Image Quality Assessment Based on Gradient Similarity

Anmin Liu, Weisi Lin, *Senior Member, IEEE*, and Manish Narwaria

Abstract—In this paper, we propose a new image quality assessment (IQA) scheme, with emphasis on gradient similarity. Gradients convey important visual information and are crucial to scene understanding. Using such information, structural and contrast changes can be effectively captured. Therefore, we use the gradient similarity to measure the change in contrast and structure in images. Apart from the structural/contrast changes, image quality is also affected by luminance changes, which must be also accounted for complete and more robust IQA. Hence, the proposed scheme considers both luminance and contrast–structural changes to effectively assess image quality. Furthermore, the proposed scheme is designed to follow the masking effect and visibility threshold more closely, i.e., the case when both masked and masking signals are small is more effectively tackled by the proposed scheme. Finally, the effects of the changes in luminance and contrast–structure are integrated via an adaptive method to obtain the overall image quality score. Extensive experiments conducted with six publicly available subject-rated databases (comprising of diverse images and distortion types) have confirmed the effectiveness, robustness, and efficiency of the proposed scheme in comparison with the relevant state-of-the-art schemes.

Index Terms—Contrast masking, gradient similarity, human visual system (HVS), image quality assessment (IQA), structural similarity (SSIM).

I. INTRODUCTION

DIGITAL images are usually affected by a wide variety of distortions during acquisition and processing, which generally results in loss of visual quality. Therefore, image quality assessment (IQA) is useful in many applications such as image acquisition, watermarking, compression, transmission, restoration, enhancement, and reproduction. The goal of IQA is to calculate the extent of quality degradation and is thus used to evaluate/compare the performance of processing systems and/or optimize the choice of parameters in processing. For example, the well-cited structural similarity (SSIM) index [1] has been used in image and video coding [9], [31].

The human visual system (HVS) is the ultimate receiver of the majority of processed images, and evaluation based on subjective experiments (as formally defined in ITU-R Recommendation BT.500 [8]) is the most reliable way of IQA. However,

subjective evaluation is time consuming, laborious, expensive, and non-repeatable; as a result, it cannot be easily and routinely performed for many scenarios, e.g., selection the prediction mode in H.264 video coding. These limitations have led to the development of objective IQA measures that can be easily embedded in image processing systems.

The simplest and most widely used IQA scheme is the mean squared error (MSE)/peak signal-to-noise ratio (PSNR) (which is calculated using MSE). It is popular due to its mathematical simplicity and it being easy to optimize. It is, however, well known that MSE/PSNR does not always agree with the subjective viewing results, particularly when distortion is not additive in nature [41]. This is not surprising given that it is simply an average of the squared pixel differences between the original and distorted images.

Aimed at accurately and automatically evaluating the image quality in a manner that agrees with subjective human judgments (HVS-oriented), regardless of the type of distortion corrupting the image, the content of the image, or the strength of the distortion, substantial research effort has been directed toward developing IQA schemes over the years, as reviewed in [1] and [44]. The well-known schemes proposed in recent ten years include PSNR-HVS-M [5], SSIM [1], visual information fidelity (VIF) [3], visual signal-to-noise ratio (VSNR) [4]), and the recently proposed most apparent distortion (MAD) [14]; the schemes based on the SSIM are reviewed in Section II.

In PSNR-HVS-M [5], MSE/PSNR in the discrete cosine transform domain is modified so that errors are weighted by the corresponding visibility threshold (which accounts for the masking effects of the HVS). However, as pointed out in [1], there is no clear psychovisual evidence that the error visibility threshold based scheme is applicable to suprathreshold distortion.

The schemes proposed in [1] and [3] are based on the high-level property of the images (e.g., structure information [1] or statistical information [3]). They have demonstrated success for images containing suprathreshold distortions [14], and as a tradeoff, these schemes generally perform less well on images containing near-threshold distortions since such schemes do not adequately account for HVS' masking property. In [1], the SSIM assumes that the HVS is highly adapted for extracting structural information from a scene, and the SSIM is measured as the correlation between the two image blocks. The VIF [3] views the IQA problem as an information fidelity problem, and the images are modeled using Gaussian scale mixtures to measure the amount of image information.

