers to propose effective objective metrics for distorted SCIs.

IV. OBJECTIVE QUALITY ASSESSMENT OF SCIS

As aforementioned in Sec. II, due to the different properties of textual and pictorial regions in SCIs, the same distortion in different regions may lead to different visual perception of human beings. Hence, it is natural and reasonable to separately handle each part, and then combine them together with differentiation. In this section, we propose a novel scheme (SPQA) to objectively evaluate the visual quality of distorted SCIs, considering the visual differences between textual and pictorial regions. The diagram of the proposed scheme is illustrated in Fig. 11. One reference SCI X and

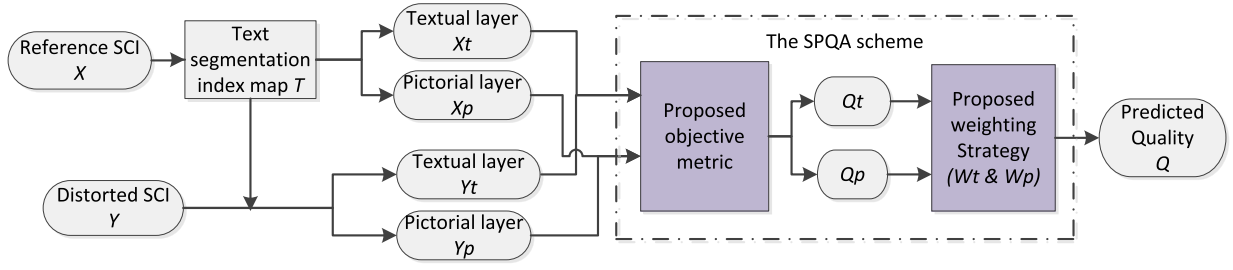


Fig. 11. Diagram of the proposed SPQA scheme. The SPQA mainly contains two algorithms highlighted in the figure.



Fig. 12. DMOS values of some examples in SIQAD. The scale of the DMOS values is from 0 to 100. A higher value represents worse visual quality of the image (refer to the images at the original resolution for better visual comparison). (a) Reference image: cim11. (b) cim11_3_5, DMOS:63.98. (c) cim11_4_7, DMOS:37.50. (d) cim11_4_1, DMOS:76.54.

its distorted version Y are firstly segmented into textual and pictorial layers based on their text segmentation index map T . The quality of the textual and pictorial layers is then separately evaluated by the proposed objective metric (to be introduced in Sec. IV-A). A novel weighting strategy, derived from the correlation analysis of subjective scores, is proposed in Sec. IV-B to integrate the two quality scores Q_t and Q_p to the final visible quality score Q of the distorted SCI.

A. Quality Evaluation of Textual and Pictorial Regions

It is known that the HVS is relevant to image luminance, contrast and sharpness. They change along with various image distortions, such as noise corruption, blur, quantization and compression artefacts. Hence, they have been widely investigated in the FR NIQA. In SSIM [24], the product of three components of similarity between the reference patch x and its distorted version y is computed to estimate the image local quality:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (4)$$

where $l(x, y)$, $c(x, y)$ and $s(x, y)$ are luminance, contrast and structural similarity; α , β and γ are positive constants used to adjust the relative importance of these three components. A simple setting ($\alpha = \beta = \gamma = 1$) is adopted in SSIM and most of its variations [24]. Liu et al. [27] used gradient similarity to replace the contrast/structural similarity in SSIM, and proposed a weighting strategy to combine the luminance and gradient similarity as follows:

$$q = (1 - W) \times g(x, y) + W \times e(x, y) \quad (5)$$

where q is the quality score of the distorted patch y ; $e(x, y)$ and $g(x, y)$ are luminance and gradient similarity.

$W = 0.1 \times g(x, y)$ is used as weighting value to highlight the contribution of the gradient similarity to the final quality. In [28], the authors found that, without any additional information, using the image gradient similarity alone can yield highly accurate quality prediction.

However, these interaction schemes of the properties cannot work well for SIQA, since the HVS perception to textual and pictorial regions are different. As illustrated in Sec. III-D, the distortions in textual regions are not always playing the same role to the overall quality. For example, subjects can easily notice luminance and contrast change in pictorial regions. However, they prefer to give high quality scores to texts with high integrity and clear shape, even though their color intensity or contrast has been greatly changed. Conversely, subjects are sensitive to blurring artifacts appearing on textual regions. As illustrated in Fig. 12, there is motion blur appearing on the image in (b) and color intensity change occurring on the image in (c). We can see that the background content and color intensity of texts in (c) are much different from the reference image in (a), while the background and contrast of texts in (b) are well maintained. However, subjective tests show that humans are more satisfied with (c) than (b), which can be reflected from their DMOS values: 63.98 for (b) and 37.50 for (c). Therefore, in these cases, we should reduce the effect of the luminance change to the overall quality of textual regions. However, with much luminance change, as displayed in Fig. 12 (d), subjects give low quality scores to this image at their first impression. Hence, for these cases, the effect of the luminance change in textual regions to the overall quality should be enhanced.

