Objective Quality Assessment of Screen Content Images by Uncertainty Weighting

Yuming Fang, *Member, IEEE*, Jiebin Yan, Jiaying Liu, *Member, IEEE*, Shiqi Wang, Qiaohong Li, and Zongming Guo, *Member, IEEE*

***Abstract*— In this paper, we propose a novel full-reference objective quality assessment metric for screen content images (SCIs) by structure features and uncertainty weighting (SFUW). The input SCI is first divided into textual and pictorial regions. The visual quality of textual regions is estimated based on perceptual structural similarity, where the gradient information is adopted as the structural feature. To predict the visual quality of pictorial regions in SCIs, we extract the structural features and luminance features for similarity computation between the reference and distorted pictorial patches. To obtain the final visual quality of SCI, we design an uncertainty weighting method by perceptual theories to fuse the visual quality of textual and pictorial regions effectively. Experimental results show that the proposed SFUW can obtain better performance of visual quality prediction for SCIs than other existing ones.**

***Index Terms*— Visual quality assessment, screen content image, full-reference quality assessment, uncertainty weighting.**

I. Introduction

With the increasing requirement of the transmission for complicated screen interfaces among clients, there is one type of images emerging over Internet, which is called screen content image (SCI). Generally, the SCI is a mixture of pictorial and computer generated textual/graphical regions. It has been widely used in various multimedia applications, including information sharing system between computer and smart devices [1], cloud computing systems [2], [3], remote conference, product advertising, *etc.* With the popularity of smart phones, more and more users would like to share different information with each other by rendering various visual content as the form of SCI, where various multimedia processing methods might be involved, such as coding [10], [11], [12], [8], transmission [3], *etc*. Currently, there are a large number of image processing algorithms proposed for SCI, including Just Noticeable Difference estimation (JND) for SCIs [54], SCI compression [4], [6], [7], SCI quality assessment [5], [17], [24], [52], [53], SCI segmentation [9], *etc*.

During the procedure of SCI acquisition, processing, transmission, *etc*., various distortions might be involved. When SCI is created by the camera in smart phones, the noise and blurring distortions might be generated due to the camera motion and different environments. For the transmission of SCI over Internet, the compression distortion might be created due to the image coding for efficient transmission. Thus, the visual quality assessment method, which can be used to estimate the perceptual quality of SCI, is much desired to serve as a benchmark for various SCI based systems.

In the past decades, there have been various image quality assessment (IQA) methods designed for visual content. According to available reference information, visual quality assessment (VQA) metrics can be classified into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR) approaches [21], [22]. Among these IQA metrics, FR metrics require the complete reference information to predict visual quality of images; RR metrics need part of reference information to estimate visual quality of images; NR metrics do not require any reference information for visual quality assessment of images. For all these three types of IQA methods, there have been various studies designing related metrics for IQA [21].

Most of traditional IQA metrics are FR approaches. Traditional signal fidelity methods such as Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE) and Mean Absolute Error (MAE) predict visual quality of images by simply computing pixel differences between the reference and distorted images. These signal fidelity methods are widely used for VQA in both industry and academia due to their simple and efficient implementation. However, they do not consider the properties of the Human Visual System (HVS), and thus, they might not obtain accurate quality prediction results as human beings perceive [21], [22]. To overcome the drawbacks of these existing metrics, many advanced perceptual IQA metrics have been proposed for various multimedia processing applications during the past decade [21].

Wang *et al.* [23] proposed the well-known FR metric of structural similarity (SSIM) by considering the characteristics of human beings' perception on image structure. Following this perceptual VQA metric, there are various types of FR VQA metrics proposed in recent ten years [18], [21].

Manuscript received August 24, 2016; revised December 29, 2016 and February 8, 2017; accepted February 8, 2017. Date of publication February 14, 2017; date of current version March 9, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61571212 and Grant 61472011 and in part by the Natural Science Foundation of Jiangxi under Grant 20161ACB21014. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stefan Winkler. *(Corresponding author: Jiaying Liu.)*

Y. Fang and J. Yan are with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China (e-mail: fa0001ng@e.ntu.edu.sg).

J. Liu and Z. Guo are with the Institute of Computer Science and Technology, Peking University, Beijing 100080, China (e-mail: liujiaying@pku.edu.cn; guozongming@pku.edu.cn).

S. Wang is with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore 637553 (e-mail: wangshiqi@ntu.edu.sg).

Q. Li is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 637553 (e-mail: qli013@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.2017.2669840

Sheikh *et al.* [32] used natural scene statistics to design a FR metric of information fidelity criterion (IFC) for images. The visual information fidelity (VIF) is designed by considering additive noise and blur distortion in [26], where most distortion types can be composed of these two distortions. Chandler *et al.* proposed a FR metric of visual signal-to-noise ratio (VSNR) by wavelet coefficients [35]. Larson and Chandler [34] proposed a FR metric called most apparent distortion (MAD) by local luminance and contrast masking, and local statistical features of images. Zhang *et al.* [25] adopted phase congruency and gradient magnitude to design a FR metric called feature similarity (FSIM). Gradient similarity (GSM) is used to measure the difference in contrast and structure for FR IQA metric design in [28]. Wu *et al.* [27] built a FR IQA model by dividing the input image into two different regions based on internal generative mechanism. Recently, there are many studies using machine learning techniques to design FR IQA metrics [36], [37].

Besides the FR IQA metrics, many RR and NR IQA metrics have been proposed in the past decade [42]. Rehman and Wang [38] proposed a RR metric by structural similarity in, where statistical features from a multi-scale and multi-orientation divisive normalization transform are extracted for distortion measure. Narwaria *et al.* [39] used the phase and magnitude of discrete fourier transform (DFT) to design a RR IQA metric. Wu *et al.* [13] designed a RR metric based on visual information fidelity. Ma *et al.* [19] designed a metric for multi-exposure image fusion. For NR IQA, there have been also various studies investigating how to evaluate the image quality without any reference information [15], [16], [41]. Ong *et al.* [40] designed a NR metric to measure the visual quality of image blur by the average extent of edges in the image. Zhai *et al.* [41] proposed a NR IQA metric based on the extracted features from DCT domain. Gu *et al.* [14] computed the related features to design a NR metric in the autoregressive parameter space. Fang *et al.* [15] built natural scene statistics models for NR IQA. Wu and Wang [43] proposed an efficient NR IQA metric by statistical features extracted from binary patterns of local image structures. Li *et al.* [20] designed a NR metric for image blur quality assessment by using discrete orthogonal moments. There are also other NR metrics for image quality evaluation by using machine learning technologies [30], [44].

For these IQA metrics introduced above, they are mainly designed for VQA of natural images. The properties of natural images are different greatly from those of SCIs. Generally, the SCI includes repeated patterns, thin lines and sharp edges, while natural images usually include smooth edges, thick lines and more color information. Thus, the existing IQA metrics designed for natural images are not effective for visual quality evaluation of SCI. To investigate the visual quality of SCI, Yang *et al.* conducted an user study to construct a subjective database SIQAD (screen image quality assessment database) for SCI quality assessment recently [5]. In that database, there are 20 reference and 980 distorted screen content images created by different distortion types of Gaussian noise, Gaussian blur, motion blur, contrast changing, JPEG, JPEG2000 and layer segmentation based coding [45]. According to the analysis of subjective and objective results of different VQA metrics in that study, there is much room to improve the performance of SCI quality evaluation. To investigate the perceptual SCI coding, Wang *et al.* designed an VQA metric by visual field adaptation and information content weighting [17]. However, that study mainly focuses on the VQA for SCI coding. Ni *et al.* [52] proposed VQA metrics for SCIs based on edge [53] and gradient direction information. However, these metrics do not consider the different influences of textual region and pictorial region on the visual quality of SCIs.