In [4], the VSNR deals with both detectability of distortions (low-level vision) and structural degradation based on the global precedence (midlevel visual property), and a

better tradeoff for the performance on near-threshold and suprathreshold distortions is achieved. The MAD proposed in [14] yields two quality scores, namely, visibility-weighted error and the differences in log–Gabor subbands statistics. The two scores are then adaptively combined to obtain the final quality score. Although it achieves good correlation with the human judgment, it has higher computational complexity.

The SSIM [1] is widely accepted due to its reasonably good evaluation accuracy [10], [11], pixelwise quality measurement, and simple mathematical formulation, which facilitates analysis and optimization. However, as pointed in some existing works [19]–[21], it is less effective for badly blurred images (actually, the SSIM is less capable of predicting the relative quality of blurred image and image with white noise; more details are to be discussed in Section III-B) since it underestimates the effect of edge damage and treats every region in an image equally. It is well known that edges are crucial for visual perception and play a major role in the recognition of image content [15], [16], [39]. In [40], it has been also demonstrated that edge information and differentiated distortion at the edges are important for perceptual quality gauging. Another example for the significance of edges comes from the fact that a mere sketch image can convey most information in the scene [15].

Therefore, in this paper, we propose an IQA scheme that is based on the edge/gradient similarity and demonstrate its effectiveness by careful analysis and extensive experiments. Similar to the SSIM, the proposed scheme considers luminance and contrast–structural changes. The main contributions of this paper in comparison to the existing work are as follows: 1) We demonstrate that gradient information that captures both contrast and structure of the image allows more emphasis on distortions around the edge regions in the proposed IQA scheme, leading to a more accurate assessment of image quality; 2) our scheme matches better with the contrast masking, particularly for the cases when both masked and masking signals are small; and 3) we devise an adaptive approach to integrate the different components (i.e., luminance and contrast–structure) of distortion.

The rest of this paper is organized as follows. In Section II, we provide a brief introduction of the SSIM and some SSIM-based schemes, including the existing gradient-based methods that are modified from the SSIM, as the ground for the following analysis, discussion, and comparison. Section III describes the details of the gradient similarity in the proposed scheme. In Section IV, we discuss how to combine the luminance and contrast–structural changes. The experimental results with further discussion are given in Section V. We will validate our IQA scheme with six IQA databases for benchmarking. Performance improvement is achieved, as compared with the state-of-the-art schemes [e.g., the SSIM, multiscale SSIM, VIF, VSNR, MAD, and information content weighted (IW)-SSIM]. Finally, conclusions are drawn in Section VI.

II. SSIM INDEX AND RELATED SCHEMES

As proposed in [1], the SSIM assumes that natural images are highly structured, and the HVS is sensitive to structural distortion. The structural information in an image is defined as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast [1].

The SSIM is calculated for each overlapped image block by using a pixel-by-pixel sliding window, and therefore, it can provide the distortion/similarity map in the pixel domain. It has been also extended using the multiscale analysis [2], complex wavelets [17], and discrete wavelets [18]. For any two image blocks x and y, the SSIM models the distortion/similarity between them as three complementary components, namely, luminance similarity, contrast similarity, and structural similarity, and these three components are mathematically described as (1)–(3) below, respectively

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{1}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{3}$$

where μ_x, μ_y, σ_x^2, σ_y^2, and σ_{xy} are the mean of x, the mean of y, the variance of x, the variance of y, and the covariance of x and y, respectively; C_1, C_2, and C_3 are claimed as small constants to avoid the denominator being zero.

The SSIM for the image blocks is given as

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

where α, β, and γ are positive constants used to adjust the relative importance of the three components. The higher the value of $\text{SSIM}(x,y)$ is, the more similar image blocks x and y are. Obviously, if y (assumed to the image block under evaluation) and x (assumed to the reference image block) are the same, then (1)–(4) will all give the value of 1, indicating no change in block y with respect to block x. The overall image quality score is determined using the mean of the local SSIM (i.e., the SSIM for each image block) [1] or calculated as the IW average of the local SSIM [32].