Based on the above analysis, we propose a new scheme for quality evaluation of distorted SCIs. In the proposed scheme,

0	0	0	0	0
1	3	8	3	1
0	0	0	0	0
-1	-3	-8	-3	-1
0	0	0	0	0

0	0	1	0	0
0	8	3	0	0
1	3	0	-3	-1
0	0	-3	-8	0
0	0	-1	0	0

0	0	1	0	0
0	0	3	8	0
-1	-3	0	3	1
0	-8	-3	0	0
0	0	-1	0	0

0	1	0	-1	0
0	3	0	-3	0
0	8	0	-8	0
0	3	0	-3	0
0	1	0	-1	0

Fig. 13. Filters for calculating the sharpness values.

sharpness and luminance similarity between the reference and distorted SCIs is computed. Sharpness is computed since it is a good measure to summarize various distortions appearing in images [28], [53]. The luminance similarity of textual regions is adaptively integrated to the sharpness similarity, while only sharpness similarity is considered for pictorial regions. For one SCI X and its distorted version Y , given its text segmentation index map T , their textual layers (X_t , Y_t) and pictorial layers (X_p , Y_p) are calculated by $X_t = X \cdot T$, $X_p = X \cdot (1 - T)$, $Y_t = Y \cdot T$ and $Y_p = Y \cdot (1 - T)$. The luminance similarity map $S_l(X_t, Y_t)$ between the textual layers X_t and Y_t is calculated as follows:

$$S_l(X_t, Y_t) = \frac{2 \cdot \mu_{x_t} \cdot \mu_{y_t} + c_1}{\mu_{x_t}^2 + \mu_{y_t}^2 + c_1} \quad (6)$$

where μ_{xt} and μ_{yt} denote the local mean values for each pixel in the textual layers X_t and Y_t . c_1 is a parameter to avoid the instability when the denominator is close to zero.

To compute the sharpness of images, we use the multi-directional filters $\{h_k\}_{k=1,2,3,4}$ illustrated in Fig. 13. These filters can capture the local variations of images at four directions, including horizontal and vertical directions. The sharpness of one image X is measured by the summary of the first two maximum filtering results:

$$s(X) = |X \cdot h_a| + |X \cdot h_b|; \quad (7)$$

where a and b are the index of the filter that lead to the first two maximum results; $|\cdot|$ represents the absolute value of the convolution of X and h_k . Thus, the sharpness similarity between X_t and Y_t , and X_p and Y_p , are computed as:

$$S_s^t(X_t, Y_t) = \frac{2 \cdot s(X_t) \cdot s(Y_t) + c_2}{s(X_t)^2 + s(Y_t)^2 + c_2} \quad (8)$$

$$S_s^p(X_p, Y_p) = \frac{2 \cdot s(X_p) \cdot s(Y_p) + c_2}{s(X_p)^2 + s(Y_p)^2 + c_2} \quad (9)$$

where c_2 is a parameter to avoid the instability when the denominator is close to zero.

The quality map for the pictorial part Q^p_map is measured by the sharpness similarity between pictorial regions.

$$Q^p_{map} = S^p_s(X_p, Y_p) \quad (10)$$

The quality map for the textual part Q^t_map can be calculated by integrating the luminance and sharpness similarity maps as follows.

$$Q^t_map = [S_l(X_t, Y_t)]^\alpha \cdot [S_s^t(X_t, Y_t)]^\beta \quad (11)$$

where $\alpha > 0$ and $\beta > 0$ are parameters used to adjust the effect of the two components. In this paper, we set $\beta = 1$ to simplify this definition, since the structural difference is

important to both textual and pictorial regions. α is used to adjust the effect of the luminance component when the textual layers are processed. As illustrated in Fig. 12, human beings are not sensitive to intensity change derived from some degree of quantization or contrast change, we calculate the difference between the textual layers to measure the degree of the intensity change. The difference is measured as follows:

$$d = (2 \cdot v_1 \cdot v_2)/(v_1^2 + v_2^2); \quad (12)$$

where $v_1 = \max(X_I) - \min(X_I)$ and $v_2 = \max(Y_I) - \min(Y_I)$. When the intensity change is small, the effect of the luminance similarity to the visual quality should be reduced; when the change is large, the effect of the luminance similarity should be enhanced. Hence, the value of α can be determined by d and the threshold δ as follows:

$$\alpha = \begin{cases} d & \text{if } d > \delta \\ 1/d & \text{if } d \leq \delta \end{cases} \quad (13)$$

B. Proposed Weighting Strategy

As aforementioned, it is challenging to establish an uniform formula to account for the interaction of the three regions. There are many factors affecting human perception when viewing SCIs, including area ratio and position of texts, size of characters, content of pictures, etc. As an initial attempt towards solving this problem, we initially investigate a statistical property of SCIs that reflects impairments of test images, rather than any specific factor. Here, image activity measure is adopted to calculate the weights. Image activity values reflect the variation of image content, which can be used to differentiate images [54], [55]. Based on the activity measure and the segmentation algorithm proposed in [41], we propose a novel model to compute two weights ($\mathcal{W}_t$ and $\mathcal{W}_p$) that can measure the effect of textual and pictorial regions to the quality of the entire image. In particular, given one reference SCI and its text segmentation index map T in which textual pixels are marked by one and pictorial pixels by zero, we calculate the activity map A of the corresponding distorted SCI [41]. The activity maps $A_t = A \times T$ and $A_p = A \times (1 - T)$ of the textual and pictorial regions can be calculated. Considering the human visual acuity in the HVS (the human eyes have high visual acuity to points closed to the fixation center, and the visual acuity decreases with the distance increase from the fixation point), a Gaussian mask G is used to weight the activity values. Based on the weighted activity map, two values $\mathcal{W}_t$ and $\mathcal{W}_p$ for the textual and pictorial parts are computed as Eq. (14) and (15), which are subsequently employed as weights to combine the two quality scores.

$$\mathcal{W}_t = \frac{\sum_{i=1}^m \sum_{j=1}^n (A \cdot T \cdot G)_{i,j}}{\sum_{i=1}^m \sum_{j=1}^n (T)_{i,j}} \quad (14)$$

and

$$\mathcal{W}_p = \frac{\sum_{i=1}^m \sum_{j=1}^n (A \cdot (1 - T) \cdot G)_{i,j}}{\sum_{i=1}^m \sum_{j=1}^n (1 - T)_{i,j}} \quad (15)$$

where m and n represent the dimensions of the images. The weighting maps for textual and pictorial parts of one SCI example are shown in Fig. 14.