In this study, we propose a FR VQA metric for SCI based on structure features and uncertainty weighting (SFUW). We first divide the SCI into textual and pictorial regions by the text segmentation method. It is well known that the HVS is sensitive to edge information, which is the basic component of characters in textual regions. Thus, we extract the structure feature of textual regions by gradient information for feature representation of VQA for textual regions. For natural images, existing studies have shown that image structure represents the primary visual information and the HVS is highly adapted to obtain structural information for visual perception and understanding [46], [47]. Besides the structure information, the HVS is also highly sensitive to luminance variation in a natural scene, which might lead to visual distortion [48], [49]. Thus, we extract the structure and luminance features for visual quality prediction of pictorial regions. The overall quality of the SCI is predicted by fusing the visual quality of pictorial and textual regions with uncertainty weighting.

There is much difference between SFUW and other existing related methods. In SPQA [5], the quality score of the textual region is computed by luminance and sharpness information, while the quality score of pictorial region is predicted by sharpness information only. The quality scores of textual region and pictorial region are combined linearly to obtain the overall quality of SCIs. Although the weighting values for textual region and pictorial region are different in SPQA, the weighting values for all textual pixels (or pictorial pixels) are the same and thus, SPQA does not consider the visual perception differences of local regions in SCIs. On the contrary, SFUW computes the quality score of textual region by the similarity of local gradient information, while luminance and structure information is used to calculate the quality score of pictorial region. Compared with SPQA [5] which uses the same weighting for all textual pixels (or pictorial pixels) to get the quality score of textual region (or pictorial region), we use the uncertainty weighting for the consideration of different influences on image patches to calculate the overall visual quality of SCIs. In SQI, Wang *et al.* [17] propose a VQA metric for SCIs by incorporating viewing field adaption and local information content weighting. However, it does not consider the visual perception differences for textual region and pictorial region in SCIs. It uses the same method to calculate the quality score of textual region and pictorial region in SCIs. In sum, the main contributions of this study include the following aspects.

- According to the properties of textual regions, we extract the structure feature to estimate the visual quality of textual regions. The gradient information is used

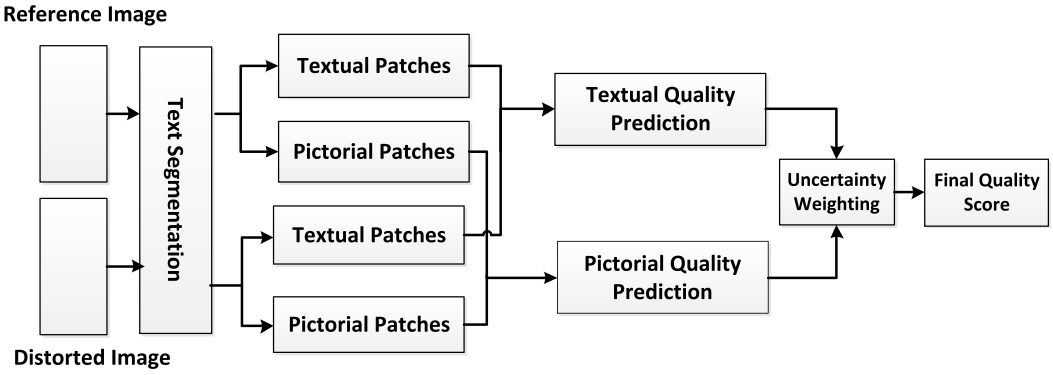

Fig. 1. The proposed framework.

to compute the structure feature of textual regions in SCI.

- For pictorial regions, we extract the luminance and structure features for quality prediction of pictorial regions. The quality map of pictorial regions is obtained by fusing those from luminance and structure features.
- To predict the visual quality of SCI, we design a novel uncertainty weighting fusion method to combine the visual quality of textual and pictorial regions.

II. Proposed Method

Our previous study [5] has shown that the statistical features of natural and textual images are different greatly. The detailed analysis of subjective data demonstrates that human perception on pictorial and textual regions is different from each other. Specifically, the distortion in textual regions are perceived differently from that of the overall SCI. Observers would be sensitive to the luminance and contrast change in pictorial regions, while for textual regions, they are more sensitive to blurring distortion than other types of distortion. Thus, it is reasonable to design different methods for visual quality assessment of pictorial and textual regions.

In this work, we propose a new VQA metric SFUW for SCIs. The framework of SFUW is shown in Fig. 1. First, we divide the input SCI into textual and pictorial patches by the text segmentation algorithm. Then the visual quality of textual patches can be estimated by the feature differences between textual patches from reference and distorted SCIs. Similarly, the visual quality of pictorial patches can be predicted by the feature differences between pictorial patches from reference and distorted SCIs. Through the uncertainty weighting, we predict the visual quality of the whole SCI by combining the visual quality of textual and pictorial patches in the SCI. We will introduce the details of SFUW in the following subsections.

A. Text Segmentation

We first use the texture segmentation method in [29] to segment the SCI into pictorial and textual regions. In that method, the authors first design a local image activity measure by variation distribution of characters, which would obtain a coarse textual layer including textual regions and a few pictorial regions with high activity. To obtain accurate segmentation results, the authors further propose a textual connected component by a scale and orientation invariant grouping algorithm to eliminate the survived pictorial regions. The details about that segmentation algorithm can be referred to the study [29].

B. VQA of Textual Patches

As we know, the content in textual regions of the SCI mainly includes various characters. Since characters are composed of various edges, we use the gradient to represent the structure feature of textual regions in SCI. The following filters with two directions are adopted to compute the gradient feature of textual regions: $h_x = [-1/2\ 0\ 1/2]$ and $h_y = [-1/2\ 0\ 1/2]'$. With these two filters, we can compute the structure feature of textual regions in reference and distorted SCIs as follows.

$$g_{rx} = h_x \bigotimes T_r, \tag{1}$$

$$g_{ry} = h_y \bigotimes T_r, \tag{2}$$

$$g_{dx} = h_x \bigotimes T_d, \tag{3}$$

$$g_{dy} = h_y \bigotimes T_d, \tag{4}$$

where T_r and T_d represent textual patches in the reference and distorted images, respectively; $\bigotimes$ denotes the convolution operation. (g_{rx}, g_{ry}) and (g_{dx}, g_{dy}) denote the gradient features with two directions for the textual patches in the reference and distorted images, respectively.