The schemes in [19]–[21] and [45] are also based on the SSIM and take into account the importance of edge. In these schemes, one or more components of the SSIM are changed to calculate the value in the edge domain (note that (1)–(3) are calculated in the pixel domain). For example, the structure comparison component has been changed [19], [45] to the gradient domain, or both the contrast and structure comparison components have been modified [20]. In [21] and [45], the luminance comparison component has not been included. As minor variants of the SSIM, these schemes are lack of due consideration of the HVS' masking and visibility characteristics.

III. PROPOSED GRADIENT SIMILARITY SCHEME

The block diagram of the proposed scheme is shown in Fig. 1, where the parts enclosed with dash lines are discussed here and the remaining parts are to be discussed in Section IV.

A. Gradient Similarity

The proposed gradient similarity is defined as

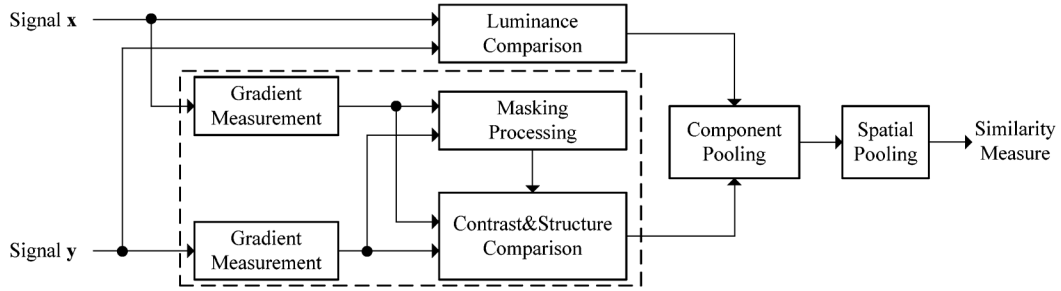$$g(x,y) = \frac{2g_xg_y + C_4}{g_x^2 + g_y^2 + C_4} \tag{5a}$$
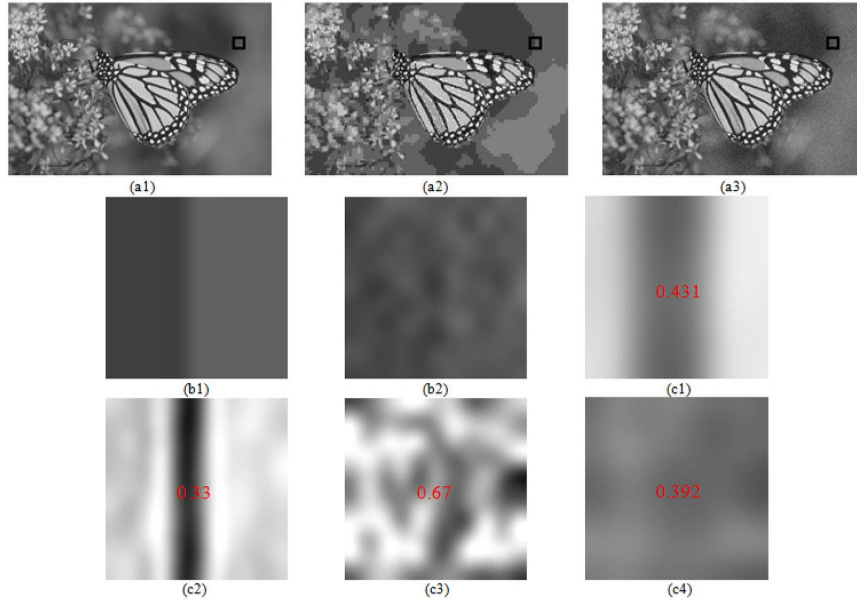
Fig. 1. Block diagram of the proposed scheme.