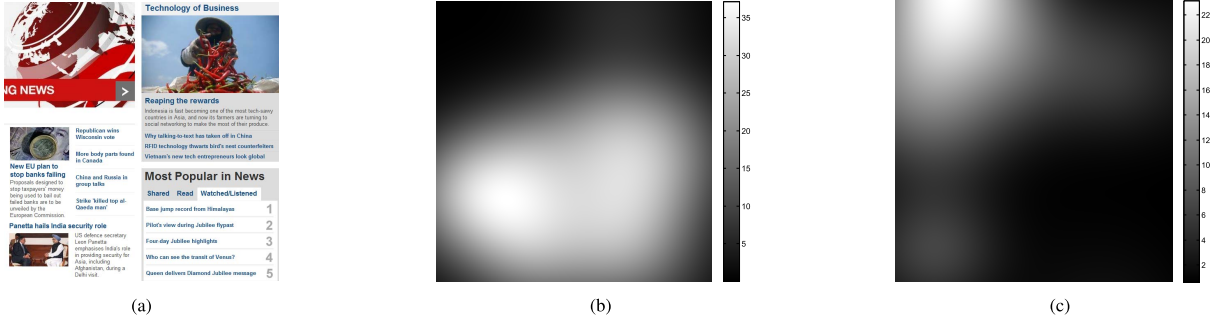


Fig. 14. Weighting maps for textual and pictorial regions of one SCI example. (a) Reference image: cim1. (b) Weighting map for textual regions. (c) Weighting map for pictorial regions.

Based on the calculated quality maps of textual layer Q^t_map and pictorial layer Q^p_map , the quality scores of the textual and pictorial regions are computed as the mean values of the corresponding regions:

$$Q_t = \frac{Q^t_map \cdot T}{\sum_{i=1}^m \sum_{j=1}^n (T)} \quad (16)$$

$$Q_p = \frac{Q^p_map \cdot (1 - T)}{\sum_{i=1}^m \sum_{j=1}^n (1 - T)} \quad (17)$$

Following the same notation above, m and n denote the dimension of the reference SCI. The final quality score Q of the distorted image Y is computed as follows:

$$Q = \mathcal{W}_t * Q_t + \mathcal{W}_p * Q_p \quad (18)$$

V. EXPERIMENTAL RESULTS

In this section, we first test the validity of the proposed weighting strategy, by applying the weighting strategy to subjective scores and some existing NIQA methods. We then investigate the effectiveness of the proposed SPQA scheme to assess the quality of SCIs in the *SIQAD*.

A. Analysis of the Proposed Weighting Strategy

1) *Applying the Weighting Strategy to Subjective Data:* Since we obtain the three sets of subjective scores for entire, textual and pictorial regions in SCIs, it is reasonable to verify the proposed weighting strategy on the basis of subjective scores. A quality score QE' of an entire SCI is predicted based on the quality scores of textual and pictorial regions, i.e., QT and QP . The QE' is computed as follows.

$$QE' = \mathcal{W}_t * QT + \mathcal{W}_p * QP \quad (19)$$

where $\mathcal{W}_t$ and $\mathcal{W}_p$ are computed as introduced in Sec. IV-B. The performance of the combination can be measured by computing the correlation between QE' and ground truth score QE . Meanwhile, we compare the proposed model with a simple averaging combination of textual and pictorial scores. In the averaging combination, the predicted quality scores QE_a is the mean of quality scores of textual and pictorial regions:

$$QE_a = 0.5 * QT + 0.5 * QP \quad (20)$$

TABLE II

COMPARISON OF TWO COMBINATION METHODS. THE PAIRED T-TEST IS APPLIED TO THE QE' AND QE_a . THE RESULT ($H = 1$, $P < 0.05$) INDICATES THAT THE QUALITY SCORES GENERATED BY THE TWO METHODS ARE STATISTICALLY DIFFERENT

	QE and QEa			QE and QE'		
Distortions	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
GN	0.9085	0.8968	6.2332	0.9436	0.9364	4.9392
GB	0.9047	0.9083	6.4672	0.9422	0.9429	5.0856
MB	0.8735	0.8755	6.3298	0.9277	0.9258	4.8551
CC	0.8522	0.8129	6.5830	0.8850	0.8321	5.8575
JPEG	0.7750	0.7473	5.9378	0.8436	0.8279	5.0458
JPEG2000	0.8171	0.7964	5.9914	0.8528	0.8441	5.4284
LSC	0.7128	0.7126	5.9845	0.7945	0.7900	5.4973
Overall	0.8631	0.8579	7.2297	0.9230	0.9164	5.5088

TABLE III

COMPARISON OF TWO COMBINATION METHODS. MORE DETAILED RESULTS OF THE NIQA METHODS ARE REPORTED IN SEC. V-B

	SSIM vs. W_SSIM		IFC vs. W_IFC		VIF vs. W_VIF	
PLCC	0.7561	0.7974	0.6395	0.5342	0.8198	0.7886
SROCC	0.7566	0.7883	0.6011	0.5224	0.8065	0.7807
RMSE	9.3676	8.5725	11.0048	11.9946	8.1969	8.8885
	FSIM vs. W_FSIM		GMSD vs. W_GMSD		SPQA	
PLCC	0.5389	0.5617	0.7387	0.7642	0.8584	
SROCC	0.5279	0.5582	0.7305	0.7692	0.8416	
RMSE	12.0583	12.2928	9.6484	9.2062	7.3421	

Table II reports the comparison results. It shows that the results with the proposed weighting strategy are more consistent with human visual perception. Although there is still room to improve the performance, the proposed weighting strategy reflects the contributions of textual and pictorial regions with a high reliability. We also checked the performance of area-ratio based weighting method, that is to say, the area ratio of textual region (pictorial region) is used to replace the $\mathcal{W}_t$ ($\mathcal{W}_p$). Since the area ratios of textual regions in the *SIQAD* just vary from 35% to 60%, the correlation result of the area-ratio weighting method is similar to the result of average combination.