With the computed structure features in Eqs. (1) - (4), we calculate the similarity between the textual patches from the reference and distorted images as follows [23]:

$$S_m^t(g_{rk}, g_{dk}) = \frac{2\mu_{g_{rk}}\mu_{g_{dk}} + C_1}{\mu_{g_{rk}}^2 + \mu_{g_{dk}}^2 + C_1} \frac{2\sigma_{g_{rk}g_{dk}} + C_2}{\sigma_{g_{rk}}^2 + \sigma_{g_{dk}}^2 + C_2} \tag{5}$$

where $k \in \{x, y\}$; $S_m^t(g_{rk}, g_{dk})$ denotes the similarity between gradient features g_{rk} and g_{dk} for image patch m; μ_{rk} and μ_{dk} are mean values of the gradient g_{rk} and g_{dk}, respectively; σ_{rk} and σ_{dk} denote the standard variance values of the gradient g_{rk} and g_{dk}, respectively; $\sigma_{g_{rk}g_{dk}}$ is the covariance of the gradient g_{rk} and g_{dk}; C_1 and C_2 are two constant values.

(a) (b) (c)

Fig. 2. The visual samples of quality maps for textual regions. The first row: three distorted SCIs degraded by gaussian noise, gaussian blur and contrast change. The second row and third row: the corresponding quality maps of textual regions based on SSIM and SFUW. For each distorted SCI, the subjective quality score (DMOS), SSIM and SFUW values are also listed. (a) DMOS=68.713, SSIM=0.413, SFUW=0.572. (b) DMOS=72.937, SSIM=0.604, SFUW=0.477. (c) DMOS=37.350, SSIM=0.801, SFUW=0.872.

According to Eq. (5), we can calculate the quality maps of the textual patch m in horizontal and vertical directions as $S_m^t(g_{rx}, g_{dx})$ and $S_m^t(g_{ry}, g_{dy})$. Through calculating the mean value of these two quality maps separately, we can obtain two quality scores for textual patch m in horizontal and vertical directions as S_{mx}^t and S_{my}^t, respectively. After we calculate the similarity between gradient features with two directions for textual patches, we estimate the visual quality score of textual patch m as follows:

$$S_m^t = \frac{1}{2}(S_{mx}^t + S_{my}^t) \tag{6}$$

where S_m^t represents the visual quality of textual patch m.

In Fig. 2, we provide some samples for the degradation of textual regions. As shown in Fig. 2, In the first row, there are three distorted SCIs degraded by gaussian noise, gaussian blur and contrast change. The second row and third row show their corresponding quality maps of textual regions by SSIM and SFUW, regardless of pure black areas which represent pictorial regions in SCIs. We also give the quality scores of SSIM and SFUW in the caption of this figure. We can compare the quality maps of distorted SCIs with their subjective and objective scores denoted by DMOS, SSIM and SFUW respectively. In the third row, compared with the first and second quality maps, there are more white pixels in the third quality map, which demonstrates more image pixels with high similarity between the reference and distorted SCIs. The lower DMOS value in the third SCI also shows better subjective quality of this image than the first and second ones.

C. VQA of Pictorial Patches

As indicated in the previous section, we extract the structure and luminance features to predict the visual quality of pictorial patches. Here, we extract the normalized luminance features to compute luminance features of pictorial patches, while structure features of pictorial patches are calculated by LBP features. To calculate the luminance feature of pictorial patches, the local contrast normalization is applied to pictorial regions of SCI to mimic early visual system and remove the redundancy information in the visual scene. We use the normalization operation as follows.

$$I'(x, y) = \frac{I(x, y) - \mu_{(x,y)}}{\sigma_{(x,y)} + C_3} \tag{7}$$

where $I'(x, y)$ and $I(x, y)$ represent the normalized and original luminance values at location (x, y) in pictorial patches; $\mu_{(x,y)}$ and $\sigma_{(x,y)}$ denote the local mean and standard deviation values of local pictorial patches; C_3 is a constant parameter. $\mu_{(x,y)}$ and $\sigma_{(x,y)}$ can be computed as:

$$\mu_{(x,y)} = \sum_{r=-R}^{R} \sum_{h=-H}^{H} \omega_{(r,h)} I(x+r, y+h) \tag{8}$$

$$\sigma_{(x,y)} = \sqrt{\sum_{r=-R}^{R} \sum_{h=-H}^{H} \omega_{(r,h)} [I(x+r, y+h) - \mu_{(x,y)}]^2} \tag{9}$$

where $\{\omega_{(r,h)} | r = -R, ..., R; h = -H, ..., H\}$ define a unit-volume Gaussian window.

With the local normalization operation in Eq. (7), we compute the normalized luminance features of pictorial patch n in the reference and distorted SCIs as I'_{rn} and I'_{dn}, respectively. The luminance similarity between pictorial patches from the reference and distorted SCIs can be calculated as follows.

$$\mathbb{S}_n^l = \frac{2I'_{rn}I'_{dn} + C_4}{I'^2_{rn} + I'^2_{dn} + C_4} \tag{10}$$

where $\mathbb{S}_n^l$ denotes the quality map of pictorial patch n in the distorted SCI from luminance feature; C_4 is a constant for numerical stability. Please note that the quality map $\mathbb{S}_n^l$ is computed in a pixel-wise manner.

We extract the structure feature of pictorial patches by using the rotation invariant uniform LBP descriptor [31] on the original SCI. The general LBP representation can be formulated as follows.

$$LBP_{K,R} = \sum_{i=0}^{K-1} t(I_i - I_c) 2^i, \tag{11}$$

$$t(I_i - I_c) = \begin{cases} 1, (I_i - I_c) \geq 0 \\ 0, (I_i - I_c) < 0 \end{cases} \tag{12}$$

where K and R denote the number of neighbors and the radius of the neighborhood; I_c is the luminance value of the center pixel in the local patch; $(I_0, I_1, ..., I_{(K-1)})$ represent the luminance values of K circularly symmetric neighborhood. Based on the study [31], we can define the local rotation invariant uniform LBP operator as:

$$LBP'_{K,R} = \begin{cases} \sum_{i=0}^{K-1} t(I_i - I_c), U(LBP_{K,R}) \leq 2 \\ K+1, Otherwise \end{cases} \tag{13}$$

$$U(LBP_{K,R}) = \|t(I_{(K-1)} - I_c) - t(I_0 - I_c)\| + \sum_{i=0}^{K-1} \|t(I_i - I_c) - t(I_{(i-1)} - I_c)\| \tag{14}$$

where U is computed as the number of bitwise transitions.

After extracting the LBP features by Eq. (13), we compute the similarity from structure feature between pictorial patches from the reference and distorted SCIs as follows.

$$\mathbb{S}_n^s = \frac{2B_{rn}B_{dn} + C_5}{B_{rn}^2 + B_{dn}^2 + C_5} \tag{15}$$

where $\mathbb{S}_n^s$ is the quality map of pictorial patch n in the distorted SCI from structure feature; B_{rn} and B_{dn}, computed by Eq. (13), are LBP features of pictorial patch n from the reference and distorted SCIs, respectively. Also, the quality map $\mathbb{S}_n^s$ is computed in a pixel-wise manner.

According to Eqs. (10) - (15), we compute the quality maps from luminance and structure features for SCI. The final quality map of pictorial patches can be obtained by combining these two from luminance and structure features as follows.

$$\mathbb{S}_n^p = \mathbb{S}_n^l \mathbb{S}_n^s \tag{16}$$

where $\mathbb{S}_n^p$ denotes the quality map of pictorial patch n.