Fig. 2. Illustration of the difference between the SSIM and the proposed scheme: (a1) original image from LIVE database; (a2) JPEG image from LIVE database, with DMOS = 83.6, SSIM = 0.769, and $g = 0.787$; (a3) white noisy image from LIVE database, with DMOS = 51.4, SSIM = 0.674, and $g = 0.795$; (b1) and (b2) being the magnified versions of image blocks indicated by blocks in (a2) and (a3), respectively; (c1) being the SSIM map for the image blocks whose central pixels are within (b1), with the SSIM value of 0.431 for (b1); (c2) being the quality map from (5a) for the image blocks whose central pixels are within (b1), with the score from (5a) of 0.33; (c3) being the quality map from (5a) for the image blocks whose central pixels are within (b2), with the score from (5a) of 0.67 for (b2); and (c4) being the SSIM map for the image blocks whose central pixels are within (b2), with the SSIM value of 0.392 for (b2). The order of (c1)–(c4) in the figure is just arranged for easy comparison between (c1) and (c4) as well as between (c2) and (c3). Note that (c1) is almost white except for edges, and this means that the error caused by JPEG compression mainly appears in the introduced false edges; on the other hand, (c4) is with low brightness; this represents that the SSIM is quite sensitive to white noise in a smooth region (note that the image block indicated by the block in (a1) is smooth), and it is a case where SSIM overestimates the white noisy distortion.

where g_x and g_y are the gradient values for the central pixel of image blocks x and y, respectively, and C_4 is the small constant as C_2 in (2) to avoid the denominator being zero (e.g., $C_4 = 10^{-5}$); $g(x, y)$ is the gradient similarity between x and y, and its value lies in [0, 1]. The initial form of the proposed scheme in (5a) is mathematically similar to the luminance/contrast comparison term of the SSIM, and as will be demonstrated, it is more effective than that in the SSIM.

An example is shown in Fig. 2. These images have been taken from the Laboratory for Image and Video Engineering (LIVE) database [23]. In this figure, (a1) is the original image, (a2) is the JPEG compressed image, and (a3) is the images with white noise; (b1) and (b2) are the amplified versions of the blocks corresponding to the same region, as highlighted in (a2) and (a3), respectively: (b1) is an edge (to be exact, due to a false edge caused by JPEG compression) block, and (b2) is a noisy block; (c1) and (c4) show the SSIM map (darker gray levels

represent larger distortion) for the image blocks whose central pixels are within (b1) and (b2), respectively; (c2) and (c3) are the quality maps from (5a) for the image blocks whose central pixels are within (b1) and (b2), respectively.

As it is shown in Fig. 2, blocking in (a2) is much more annoying than white noise in (a3). This is also confirmed by the subjective differential mean opinion score (DMOS) values, which are 83.6 for (a2) and 51.4 for (a3) in LIVE database for the whole images. Note that a smaller DMOS indicates higher quality. However, Fig. 2(c1) and (c4) have similar brightness (i.e., the SSIM value) for the distortion regions, and this is not consistent with the HVS perception (i.e., the DMOS values). The quality values for (b1) and (b2) are given in the images for (c1) and (c4), respectively, and they are similar. On the contrary, as shown in (c2) and (c3), the proposed $g(x, y)$ can measure the relative quality loss (mainly the quality loss around edge and that in the nonedge regions) better than the SSIM,

Fig. 3. Operators for calculating the gradient value.

TABLE I
GRADIENT AND STANDARD DEVIATION FOR DIFFERENT IMAGE BLOCKS IN
Fig. 2

type	(b1)	(b2)
g_x	32	14.7
σ_x	15.42	12.08

with the evidence that the given value for block (b1) by (5a) is much lower than that for (b2). This is the reason why the quality scores in (5a) for (a2) and (a3) are more consistent with the DMOS values compared with the SSIM scores. The effectiveness of the proposed scheme will be further demonstrated through extensive experiments in Section V.

Gradient value g_x (same for g_y) is calculated as the maximum weighted average of difference for the block as [43]

$$g_x = \max_{k=1,2,3,4} \{\text{mean2}(|x \cdot M_k|)\} \qquad (6)$$

with M_k ($k = 1, 2, 3, 4$), as shown in Fig. 3, where the weighting coefficient decreases as the distance from the central pixel increases, and mean2$(\cdot)$ is the mean value for a matrix. The Sobel or Prewitt operator is not used since the kernels are too small (3×3) to include sufficient neighboring information and only two directions (horizontal and vertical) are considered. $g(x, y)$ can be interpreted either as a blockwise version (i.e., gradient similarity for image blocks x and y) or a pixelwise version (i.e., the gradient similarity for the central pixels of image blocks x and y). In the pixel version, we define

$$g(x_c, y_c) = g(x, y) \qquad (5b)$$

where x_c and y_c are the central pixels of image blocks x and y, respectively.