2) *Applying the Weighting Strategy to Some Existing NIQA Metrics:* In this section, we apply the weighting strategy to some representative NIQA metrics, such as SSIM [24], VIF [25], IFC [56], FSIM [26] and GMSD [28]. In particular, we firstly separate SCIs into textual and pictorial layers,

TABLE IV

CORRELATION RESULTS OF THE DMOS VALUES AND THE OBJECTIVE SCORES GIVEN BY 12 METRICS. THE PAIRED T-TEST IS APPLIED TO THE PROPOSED SPQA AGAINST THE 11 NIQA METHODS. THE RESULTS ($H = 1$, $P < 0.05$) FOR EACH PAIR INDICATES THAT THE SPQA IS SIGNIFICANTLY BETTER THAN THE TESTED 11 NIQA METHODS

	Distortions	PSNR	SSIM	MSSIM	IWSSIM	VIF	IFC	VSNR	MAD	FSIM	GSIM	GMSD	SPQA
PLCC	GN	0.9053	0.8806	0.8783	0.8804	0.9011	0.8791	0.8840	0.8852	0.7428	0.8448	0.8956	0.8921
	GB	0.8603	0.9014	0.8984	0.9079	0.9102	0.9061	0.8890	0.9120	0.7206	0.8831	0.9094	0.9058
	MB	0.7044	0.8060	0.8240	0.8414	0.8490	0.6782	0.7829	0.8361	0.6874	0.7711	0.8436	0.8315
	CC	0.7401	0.7435	0.8371	0.8404	0.7076	0.6870	0.7667	0.3933	0.7507	0.8077	0.7827	0.7992
	JPEG	0.7545	0.7487	0.7756	0.7998	0.7986	0.7606	0.7972	0.7662	0.5566	0.6778	0.7746	0.7696
	J2K	0.7893	0.7749	0.7951	0.8040	0.8205	0.7963	0.8170	0.8344	0.6675	0.7242	0.8509	0.8252
	LSC	0.7805	0.7307	0.7729	0.8155	0.8385	0.7679	0.7982	0.8184	0.5964	0.7218	0.8559	0.7958
	Overall	0.5869	0.7561	0.6161	0.6527	0.8198	0.6395	0.5982	0.6191	0.5389	0.5663	0.7387	0.8584
SROCC	GN	0.8790	0.8694	0.8679	0.8743	0.8888	0.8717	0.8662	0.8721	0.7373	0.8404	0.8856	0.8823
	GB	0.8573	0.8921	0.8883	0.9060	0.9059	0.9106	0.8827	0.9087	0.7286	0.8796	0.9119	0.9017
	MB	0.7130	0.8041	0.8238	0.8421	0.8492	0.6737	0.7799	0.8357	0.6641	0.7753	0.8441	0.8255
	CC	0.6828	0.6405	0.7506	0.7563	0.6433	0.6396	0.6694	0.3907	0.7175	0.7148	0.6378	0.6154
	JPEG	0.7569	0.7576	0.7787	0.7978	0.7924	0.7636	0.8084	0.7674	0.5879	0.6796	0.7712	0.7673
	J2K	0.7746	0.7603	0.7855	0.7998	0.8131	0.7980	0.8112	0.8382	0.6363	0.7125	0.8436	0.8152
	LSC	0.7930	0.7371	0.7711	0.8214	0.8463	0.7713	0.8088	0.8154	0.5979	0.7145	0.8592	0.8003
	Overall	0.5604	0.7566	0.6115	0.6545	0.8065	0.6011	0.5743	0.6067	0.5279	0.5551	0.7305	0.8416
RMSE	GN	6.3372	7.0679	7.1309	7.0744	6.4673	7.1096	6.9721	6.9391	9.9860	7.9811	6.6354	6.7394
	GB	7.7376	6.5701	6.6638	6.3619	6.2859	6.4193	6.9506	6.2269	10.5230	7.1210	6.3111	6.4301
	MB	9.2287	7.6967	7.3675	7.0260	6.8704	9.5544	8.0897	7.1322	9.4432	8.2788	6.9816	7.2223
	CC	8.4591	8.4116	6.8818	6.8184	8.8876	9.1407	8.0760	11.5652	8.3190	7.4160	7.8294	7.6184
	JPEG	6.1665	6.2295	5.9311	5.6404	5.6551	6.1004	5.6726	6.0380	7.8072	6.9085	5.9427	6.0000
	J2K	6.3819	6.5691	6.3040	6.1804	5.9412	6.2875	5.9929	5.7276	7.7404	7.1675	5.4591	5.8706
	LSC	5.3336	5.8253	5.4141	4.9379	4.6497	5.4657	5.1429	4.9025	6.8486	5.9046	4.4121	5.1664
	Overall	11.5898	9.3676	11.2744	10.8444	8.1969	11.0048	11.4706	11.2409	12.0583	11.7980	9.6484	7.3421