With the computed visual quality map of each pictorial patch, the visual quality score of each pictorial patch is represented by the average value of the visual quality map as follows.

$$S_n^p = \frac{1}{N} \sum_{(x,y)} \mathbb{S}_n^p(x, y) \tag{17}$$

where N is the total amount of pixels in the pictorial patch; (x, y) is the pixel location in the pictorial patch; S_n^p denotes the visual quality score of pictorial patch n in the distorted SCI.

In Fig. 3, we give some visual samples to demonstrate the visual degradation of pictorial regions in SCI. The first row of this figure includes three distorted SCIs degraded by motion blur, gaussian blur and contrast change, while the corresponding quality maps of pictorial regions in these three SCIs are shown in the second row and third row calculated by SSIM and SFUW. In the third row, from the third quality map, we can observe that the there are more white pixels with high similarity between the reference and distorted SCIs than the other two. This is consistent with the subjective quality denoted by DMOS values in these three SCIs, which show that the DMOS value of the third SCI is lower than those of the other two. However, in the second row, there are more white pixels with highly similarity in the fist quality map calculated by SSIM and less white pixels in the third quality map. That means the predicted quality of the first distorted SCI is better than the predicted quality of the third one by SSIM, which is opposite to the subjective results from DMOS values.

D. Pooling Strategy by Uncertainty Weighting

According to Eqs. (6) and (17), we can compute the visual quality scores of textual and pictorial patches in distorted SCIs. The overall visual quality of the distorted SCI can be predicted by combining visual quality scores of textual and pictorial patches in the SCI. We propose to use the uncertainty weighting to fuse the visual quality of textual and pictorial parts to obtain the overall quality of SCIs based on the characteristic of the HVS. Generally, the HVS is more sensitive to high-frequency information (such as edge information) than other smooth regions in the visual scene. Thus, the distortion in the high-frequency regions with more edge information is more sensitive to the HVS than that in other smooth regions. In this study, we measure the edge information degree of different patches in distorted SCIs by

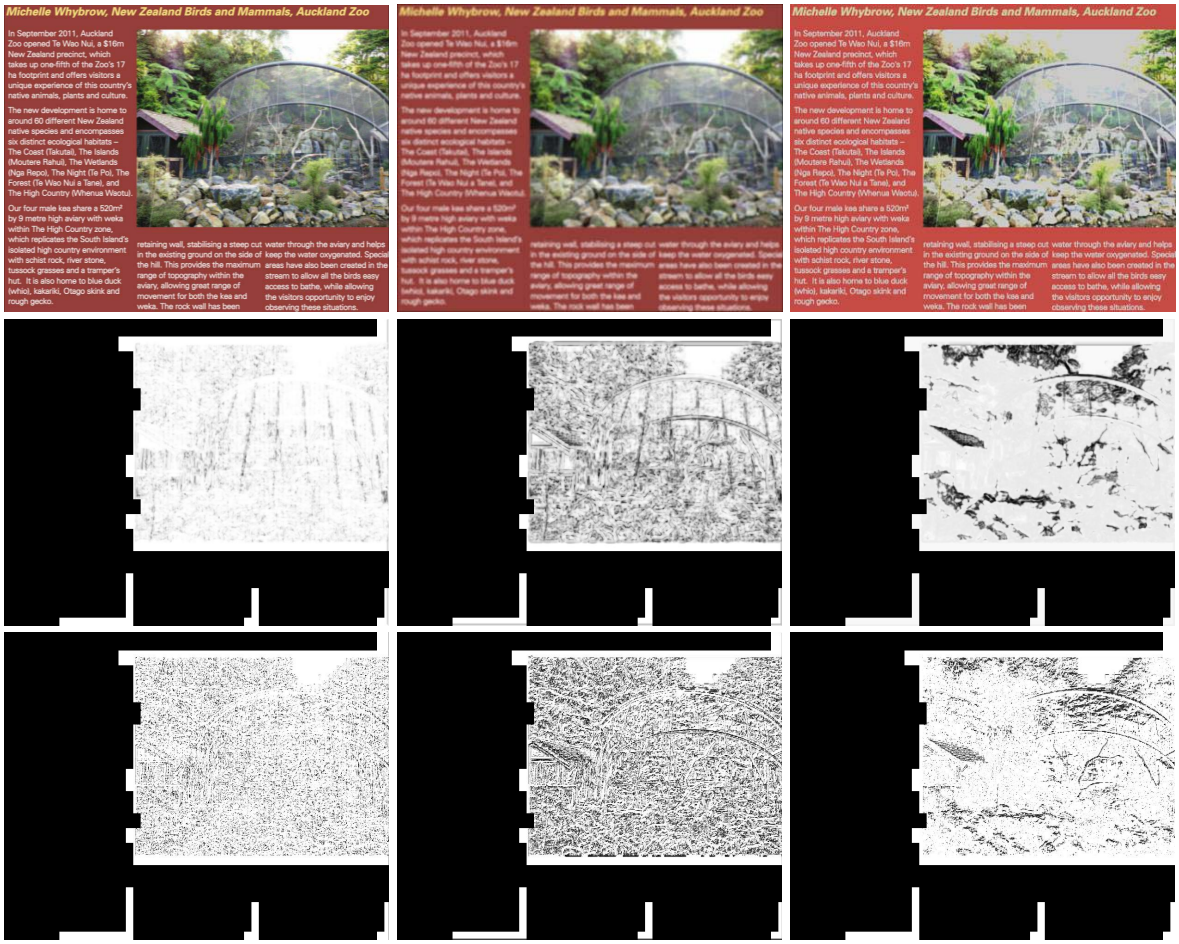

Fig. 3. The visual samples of quality maps for pictorial regions. The first row: the distorted SCIs degraded by motion blur, gaussian blur, and contrast change. The second row and third row: the corresponding quality maps of pictorial regions calculated by SSIM and SFUW. For each distorted SCI, the subjective quality score (DMOS), SSIM and SFUW values are also listed.(a) DMOS=52.956, SSIM=0.983, SFUW=0.808. (b) DMOS=72.776, SSIM=0.632, SFUW=0.501. (c) DMOS=45.872, SSIM=0.883, SFUW=0.870.

entropy of the gradient information, which is used as the uncertainty weighting in the proposed SFUW.

To measure the edge information degree in different patches of any distorted SCI, we can quantify the perceptual uncertainty v of any image patch (textual or pictorial patch) by the entropy of the gradient information as follows:

$$v = -\sum_{j=0}^{L} p_j * log_2 p_j \tag{18}$$

where L is the maximum pixel value in the gradient patch, p_j denotes the probability of the pixel value equal to j in the gradient patch. It can be computed as follows:

$$p_j = \frac{N_j}{N}, \tag{19}$$

where N_j represents the number of pixel value equal to j in the gradient patch; N is the total number of pixels in the gradient patch. we set N to 256 and j is in the range of 0 and L.

Based on the uncertainty weighting calculation in Eq. (18), we can estimate the overall quality score of textual regions in SCI as follows:

$$S^t = \frac{\sum_{m=1}^{N_t} v_m * S_m^t}{\sum_{m=1}^{N_t} v_m}, \tag{20}$$

where v_m denotes the perceptual uncertainty of textual patch m which is computed by Eq. (18); S_m^t is the visual quality score of textual patch m which is predicted by Eq. (6); N_t is the number of textual patches in the distorted SCI.