The proposed formulation for $g(x, y)$ is able to measure both image contrast (the degree of signal variation) change and image structure (structure of objects in the scene) change since the gradient value (i.e., g_x and g_y) is a contrast-and-structure variant feature. To illustrate this, first, consider $y = \delta \cdot x$ with δ is a constant and $\delta \neq 1$, and then we have $g_y = \delta \cdot g_x$; therefore, $g_y \neq g_x$ since $\delta \neq 1$, and it is shown that g_x is a contrast variant feature.

To further demonstrate that g_x is a structure variant feature, we have shown in Table I the gradient values for (b1) and (b2) in Fig. 2. The standard deviation values (i.e., σ_x, as used in the SSIM to measure the contrast change) for the image blocks are also given for a comparison. In Table I, we can see that the values of σ_x for (b1) and (b2) are very close; however, (b1) and (b2) have very different structures, i.e., one is edge and the other

is weak texture. As expected, the gradient values for (b1) and (b2) are different. Although both standard deviation and gradient can be used to measure the contrast, their difference is as follows. Given a group of pixel values, the standard deviation for the pixels is a constant, no matter how these pixels are positioned (i.e., independent of pixel positioning), whereas the gradient values for the same group of pixels change according to the positioning of these pixels.

Note that $g(x, y)$ can be rewritten as

$$
g(x, y) = \frac{\max(g_x, g_y)^2 \cdot \left[2\frac{\min(g_x,g_y)}{\max(g_x,g_y)} + \frac{C_4}{\max(g_x,g_y)^2}\right]}{\max(g_x, g_y)^2 \cdot \left[1 + \left(\frac{\min(g_x,g_y)}{\max(g_x,g_y)}\right)^2 + \frac{C_4}{\max(g_x,g_y)^2}\right]}
$$

$$
= \frac{2\frac{\min(g_x,g_y)}{\max(g_x,g_y)} + K}{1 + \left(\frac{\min(g_x,g_y)}{\max(g_x,g_y)}\right)^2 + K}
$$

$$
= \frac{2\left[1 - \frac{|g_x - g_y|}{\max(g_x,g_y)}\right] + K}{1 + \left(1 - \frac{|g_x - g_y|}{\max(g_x,g_y)}\right)^2 + K}
$$

$$
= \frac{2(1 - R) + K}{1 + (1 - R)^2 + K} \qquad (7)
$$

with the masked gradient change defined as

$$
R = \frac{|g_x - g_y|}{\max(g_x, g_y)} \qquad (8)
$$

where $K = C_4 / \max(g_x, g_y)^2$, and it is close to zero (when C_4 is small enough, e.g., $C_4 = 10^{-5}$ as aforementioned); R is the gradient change (i.e., $|g_x - g_y|$) relative to the masking gradient (note that either g_x or g_y is a masking signal, and here, we use the larger one, i.e., $\max(g_x, g_y)$ as the masking gradient) and lies in the range of [0, 1]. As it is shown in (8), $R = 0$ when $g_x = g_y$, i.e., the gradient of test image block y is the same as that of reference image block x; $R = 1$ when either g_x or g_y is zero (therefore, $|g_x - g_y| = \max(g_x, g_y)$), i.e., when either false gradient is created from an originally smooth block or an originally nonsmooth block is changed into a complete smooth one.

It is revealed in (7) and (8) that, with the same amount of gradient change, the proposed $g(x, y)$ is less sensitive to the case of higher masking contrast than that of lower masking contrast, and this is consistent with the contrast masking of the HVS for high masking contrast, as will be further discussed in Section III-C.