and then substitute the objective evaluation part in the SPQA scheme with the NIQA metrics. The quality of textual and pictorial layers is evaluated by the NIQA metrics separately, and then is combined to estimate the final overall quality via the proposed weighting strategy. The modified NIQA methods are marked by weighted metrics, e.g., W_SSIM , W_FSIM , W_GMSD , W_IFC and W_VIF . The correlation between overall DMOS and predicted scores by the modified NIQA metrics is computed and reported in Table III. From this table, we can see that the performance of some modified NIQA metrics are improved when the proposed weighting strategy is integrated, such as W_SSIM , W_FSIM and W_GMSD . However, the improvement is still far away from satisfaction in evaluating the visual quality of distorted SCIs. As to the W_IFC and W_VIF , the performance drops somehow. Therefore, new objective metrics specific for SCI quality assessment is desired, and the proposed SPQA at some extent has filled this requirement. Overall, the proposed SPQA with the weighting strategy works much better than other relevant existing objective metrics.

B. Performance of the Proposed SPQA on the SIQAD

In this section, we use the images in the SIQAD to conduct the comparison experiments by using the proposed SPQA

and other existing ones. The following 11 state-of-the-art NIQA metrics are adopted: PSNR, SSIM [24], MSSIM [57], IWSSIM [58], VIF [25], IFC [56], VSNR [59], MAD [21], FSIM [26], GSIM [27] and GMSD [28]. These metrics are implemented using the codes on their websites. We apply all the metrics to the grayscale version of images, and compute the correlations between the predicted scores and DMOS values in terms of PLCC, RMSE and SROCC. Meanwhile, the correlations of specific distortions are calculated, to investigate the effectiveness of objective methods for different types of distortions. We set $c_1 = 0.0026$, $c_2 = 0.0062$, and δ is experimentally set to 0.95 in the experiments.

We report the correlation results in Table IV, where the first two with the best performance are marked with the **bold** font. It is shown that the proposed SPQA achieves the highest overall correlation with DMOS values. Correlations between the SPQA scores and DMOS values for different distortion types are distinct from each other, as most of the other metrics. Particularly, there are much higher values for the first three distortions (i.e., GN, GB and MB) than others. The reason is that observers are sensitive to such kinds of distortions allocated in the entire image, and are able to distinguish the images with different distortion levels. For the remaining four types, especially for the CC case, the correlation results are not so high. The reason is that the contrast change only affects