Similarly, we predict the overall quality of pictorial regions by combining these of all pictorial patches as follows:

$$S^p = \frac{\sum_{n=1}^{N_p} v_n * S_n^p}{\sum_{n=1}^{N_p} v_n}, \tag{21}$$

where v_n denotes the perceptual uncertainty of pictorial patch n which is computed as Eq. (18); S_n^p is the visual quality scores of pictorial patch n predicted by Eq. (17); N_p is the number of pictorial patches in the distorted SCI.

After computing the visual scores of textual and pictorial regions as Eqs. (20) and (21), we predict the overall visual

quality S of the SCI by combining them as follows:

$$S = \omega_t S^t + \omega_p S^p \tag{22}$$

$$\omega_t = \frac{\bar{v_t}}{\bar{v_t} + \bar{v_p}} \tag{23}$$

$$\omega_p = \frac{\bar{v_p}}{\bar{v_t} + \bar{v_p}} \tag{24}$$

where $\bar{v_t}$ and $\bar{v_p}$ are the averaging information measure of textual and pictorial regions in the SCI, respectively. They are computed as follows:

$$\bar{v_t} = \frac{1}{N_t} \sum_{m=1}^{N_t} v_m, \tag{25}$$

$$\bar{v_p} = \frac{1}{N_p} \sum_{n=1}^{N_p} v_n. \tag{26}$$

III. Experimental Results

A. Evaluation Methodology

To demonstrate the advantages of SFUW, we use the image database in [5] to conduct the comparison experiments. As indicated previously, this database includes 20 reference SCIs in total. For each SCI in this database, there are seven distortion types (Gaussian Noise, Gaussian Blur, Motion Blur, Contrast Change, JPEG, JPEG2000, and Layer Segmentation Based Coding) with seven degradation levels and thus there are 980 distorted SCIs in total. These reference images obtained from webpages, slides, PDF files and digital magazines are diverse with the visual content.

The 11-category Absolute Category Rating (ACR) is used in the subjective experiment. In total, there were 96 subjects involved in the test and each image was rated by at least 30 subjects. The participants' ages range from 19 to 38 years. After the raw subjective scores were obtained, outliers were removed to obtain the DMOS values.

Here, we adopt three commonly used methods to compute the correlation between the subjective and objective scores: SRCC (Spearman Rank-order Correlation Coefficient), PLCC (Pearson Linear Correlation Coefficient), and RMSE (Root Mean Squared Error). SRCC can be used to evaluate the prediction monotonicity, while PLCC can be adopted to assess the prediction accuracy. RMSE is a measure of deviation between the objective and subjective scores. Generally, a better visual quality assessment method has higher SRCC and PLCC values, and lower RMSE value. Given the ith image in the database (with N images in total), its objective and subjective scores are o_i and s_i. We can estimate the PLCC as follows.

$$PLCC = \frac{\sum_{i=1}^{N}(o_i - \overline{o})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(o_i - \overline{o}) * \sum_{i=1}^{N}(s_i - \overline{s})}}, \tag{27}$$

where $\overline{o}$ and $\overline{s}$ denote the mean values of o_i and s_i, respectively.

SRCC can be computed as follows.

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N(N^2-1)}, \tag{28}$$

TABLE I

Experimental Results by Different Weighting Methods. AW: Avergae Weighting; ICW: Information Content Weighting [17]; UW: Uncertainty Weighting

Components	AW	ICW	UW
PLCC	0.770	0.843	0.891
SRCC	0.741	0.829	0.880
RMSE	9.141	7.703	6.499

where e_i is the difference between the ith image's ranks in subjective and objective results.

RMSE can be calculated as follows.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(o_i - s_i)^2}{N}} \tag{29}$$

The estimated quality scores by different IQA metrics might have different ranges, we use a five-parameter mapping function to nonlinearly regress the quality scores into a common space as follows:

$$f(x) = \beta_1 (\frac{1}{2} - \frac{1}{1 + e^{(\beta_2 (x - \beta_3))}}) + \beta_4 x + \beta_5 \tag{30}$$

where $(\beta_1, ..., \beta_5)$ are the parameters to be fitted.

B. Experiment 1

In this section, we conduct the comparison experiment to demonstrate the advantages of the uncertainty weighting used in SFUW. From the pooling strategy by uncertainty weighting of image patches in the previous section, we can see that the patches with more complicated gradient information would be assigned larger weighting for textual or pictorial regions. From the pooling strategy for quality scores of textual and pictorial regions, we can see that the weighting would be larger with more complicated gradient information in textual or pictorial regions. In the experiments, we set the size of patch to $16 * 16$, $c_1 = c_3 = c_4 = 6.5025$, $c_2 = c_5 = 58.5225$.

A common weighting method in IQA is the average weighting to obtain the overall visual quality of the image from those of all image patches in the image. Here, we use the average weighting method to conduct the comparison experiment to demonstrate the effect of uncertainty weighting in SFUW. In the average weighting, we assign the same weighting to the pixels in each patch, regardless of the patch belonging to textual or pictorial region. The overall quality score of the distorted SCI will be calculated by averaging the visual scores of all patches. Besides the average weighting, we also use the information content weighting method proposed in [17] in this comparison experiment. In that study, the concept of information content is used to represent the contribution of each image patch to the overall quality of the SCI.

We provide the experimental results of visual quality scores calculated by average weighting (AW), information content weighting (ICW) and the proposed uncertainty weighting in Table I. From this table, we can see that the performance of average weighting which sets the weighting value of each patch to be the same is worst. The information content weighting method in [17] can obtain better performance than

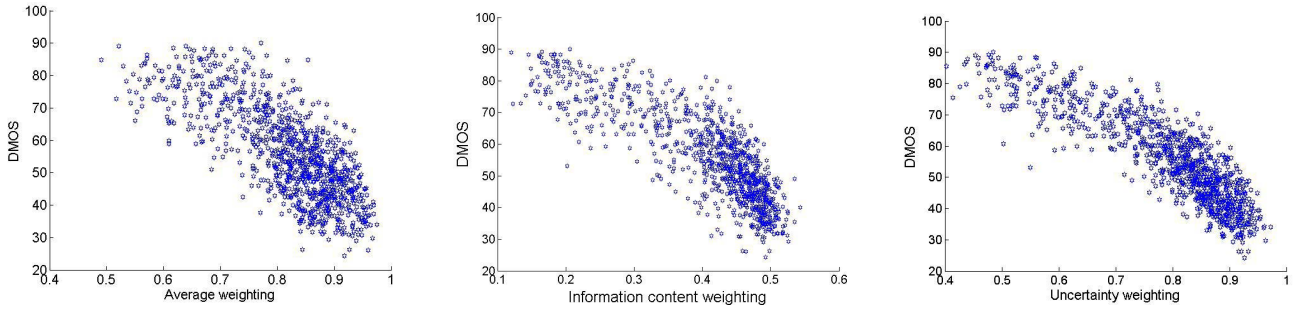

Fig. 4. Comparison results of different weighting methods. The scatter plots of objective quality scores against subjective scores for different weighting methods.