B. Further Analysis for the Proposed Scheme and SSIM

Consider Fig. 4 in which the x-axis represents the predicted value from the scheme under consideration and the y-axis represents the subjective DMOS. We show the five-parameter logistic mapping curve between the objective outputs and the subjective DMOS [the logistic mapping is to be discussed in Section V-A as (13)] as the middle curve. The upper curve denotes $T\%$ (here, $T = 20$ as an example in the figure) the upper bound (i.e., with DMOS value being $T\%$ higher than the middle curve); the lower curve denotes the lower bound (i.e., with DMOS value being $T\%$ lower than the middle curve). In other words, the DMOS value of the images corresponding to the data points lying between the upper and lower curves can be "correctly" predicted
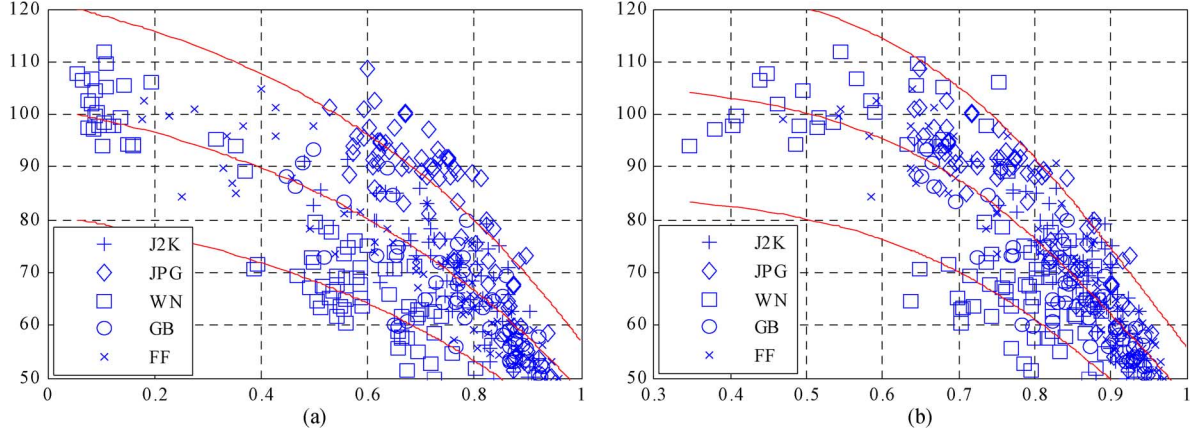
Fig. 4. Predicted value from schemes under consideration (x-axis) and the subjective DMOS (y-axis; with DMOS > 50) for LIVE database. (a) DMOS versus the SSIM value. (b) DMOS versus the score from (5a).

by the IQA scheme if $T\%$ measurement error is allowed; the DMOS value of the remaining images cannot be "correctly" predicted from the IQA scheme. The more the data points are between the bounding curves, the better the scheme is. If a data point lies below the lower bound, it means that the visual distortion is overestimated [such as Fig. 2(c4)] since the scheme predicts a higher score (i.e., lower image quality) than the actual DMOS value; the visual distortion is underestimated for the data points lying above the upper-bound curve.

Fig. 4(a) shows the plot of DMOS versus SSIM for high-distortion images (i.e., with DMOS values bigger than 50) from LIVE database, where J2K, JPG, WN, GB, and FF represent the distortion types of JPEG 2000 compression, JPEG compression, white Gaussian noise, Gaussian blur, and fast fading, respectively. In Fig. 4(a), some (more than 60 in number) data points lie outside the bounding curves; note that these data points mainly correspond to the images with WN distortion, and therefore, we have statistically demonstrated that the SSIM shows higher sensitivity to white noise, i.e., it overestimates the visual distortion with WN. It can be also observed in Fig. 4(a) that the SSIM tends to underestimate the visual distortion caused by JPEG compression.

The similar plot with the data given by (5a) is shown in Fig. 4(b), and only 29 data points lie outside the bounding curves. The effect of WN is also overestimated but to a lesser extent; in addition, there is no underestimation for JPG error due to the proper emphasis on the edge information.