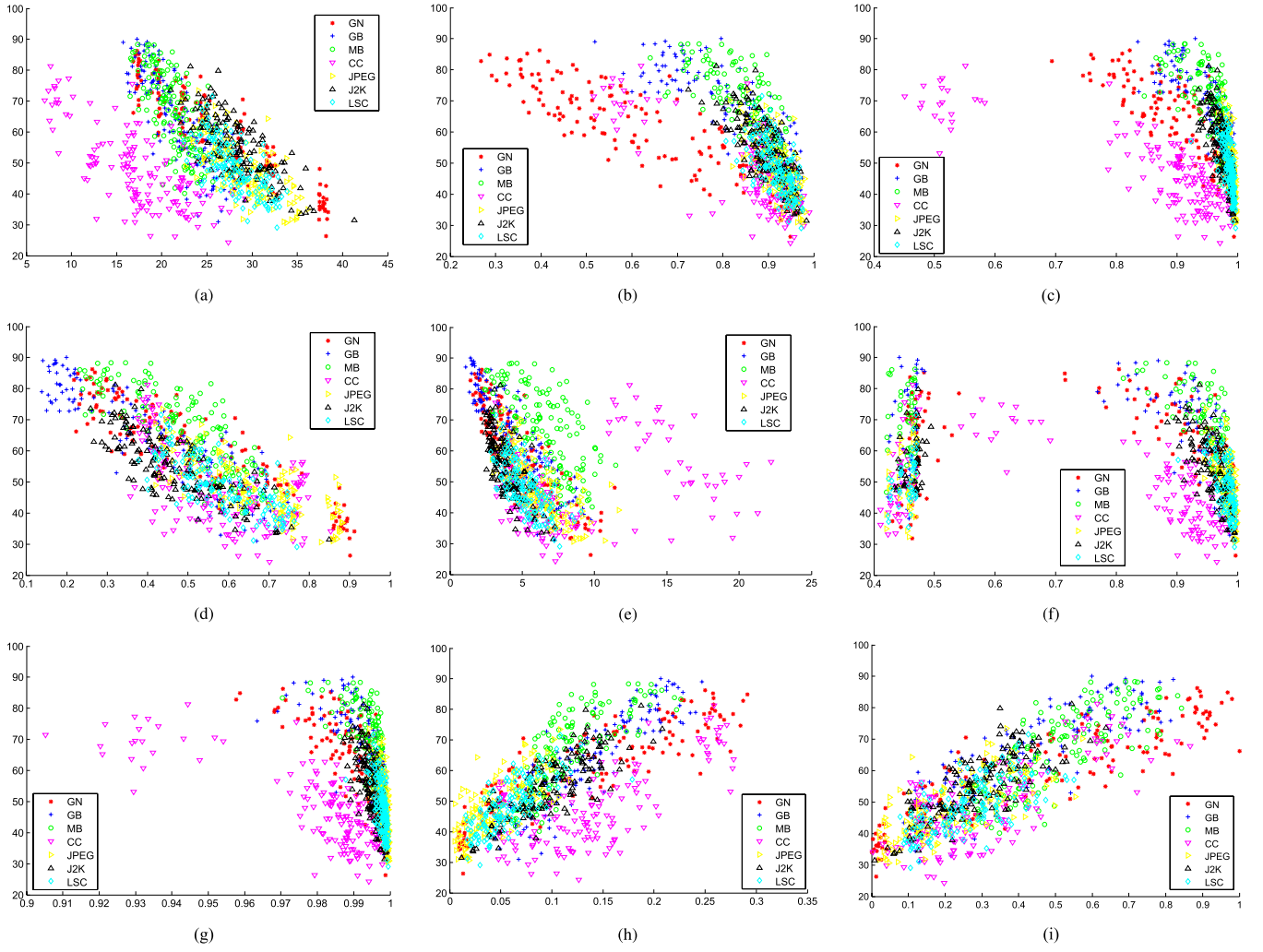


Fig. 15. Scatter of predicted quality scores by some metrics against the DMOS values on the SIQAD. The vertical axis in each figure is the DMOS values. (a) PSNR. (b) SSIM. (c) MSSIM. (d) VIF. (e) IFC. (f) FSIM. (g) GSIM. (h) GMSD. (i) SPQA.

the intensity of texts, but not the integrity of texts about which subjects care more. By contrast, the NIQA metrics take the intensity variation into account, resulting in the inconsistency with DMOS values. Taking the VIF for instance, it performs well for quality evaluation of SCIs with some distortion types, such as GN, blurring and compression artifacts. That is because the influence of such distortion type to textual and pictorial regions is similar. In other word, the visual information loss of these two regions increases along with the increase of degradation level. However, for the CC case, the visual information loss of textual regions computed by the VIF does not change highly consistently with the change of the degradation level. The proposed SPQA has taken this situation into account, and thus the predicted quality of all the test images has higher consistency with human perception compared with other existing metrics.

In addition, we test the performance of the parameter setting of α in the proposed SPQA. We combine the luminance and sharpness similarity simply by $\alpha = \beta = 1$, and mark this method as *SPQA_S*. The correlation results are as follows: $PLCC = 0.8243$, $SROCC = 0.8029$, $RMSE = 8.0254$,

from which we can see that the performance of *SPQA_S* is not as good as that of SPQA without the adaptive adjustment of α . Although the adjustment might be over-estimated for some cases (e.g., GN and JPEG), resulting in the performance drop for the single distortion type, the overall visual quality of images from different distortion types will be highly consistent with human visual perception.

In Fig. 15, we also provide the scatter plots of the predicted quality scores against the DMOS values for some representative objective metrics (such as PSNR, SSIM, MSSIM, VIF, IFC, FSIM, GSIM, GMSD and SPQA) on the SIQAD. The seven kinds of distortions (GN, GB, MB, CC, JPEG, JPEG2000 and LSC) are separately displayed with different markers. From Fig. 15, it can be observed that the predicted scores by the SPQA have the most centralized distribution than others. In most of other metrics, the distribution of predicted scores on all distortion types is somehow dispersive. For example, for PSNR and GSIM, the distribution of predicted scores on the CC distortion deviates much from the distribution on other kinds of distortions, degrading their overall performance.



Fig. 16. Visual quality comparison of SCIs with different distortion types. The DMOS values and the quality scores predicted by four different metrics (PSNR, SSIM, VIF and SPQA) are provided for comparison. (a) Reference image (cropped from ‘cim13’ in *SIQAD*). (b) Image with CC, DMOS: 40.2294, PSNR: 20.5616, SSIM: 0.8595, VIF: 0.5850, SPQA: 0.2546. (c) Image with GB, DMOS: 48.7758, PSNR: 22.6598, SSIM: 0.9054, VIF: 0.5291, SPQA: 0.3037. (d) Image encoded by JPEG, DMOS: 51.2387, PSNR: 24.6442, SSIM: 0.8653, VIF: 0.4599, SPQA: 0.3253. (e) Image with GN, DMOS: 65.8586, PSNR: 24.4163, SSIM: 0.6302, VIF: 0.4900, SPQA: 0.4736. (f) Image with MB, DMOS: 79.8107, PSNR: 19.7835, SSIM: 0.8341, VIF: 0.4804, SPQA: 0.5488.

In Fig. 16, a reference SCI (a) and its several distorted versions (b)-(f) are given for visual quality comparison. We can see that, from (b) to (f), the DMOS values of these images increase, indicating the descending of the visual quality. However, the three measures (PSNR, SSIM and VIF) do not have the same changing tendency, and this means that they cannot achieve high consistency with the DMOS values in these cases. These three metrics generally capture the practical variations occurring in the distorted images, without considering the different perception of viewers to different regions in SCIs. For instance, in the subjective test, observers prefer to give high scores to images with clear and unbroken textual regions, even though their intensity values have been changed. Compared with images in (c) and (d), the image (b) is with the highest visual quality. However, PSNR and SSIM values of (c) and (d) are higher than those of (b). Additionally, most subjects have a bad impression on the blurring effect at the first sight, and thus give low scores to the blurred images. As shown in Fig. 16 (d)-(f), the images in (d) and (e) have better visual quality than image (f) with severe motion blur. However SSIM value of (e) and VIF value of (d) are lower. This phenomenon can also be observed in Fig. 8, where most of the DMOS values for blurred images (from the first eight to the twenty-one points) are higher than other images.