TABLE II
EXPERIMENTAL RESULTS (PLCC) OF SFUW AND OTHER EXISTING METHODS ON DIFFERENT DISTORTION TYPES

	PLCC												
Distortions	PSNR	SSIM	MSSIM	IWSSIM	VIF	IFC	MAD	GMSD	SPQA	SQI	EMSQA	GSS	SFUW
GN	**0.905**	0.886	0.885	0.888	**0.902**	0.880	0.885	**0.899**	0.892	0.883	0.889	0.865	0.887
GB	0.860	0.901	0.900	0.908	0.910	0.904	0.905	0.910	0.906	**0.920**	**0.916**	0.907	**0.923**
MB	0.704	0.806	0.807	0.842	0.850	0.692	0.836	0.844	0.831	**0.879**	**0.875**	0.831	**0.878**
CC	0.753	0.756	**0.837**	**0.841**	0.708	0.687	0.387	0.783	0.799	0.772	0.769	0.609	**0.829**
JPEG	0.770	0.749	0.786	**0.799**	**0.801**	0.763	0.735	0.775	0.770	**0.822**	0.790	0.795	0.757
J2K	0.789	0.775	0.800	**0.840**	0.822	0.794	**0.831**	**0.851**	0.825	0.827	0.785	0.813	0.815
LSC	0.781	0.731	0.789	0.816	**0.841**	0.773	0.819	**0.856**	0.796	**0.831**	0.775	0.808	0.759
Overall	0.587	0.591	0.620	0.654	0.821	0.640	0.619	0.726	0.858	0.864	0.865	0.846	**0.891**

the average weighting method. From the PLCC, SRCC and RMSE values, we can observe that SFUW by uncertainty weighting can obtain the best performance among the compared ones. Meanwhile, we also show the scatter plots describing the distributions of objective quality scores against the DMOS values from these three different weighting methods in Fig. 4. From this figure, we can observe that the points in the first scatter plot by average weighting are more dispersive than the second scatter plot by information content weighting, which demonstrates that the information content weighting method can distinguish the visual distortion with different degrees better than the average weighting method. From the last scatter plot, we can observe that the points by the proposed uncertainty weighting is more centralized than the other two, showing that the objective quality scores obtained by SFUW are more consistent with the subjective scores.

C. Experiment 2

In this experiment, we perform the comparison experiments by using SFUW and the following existing visual quality metrics: PSNR, SSIM [23], IWSSIM [51], MSSIM [50], VIF [26], IFC [32], MAD [34], GMSD [33], SPQA [5], SQI [17], EMSQA [53] and GSS [52]. We compute the correlations in terms of PLCC, SRCC and RMSE values between the subjective scores and the predicted objective scores from the used compared metrics. The experimental results are shown in Tables II, III, and IV.

From Tables II, III, and IV, we can observe that GMSD can obtain better performance than PSNR, SSIM, IFC and MAD in visual quality prediction of SCIs. Obviously, SPQA, SQI, EMSQA and GSS can obtain much better performance than other existing ones. These four metrics are designed specifically for the SCI. Among all the compared metrics, SFUW can obtain the best performance on visual quality prediction of SCIs, which can be demonstrated by the highest PLCC and SRCC values in Tables II and III, and the lowest RMSE value in Table IV.

It is well known that the advanced perceptual metrics such as SSIM, MAD and IFC can obtain promising performance of natural images. However, the performance of these metrics on SCI is not good enough, as demonstrated by the experimental results in Tables II, III, and IV. These experimental results demonstrate that these well-known metrics cannot work well for textual regions in SCI, and thus, we have to design effective metrics for VQA of SCI. As mentioned in [5], MSCN coefficient can be represented by naturalness of image pixels. The distributions of nature images and screen content images are different greatly. As demonstrated by [5], the coefficients of nature images follow a Gaussian distribution, while a sharp pimpling appears in the distribution of screen content images. Thus, MSCN can not obtain good performance in VQA of SCI. In SFUW, we extract the different features of textual and pictorial regions separately for VQA of these two types of regions. The final quality score of the distorted SCI is estimated by fusing those of textual and pictorial regions with uncertainty weighting. Both the perceptual features extracted for textual and pictorial regions specifically and the uncertainty weighting can contribute much to the promising performance of SCI quality estimation in SFUW.

We also provide the scatter plots of the distributions of predicted quality scores against the subjective scores (DMOS)

TABLE III
EXPERIMENTAL RESULTS (SRCC) OF SFUW AND OTHER EXISTING METHODS ON DIFFERENT DISTORTION TYPES

	SRCC												
Distortions	PSNR	SSIM	MSSIM	IWSSIM	VIF	IFC	MAD	GMSD	SPQA	SQI	EMSQA	GSS	SFUW
GN	0.879	0.869	0.870	0.874	**0.889**	0.869	0.872	**0.886**	**0.882**	0.860	0.875	0.852	0.869
GB	0.858	0.893	0.890	0.906	0.906	0.907	0.909	0.912	0.902	**0.924**	**0.915**	0.905	**0.917**
MB	0.713	0.804	0.803	0.842	0.850	0.689	0.836	0.844	0.823	**0.881**	**0.879**	0.840	**0.874**
CC	0.683	0.641	**0.750**	**0.754**	0.527	0.640	0.391	0.544	0.611	0.668	0.631	0.597	**0.722**
JPEG	0.757	0.758	0.785	**0.798**	0.793	0.767	0.767	0.771	0.767	**0.819**	0.787	**0.797**	0.750
J2K	0.775	0.760	0.788	0.799	0.815	0.794	**0.838**	**0.844**	0.815	**0.817**	0.776	0.814	0.812
LSC	0.793	0.737	0.787	0.821	**0.850**	0.777	0.815	**0.859**	0.800	**0.843**	0.780	0.816	0.754
Overall	0.561	0.582	0.611	0.655	0.807	0.601	0.607	0.731	0.842	0.855	0.850	0.836	**0.880**

TABLE IV
EXPERIMENTAL RESULTS (RMSE) OF SFUW AND OTHER EXISTING METHODS ON DIFFERENT DISTORTION TYPES

	RMSE												
Distortions	PSNR	SSIM	MSSIM	IWSSIM	VIF	IFC	MAD	GMSD	SPQA	SQI	EMSQA	GSS	SFUW
GN	**6.338**	6.910	6.942	6.851	**6.456**	7.090	6.955	6.521	6.739	**6.186**	6.832	7.497	6.876
GB	7.738	6.570	6.631	6.353	6.287	6.287	6.453	6.310	6.430	**5.802**	**6.071**	6.383	**5.592**
MB	9.229	7.697	7.682	7.020	6.843	9.392	7.132	6.982	7.222	**6.127**	**6.288**	7.225	**6.236**
CC	8.282	8.238	**6.877**	**6.804**	8.879	9.141	11.600	7.828	7.618	7.996	8.044	9.976	**7.048**
JPEG	6.000	6.230	5.816	**5.640**	**5.632**	6.070	6.369	5.941	6.000	**5.502**	5.756	5.702	6.143
J2K	6.382	6.563	6.233	6.178	5.918	6.325	**5.781**	**5.459**	**5.871**	5.872	6.349	6.051	6.023
LSC	5.330	5.825	5.243	4.938	**4.611**	5.408	4.900	**4.411**	5.166	**4.898**	5.395	5.022	5.555
Overall	11.590	11.545	11.236	10.833	8.180	10.992	11.241	9.642	7.342	7.198	7.186	7.631	**6.499**

in Fig. 5. From this figure, we can observe that the points from PSNR, SSIM, IWSSIM, IFC, and MAD are dispersive, which demonstrate that the predicted objective scores from these metrics is not much consistent with the subjective scores. Compared with these metrics, the points from VIF, GMSD and SQI are more centralized, which demonstrate that the predicted objective scores by these three metrics are more consistent with the subjective scores than other existing ones. From these scatter plots, the points from SFUW are the most centralized, which shows that SFUW can predict the best visual quality of SCI among the compared metrics. This is consistent with the experimental results in Tables II, III, and IV.