C. Modified Gradient Similarity

Here, we examine more closely on how well (7) and (8) work. With a test image block y and its reference block x, the gradient (or contrast) of y can be regarded as the gradient of x being combined with an error signal $g_y - g_x$; alternatively, the gradient of x can be regarded as that of y being combined with an error signal $g_x - g_y$. When two visual signals combine, there is masking of one signal over the other, and a visibility threshold [12], [14], [22] can be determined for the masked contrast against the masking contrast.

In general, the visibility threshold C_{Tr} (below which the change is not noticeable to the HVS) for the masked signal

increases with the masking contrast [12], [14], [22]. Masked gradient change R defined in (8) is a largely reasonable choice since, when the masking signal is higher (i.e., $\max(g_x, g_y)$ is higher), the perceptual effect of the masked signal is lower for a same amount of $|g_x - g_y|$ (resulting in a lower R) due to higher C_{Tr}. As discussed in Section III-A, the lower value of R, the higher value for $g(x, y)$ (therefore, the higher predicted quality of test block y).

The only problem for the formulation with (7) and (8) is the case with both the following conditions being met: (i) $|g_x - g_y|$ causes distortion that is smaller than or close to C_{Tr}; and (ii) $\max(g_x, g_y)$ is comparable to $|g_x - g_y|$ in value. With case (i), R should be 0, and therefore, $g(x, y)$ should be 1, because the difference between x and y is not visible; however, if this situation coincides with case (ii), (8) can give a value to R that is significantly bigger than 0, and this overestimates the image distortion.

As a general scheme, we need to make the necessary provision for overcoming the aforementioned overestimation. To this end, K in (7) is modified to satisfy the following conditions: (A) K does not cause too much change to $g(x, y)$ when cases (i) and (ii) do not simultaneously happen; and (B) $g(x, y)$ is close to 1 for the cases with both (i) and (ii). For (A), K should be comparable with (several times of) $2(1 - R)$ and $1 + (1 - R)^2$ in (7); for (B) K should be much larger than (e.g., more than 20 times of) $2(1 - R)$ and $1 + (1 - R)^2$. Note that both $2(1 - R)$ and $1 + (1 - R)^2$ lie in the range of $[0, 2]$, and therefore, a reasonable K value is $K < 20$ for (A) and $K > 40$ for (B). Obviously, K must be adaptive to $\max(g_x, g_y)$.

In this paper, we have chosen K as $K = K'/\max(g_x, g_y)$, where K' is a positive constant and called as a masking parameter. Typically, $\max(g_x, g_y)$ is around 50 and 5 for (A) and (B), respectively; hence, the choice of K' is between 200 and 1000 (we choose $K' = 200$ in the current stage of the work). Substituting the fine-tuned K into (7), we get

$$g(x, y) = \frac{2(1 - R) + K'/\max(g_x, g_y)}{1 + (1 - R)^2 + K'/\max(g_x, g_y)} \qquad (9)$$

where $g(x, y)$ in (9) is still in the range of $[0, 1]$.

To illustrate the effect of fine tuning for K, Fig. 5(a) and (b) shows two simple image blocks that have gradient values of 1

200	201	200	200	200
200	201	200	200	200
200	201	200	200	200
200	201	200	200	200
200	201	200	200	200

200	204	200	200	200
200	204	200	200	200
200	204	200	200	200
200	204	200	200	200
200	204	200	200	200

(a)   (b)

Fig. 5.   Simple example to demonstrate the benefit of the modification for K.

and 4, respectively; the visibility threshold C_{Tr} for (a) is 4.7 using the method in [12]. In the figure, we can see that (a) and (b) are very similar, and this represents a case for condition (B). However, the quality score given by (5a) is 0.47 (which means that (a) and (b) are not similar). On the contrary, the score given by (9) is 0.989, which is more reasonable.

IV. INTEGRATION FOR OVERALL IMAGE QUALITY

In the previous section, we have proposed the gradient similarity-based scheme considering better (in comparison with the SSIM) masking effect and visibility threshold. For the parts within the dash box in the block diagram in Fig. 1, the "Masking Processing" module is used to determine the value of K, and the "Contrast&Structure Comparison" module is used to get the value of $g(x, y)$ as in (5a); (9) is the combination of these two modules. In (9), the gradient information of the reference and distorted images is used for evaluating the contrast and structural similarity. We also need to account for luminance similarity. In the rest of this section, we discuss how to tackle the luminance similarity and the integration of such feature with the gradient similarity measurement.