C. More Analysis on the SPQA Metric

When the proposed SPQA algorithm is used to predict the visual quality scores, other doubts may raise: for example, does the predicted textual score Q_t have high correlation with

TABLE V
CORRELATION RESULTS BETWEEN SUBJECTIVE (QE, QT AND QP) AND PREDICTED QUALITY SCORES (Q, Qt AND Qp), RESPONDING TO ENTIRE, TEXTUAL AND PICTORIAL REGIONS IN SCIS

subjective & objective scores	PLCC	SROCC	RMSE
QT & Qt	0.8493	0.8459	7.0018
QP & Qp	0.8164	0.8075	8.0012
QE & Qt	0.8377	0.8257	7.8157
QE & Qp	0.7527	0.7364	8.4356
QE & Q_a	0.7969	0.7837	8.6467
QE & Q	0.8584	0.8416	7.3421

the subjective textual score QT ? How about the performance if either W_t or W_p in the weighting strategy is set to zero? In order to answer these questions, we check the correlations between the subjective scores (QE, QT and QP) and predicted objective scores (Q, Qt and Qp), for example, QT and Qt , QP and Qp , QE and Qt , QE and Qp . The correlation results are given in Table V. From this table, we can find that although the predicted textual score Qt has relatively higher correlation with the ground truth textual scores QT , the result (i.e., correlation between QE and Qt) drops if just using the textual scores to estimate the overall quality scores. This also occurs when the pictorial scores are used alone to predict the overall quality scores. We also apply the average combination to the objective scores Qt and Qp , and the obtained overall quality scores are marked as Q_a in Table V.

We can find that, the objective quality scores Q computed via the proposed SPQA with the weighting strategy achieve the highest correlation results with the subjective scores.

VI. CONCLUSION

In this paper, we have carried out an in-depth study on perceptual quality assessment of distorted SCIs, from both subjective and objective perspectives. The first large-scale image database, *SIQAD*, is built to explore the subjective quality evaluation of SCIs. DMOS values of images in the database are obtained via the subjective test, and their reliability is verified. The built *SIQAD* is expected to facilitate further research in SCIs. Based upon the three subjective scores for textual, pictorial and entire regions, we find that textual regions contribute more to the quality of the entire image in most distortion cases. The proposed weighting strategy works well to account for this relationship. Combined with the weighting strategy, a new objective quality metric is constructed to separately assess the visual quality of textual and pictorial regions. The proposed integration scheme, named SPQA, outperforms existing 11 NIQA objective metrics on visual quality evaluation of distorted SCIs, as demonstrated by the experimental results.

REFERENCES

- [1] H. Shen, Y. Lu, F. Wu, and S. Li, "A high-performance remote computing platform," in *Proc. IEEE PerCom*, Mar. 2009, pp. 1–6.
- [2] T.-H. Chang and Y. Li, "Deep shot: A framework for migrating tasks across devices using mobile phone cameras," in *Proc. ACM CHI*, 2011, pp. 2163–2172.
- [3] Y. Lu, S. Li, and H. Shen, "Virtualized screen: A third element for cloud—Mobile convergence," *IEEE MultiMedia*, vol. 18, no. 2, pp. 4–11, Feb. 2011.
- [4] T. Lin and P. Hao, "Compound image compression for real-time computer screen image transmission," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 993–1005, Aug. 2005.
- [5] C. Lan, G. Shi, and F. Wu, "Compress compound images in H.264/MPGE-4 AVC by exploiting spatial correlation," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 946–957, Apr. 2010.
- [6] H. Yang, W. Lin, and C. Deng, "Learning based screen image compression," in *Proc. IEEE MMSP*, Sep. 2012, pp. 77–82.
- [7] Z. Pan, H. Shen, Y. Lu, S. Li, and N. Yu, "A low-complexity screen compression scheme for interactive screen sharing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 949–960, Jun. 2013.
- [8] Z. Pan, H. Shen, Y. Lu, and S. Li, "Browser-friendly hybrid codec for compound image compression," in *Proc. IEEE ISCAS*, May 2011, pp. 101–104.
- [9] H. Yang, S. Wu, C. Deng, and W. Lin, "Scale and orientation invariant text segmentation for born-digital compound images," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 533–547, Mar. 2015.
- [10] *Requirements for an Extension of HEVC for Coding of Screen Content*, document ISO/IEC JTC1/SC29/WG11 MPEG2013/N14174, 2014.
- [11] 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.
- [12] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, May 2011.
- [13] T. Ciszowski, W. Mazurczyk, Z. Kotulski, T. Hoßfeld, M. Fiedler, and D. Collange, "Towards quality of experience-based reputation models for future Web service provisioning," *Telecommun. Syst.*, vol. 51, no. 4, pp. 283–295, Dec. 2012.
- [14] R. Schatz and S. Egger, "An annotated dataset for Web browsing QOE," in *Proc. 6th Int. Workshop Quality Multimedia Exper.*, Sep. 2014, pp. 61–62.
- [15] D. Guse, S. Egger, A. Raake, and S. Möller, "Web-QOE under real-world distractions: Two test cases," in *Proc. 6th Int. Workshop Quality Multimedia Exper.*, Sep. 2014, pp. 220–225.
- [16] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, pp. 1–53, 2013.
- [17] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: What does the future hold?" *Multimedia Tools Appl.*, vol. 51, no. 2, pp. 675–696, Jan. 2011.
- [18] 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.
- [19] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 10, pp. 30–45, 2009.
- [20] P. L. Callet and F. Autrusseau. (2005). *Subjective Quality Assessment IRCCyN/IVC Database*. [Online]. Available: <http://www.irccyn.ec-nantes.fr/fr/ivc/db/>
- [21] 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.
- [22] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 616–625, Oct. 2012.
- [23] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-11, 2012.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.
- [30] 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.
- [31] P. Ye and D. Doermann, "Document image quality assessment: A brief survey," in *Proc. Int. Conf. Document Anal. Recognit.*, Aug. 2013, pp. 723–727.
- [32] I. Guyon, R. M. Haralick, J. J. Hull, and I. T. Phillips, "Data sets for OCR and document image understanding research," in *Handbook of Character Recognition and Document Image Analysis*, H. Bunke and P. Wang, Eds. Singapore: World Scientific, 1997, pp. 779–799.
- [33] D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, and J. Heard, "Building a test collection for complex document information processing," in *Proc. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2006, pp. 665–666.
- [34] H. Hase, "Quality evaluation of character image database and its application," in *Proc. Int. Conf. Document Anal. Recognit.*, Sep. 2011, pp. 1414–1418.
- [35] J. Kumar, P. Ye, and D. Doermann. (2013). *DIQA: Document Image Quality Assessment Datasets*. [Online]. Available: <http://lampsrv02.umiacs.umd.edu/projdb/project.php?id=73>
- [36] P. Ye and D. Doermann, "Learning features for predicting OCR accuracy," in *Proc. Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 3204–3207.
- [37] D. Kumar and A. G. Ramakrishnan, "QUAD: Quality assessment of documents," in *Proc. Int. Workshop Camera-Based Document Anal. Recognit.*, 2011, pp. 79–84.
- [38] T. Obafemi-Ajayi and G. Agam, "Character-based automated human perception quality assessment in document images," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 3, pp. 584–595, May 2012.
- [39] D. Karatzas, S. R. Mestre, J. Mas, F. Nourbakhsh, and P. P. Roy, "ICDAR 2011 robust reading competition—Challenge 1: Reading text in born-digital images (Web and email)," in *Proc. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2011, pp. 1485–1490.
- [40] 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.