D. Performance Analysis

Compared with state-of-the-art VQA metrics designed for nature image such as SSIM, IWSSIM and GMSD, the VQA metrics including SPQA, SQI and SFUW which are designed specifically for VQA of SCIs can obtain better performance in quality prediction for SCIs. The main reason is that these VQA methods designed specifically for SCIs consider the differences between visual perception of textual and pictorial regions in SCIs. For example, the HVS would be more sensitive to the distortion of motion blur in textual region than that in pictorial region. Also, the HVS is more sensitive to the distortion of contrast variation in pictorial region than that in textual region. Thus, SSIM (or IWSSIM, GMSD, *etc*.) which predicts the visual quality of both textual an pictorial regions in the same method and combines them with mean weighting strategy cannot obtain good performance for VQA of SCIs. Different from these metrics, SQI incorporates viewing field adaption and local information content, and thus, it can obtain better performance than these metrics designed for natural images. In existing VQA metrics (such as SQI, EMSQA and GSS) designed specifically for SCIs, the calculation methods for textual region and pictorial region are the same. On the contract, SFUW proposes to use different methods for visual quality prediction for textual and pictorial regions based on the characteristic of the HVS. We extract the gradient feature as the structure feature for visual quality prediction of textual regions, since the HVS is much sensitive to the structure information in textual regions. The luminance and structure features are used for visual quality estimation of pictorial regions, since the HVS is highly adapted to obtain structural information for visual perception and understanding [46], [47], and it is also highly sensitive to luminance variation in natural scenes [48], [49]. In SFUW, we use a novel weighting strategy by uncertainty weighting to combine the quality scores of image patches in SCIs. The uncertainty is estimated by the energy of gradient information in image patches, since the HVS is more sensitive to high-frequency information than other smooth regions in visual scenes.

From the experimental results, PSNR can get the best performance among all the compared methods for Gaussian Noise. The reason is that PSNR is a pixel pair comparison method and it can detect each distorted pixel in SCIs. However, the overall performance of PSNR is inferior to other methods, for it does not consider the characteristics of the HVS [21]. Meanwhile, the performance of SFUW on some specific distortion types (GB, MB and CC) is almost the best among the compared methods. This demonstrates that the used features in SFUW can effectively detect these distortion types in SCIs. SFUW might be worse than some existing methods for several specific distortion types such as LSC, since the used features

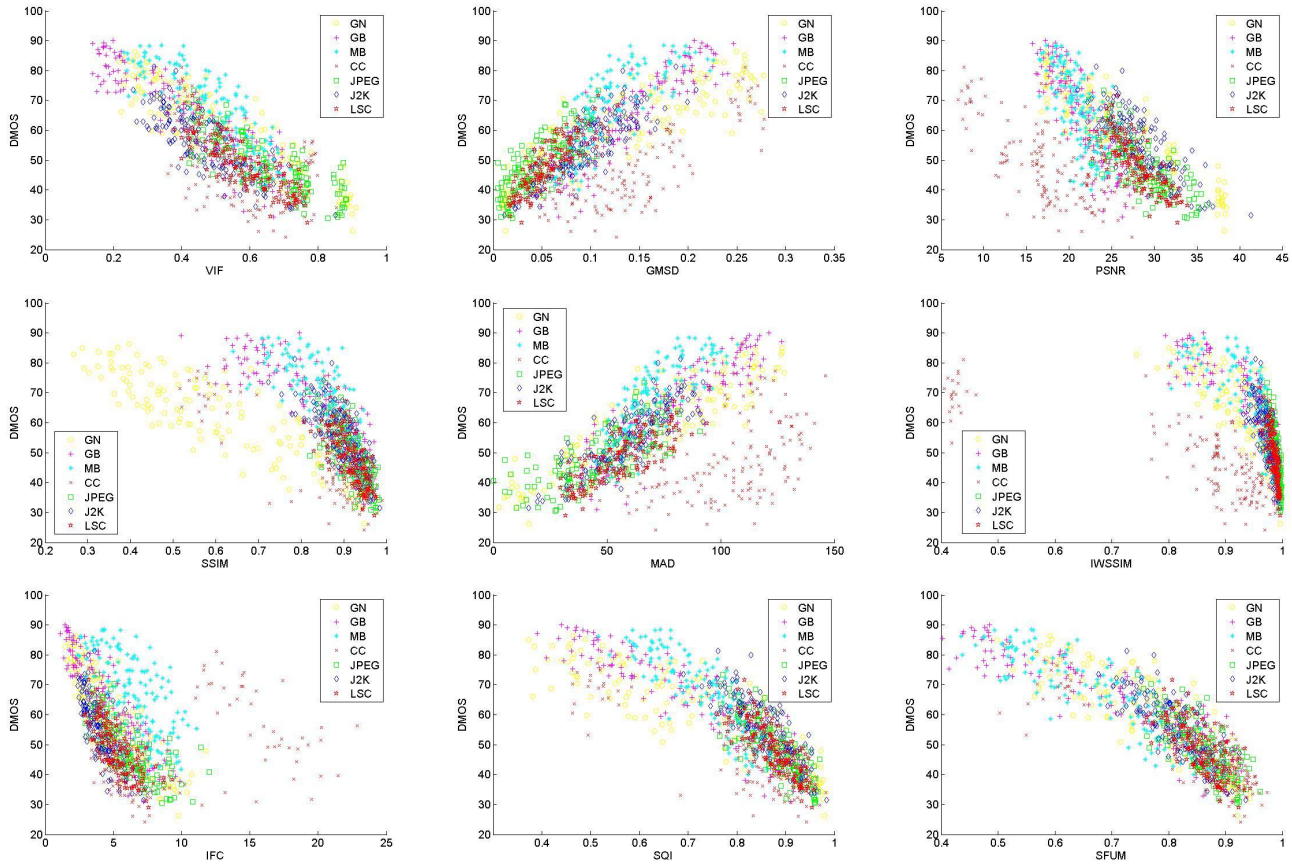

Fig. 5. The scatter plots of predicted quality scores by some metrics against the DMOS values on the SIQAD. The vertical axis in each figure is the DMOS values. From left to right, Row 1: VIF, SSIM and PSNR; Row 2: MSSIM, MAD and IWSSIM; Row 3: IFC, GMSD and SFUW.

in SFUW might not be suitable for all distortion types due to the characteristics of different visual distortion types. We will investigate into this problem further in the future to design more effective VQA metric for SCIs with different distortion types.

IV. Conclusion

In this study, we have proposed a new FR VQA metric for SCI based on uncertainty weighting. For textual regions of SCI, we extract the structure features by using the gradient information for visual quality prediction of textual regions. For pictorial regions of SCI, the luminance and structure features are computed by intensity and LBP information for visual quality prediction. The final visual quality of SCI is estimated by fusing these of textual and pictorial regions with uncertainty weighting. Experimental results demonstrate that the proposed SFUW can obtain superior performance against start-of-the-art approaches.

References

[1] T.-H. Chang and Y. Li, "Deep shot: A framework for migrating tasks across devices using mobile phone cameras," in *Proc. ACM SIGCHI Conf. Human Factors Comput. Syst.*, 2011, pp. 2163–2172.

[2] Y. Lu, S. Li, and H. Shen, "Virtualized screen: A third element for cloud-mobile convergence," *IEEE Multimedia Mag.*, vol. 18, no. 2, pp. 4–11, Feb. 2011.

[3] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, 2014.

[4] 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.

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

[6] C. Lan, J. Xu, W. Zeng, and F. Wu, "Compound image compression using lossless and lossy LZMA in HEVC," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun./Jul. 2015, pp. 1–6.

[7] 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.

[8] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-inspired divisive normalization for perceptual video coding," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, Apr. 2013.

[9] S. Minaee and Y. Wang, "Screen content image segmentation using least absolute deviation fitting," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3295–3299.

[10] J. Lei, S. Li, C. Zhu, M.-T. Sun, and C. Hou, "Depth coding based on depth-texture motion and structure similarities," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 2, pp. 275–286, Feb. 2015.

[11] J. Lei, J. Sun, Z. Pan, S. Kwong, J. Duan, and C. Hou, "Fast mode decision using inter-view and inter-component correlations for multiview depth video coding," *IEEE Trans. Ind. Informat.*, vol. 11, no. 4, pp. 978–986, Aug. 2015.

[12] S. Ma, X. Zhang, J. Zhang, C. Jia, S. Wang, and W. Gao, "Nonlocal in-loop filter: The way toward next-generation video coding?" *IEEE MultiMedia*, vol. 23, no. 2, pp. 16–26, Apr./Jun. 2016.

[13] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 1700–1705, Nov. 2013.

[14] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.

[15] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.

[16] L. Ma, L. Xu, Y. Zhang, Y. Yan, and K. N. Ngan, "No-reference retargeted image quality assessment based on pairwise rank learning," *IEEE Trans. Multimedia*, vol. 18, no. 11, pp. 2228–2237, Nov. 2016.

[17] 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.*, doi: 10.1109/MCG.2016.46, 2017, to be published.

[18] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.

[19] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

[20] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.

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

[22] 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.

[23] 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.

[24] H. Yang, Y. Fang, Y. Yuan, and W. Lin, "Subjective quality evaluation of compressed digital compound images," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 105–114, Jan. 2014.

[25] 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.

[26] 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.

[27] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.

[28] 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.

[29] 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. 519–533, Mar. 2015.

[30] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[31] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[32] 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.

[33] 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.

[34] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, 2010.

[35] 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.

[36] M. Narwaria, W. Lin, and A. E. Cetin, "Scalable image quality assessment with 2D mel-cepstrum and machine learning approach," *Pattern Recognit.*, vol. 45, no. 1, pp. 299–313, 2012.

[37] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–1807, May 2013.

[38] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.

[39] M. Narwaria, W. Lin, I. V. McLoughlin, S. Emmanuel, and L. T. Chia, "Fourier transform-based scalable image quality measure," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3364–3377, Aug. 2012.

[40] E. Ong *et al.*, "A no-reference quality metric for measuring image blur," in *Proc. IEEE Int. Symp. Signal Process. Appl.*, Jul. 2003, pp. 469–472.

[41] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu, "No-reference noticeable blockiness estimation in images," *Signal Process., Image Commun.*, vol. 23, no. 6, pp. 417–432, 2008.

[42] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[43] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 339–343.

[44] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Process., Image Commun.*, vol. 40, pp. 1–15, Jan. 2016.

[45] 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.

[46] H. Nothdurft, "Sensitivity for structure gradient in texture discrimination tasks," *Vis. Res.*, vol. 25, no. 12, pp. 1957–1968, 1985.

[47] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A*, vol. 14, no. 9, pp. 2379–2391, Sep. 1997.

[48] R. A. Frazor and W. S. Geisler, "Local luminance and contrast in natural images," *Vis. Res.*, vol. 46, no. 10, pp. 1585–1598, 2006.

[49] V. Mante, R. A. Frazor, V. Bonin, W. S. Geisler, and M. Carandini, "Independence of luminance and contrast in natural scenes and in the early visual system," *Nature Neurosci.*, vol. 8, no. 12, pp. 1690–1697, 2005.

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

[51] 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.

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

[53] 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.*, Sep. 2016, pp. 81–85.

[54] 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, Aug. 2016.

Yuming Fang (M'13) received the Ph.D. degree from Nanyang Technological University, Singapore, the M.S. degree from the Beijing University of Technology, Beijing, China, and the B.E. degree from Sichuan University, Chengdu, China. He is currently a Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He has authored or co-authored over 90 academic papers in international journals and conferences in the areas of multimedia processing. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, and 3-D image/video processing. He serves as an Associate Editor of the IEEE ACCESS and is on the Editorial Board of *Signal Processing: Image Communication*.

Jiebin Yan received the B.E degree from Beijing City University, Beijing, China. He is currently pursuing the M.S degree with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include image quality assessment and Machine learning.

Jiaying Liu (S'09–M'10) received the B.E. degree in computer science from Northwestern Polytechnic University, Xi'an, China, and the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2005 and 2010, respectively. She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored over 80 technical articles in refereed journals and proceedings. She holds 15 granted patents. Her current research interests include image/video processing, compression, and computer vision. She was a Visiting Scholar with the University of Southern California at Los Angeles, Los Angeles, from 2007 to 2008. She was a Visiting Researcher with Microsoft Research Asia in 2015, supported by Star Track for Young Faculties. She has also served as a TC member in APSIPA IVM since 2015, and APSIPA Distinguished Lecture from 2016 to 2017.

Shiqi Wang received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in computer application technology from Peking University, in 2014. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. He is currently with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. He has authored and coauthored over 80 technical articles in refereed journals and proceedings in the areas of image and image/video coding, processing, quality assessment and analysis. He has also proposed over 30 technical proposals to ISO/MPEG, ITU-T and AVS video coding standards.

Qiaohong Li received the B.E. degree and M.E. degree from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her research interests include image quality assessment, speech quality assessment, computer vision, and visual perceptual modeling.

Zongming Guo (M'09) received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively. He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication. He is the Executive Member of the China-Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, the Wang Xuan News Technology Award and the Chia Tai Teaching Award in 2008, the Government Allowance granted by the State Council in 2009, and the Distinguished Doctoral Dissertation Advisor Award of Peking University in 2012 and 2013.