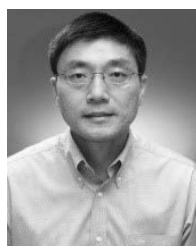
- [41] H. Yang, W. Lin, and C. Deng, "Image activity measure (IAM) for screen image segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2012, pp. 1569–1572.
- [42] H. Yang, Y. Fang, W. Lin, and Z. Wang, "Subjective quality assessment of screen content images," in *Proc. Int. Workshop Quality Multimedia Exper. (QoMEX)*, Sep. 2014, pp. 257–262.
- [43] SIQAD. [Online]. Available: <https://sites.google.com/site/subjectiveqa/>, accessed Aug. 2015.
- [44] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document ITU-T P.910, 2008.
- [45] Q. Huynh-Thu, F. Speranza, P. Corriveau, and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 1–14, Mar. 2011.
- [46] H. de Ridder, "Cognitive issues in image quality measurement," *J. Electron. Imag.*, vol. 10, no. 1, pp. 47–55, 2001.
- [47] Y. Pitrey, U. Engelke, M. Barkowsky, P. L. Callet, and R. P  pion, "Aligning subjective tests using a low cost common set," in *Proc. Euro ITV*, Lisbonne, Portugal, Jun. 2011.
- [48] K. Seshadrinathan, N. 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.
- [49] T. Hobfeld, R. Schatz, and S. Egger, "SOS: The MOS is not enough!" in *Proc. Int. Workshop Quality Multimedia Exper.*, Sep. 2011, pp. 131–136.
- [50] E. Siahhan, J. A. Redi, and A. Hanjalic, "Beauty is in the scale of the beholder: Comparison of methodologies for the subjective assessment of image aesthetic appeal," in *Proc. Int. Workshop Quality Multimedia Exper.*, Sep. 2014, pp. 245–250.
- [51] *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>, accessed Aug. 2015.
- [52] A. Gelman, "Analysis of variance—Why it is more important than ever," *Ann. Statist.*, vol. 33, no. 1, pp. 1–53, 2005.
- [53] 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.
- [54] L. Li and Z.-S. Wang, "Compression quality prediction model for JPEG2000," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 384–398, Feb. 2010.
- [55] Y.-H. Lee, J.-F. Yang, and J.-F. Huang, "Perceptual activity measures computed from blocks in the transform domain," *Signal Process.*, vol. 82, no. 4, pp. 693–707, Apr. 2002.
- [56] 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.
- [57] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.
- [58] 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.
- [59] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.



Huan Yang received the B.S. degree in computer science from the Heilongjiang Institute of Technology, China, in 2007, the M.S. degree in computer science from Shandong University, China, in 2010, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore. Her research interests include image/video processing and analysis, perception-based modeling and quality assessment, object detection/recognition, and computer vision.



Yuming Fang received the B.E. degree from Sichuan University, and the M.S. degree from the Beijing University of Technology, China. He was a (Visiting) Post-Doctoral Research Fellow with the IRCCyN Laboratory, PolyTech Nantes and University of Nantes, Nantes, France, University of Waterloo, Waterloo, Canada, and Nanyang Technological University, Singapore. He is currently an Associate Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, and 3D image/video processing. He was a Secretary of HHME2013 at the Ninth Joint Conference on Harmonious Human Machine Environment. He was also a Special Session Organizer in VCIP 2013 and the International Workshop on Quality of Multimedia Experience 2014.



Weisi Lin (SM'98) 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 more than 340 refereed papers published in international journals and conferences. He has been on the Editorial Board of the *IEEE TRANSACTIONS ON IMAGE PROCESSING*, the *IEEE TRANSACTIONS ON MULTIMEDIA* (2011–2013), the *IEEE SIGNAL PROCESSING LETTERS*, and the *Journal of Visual Communication and Image Representation*. He has been elected as an APSIPA Distinguished Lecturer (2012/13). He served as a Technical-Program Chair for 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. He is a fellow of Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.