A. Measurement for Luminance Distortion

In addition to the contrast/structure changes, luminance changes can also cause the visible distortion although they are not as annoying as contrast/structure changes. However, (9) cannot account for the luminance change/distortion since it uses only the gradient information as the input and the gradient information is not affected by the noncontrast/structure changes. For example, the gradient information is the same if a constant value is added or subtracted from the image. Therefore, we have used the squared pixel error to measure the luminance changes in this paper after normalization by the dynamic range of the pixel values; luminance similarity can be defined as

$$e(x_i, y_i) = 1 - \left(\frac{x_i - y_i}{L}\right)^2 \qquad (10)$$

where x_i and y_i are the pixels at position i in image blocks x and y, respectively, and L is the dynamic range of the pixel values (255 for 8-bit grayscale images); $e(x_i, y_i)$ is the luminance similarity between image pixels x_i and y_i and also with the range of [0, 1].

B. Adaptive Distortion Integration

The existing gradient-based schemes either fail to account for luminance distortion [21], [45] or use simple multiplication for

the integration of the contrast–structure and luminance distortions [19], [20]. However, such integration method for luminance distortion ignores the fact that contrast–structure distortion has a higher impact on visual quality than luminance distortion, and therefore, it has degraded the overall performance of the scheme as demonstrated in [7].

A general form of integration for gradient and luminance similarities to derive the overall quality indicator $q(x_i, y_i)$ for image pixel pair x_i and y_i can be given as

$$q = (1 - W(g, e)) \cdot g + W(g, e) \cdot e \qquad (11)$$

where q, g, and e are the abbreviated forms of $q(x_i, y_i)$, $g(x_i, y_i)$, and $e(x_i, y_i)$, respectively; $W(g, e)$ is the weighting function used to adjust the relative importance of the two components. $W(g, e)$ should be significantly smaller than 0.5 since, as pointed out earlier, contrast–structural distortion (i.e., g) plays a much more significant role toward final quality evaluation than the luminance distortion (i.e., e).

In addition, the simultaneous existence of multiple distortion components in a neighborhood will mask the perception of each other. Since g is more important as aforementioned and e is more likely to be masked by the contrast–structural distortion if the contrast–structural distortion is high, weight $W(g, e)$ should be adaptive with respect to g (i.e., $W(g, e)$ should be small when g is small), i.e., we define

$$W(g, e) = p \cdot g \qquad (12)$$

where p is a positive weighting parameter. Since g is in the range of [0, 1], p also has to be significantly smaller than 0.5. We chose p to be 0.1 in this paper in order to balance the following: 1) the need of accounting for luminance change for a general solution; and 2) the fact that luminance change contributes less than the structural one. We will analyze the impact of this parameter choice in Section V-E, with data from different images and distortion types, in various databases publicly available.

V. OVERALL EXPERIMENTS AND RESULTS

Here, we provide extensive experimental results to evaluate the overall accuracy, robustness, and efficiency with six publicly available subject-rated benchmark IQA databases. We also compare the performance of the proposed scheme [q as defined in (11)] with the state-of-the-art IQA schemes, including the SSIM [1], multiscale SSIM [2], VIF [3], VSNR [4], MAD [14], and IW-SSIM [32]; and the results for PSNR and PSNR-HVS-M [5] are also included for comparison.

We first introduce the databases and the evaluation criteria adopted and then present the experimental results. We will also discuss the computational complexity of the proposed scheme.

A. Databases and Evaluation Criteria

Six publicly available and subject-rated image databases in the IQA community are used, namely, LIVE [23], Tampere Image Database (TID) [24], [25], Toyama [26], A57 [27], IVC [28], [29], and CSIQ [14], [30]. The number of distorted images is 779 for LIVE, 1700 for TID2008, 168 for Toyama, 54 for A57, 185 for IVC, and 866 for CSIQ, and these images and
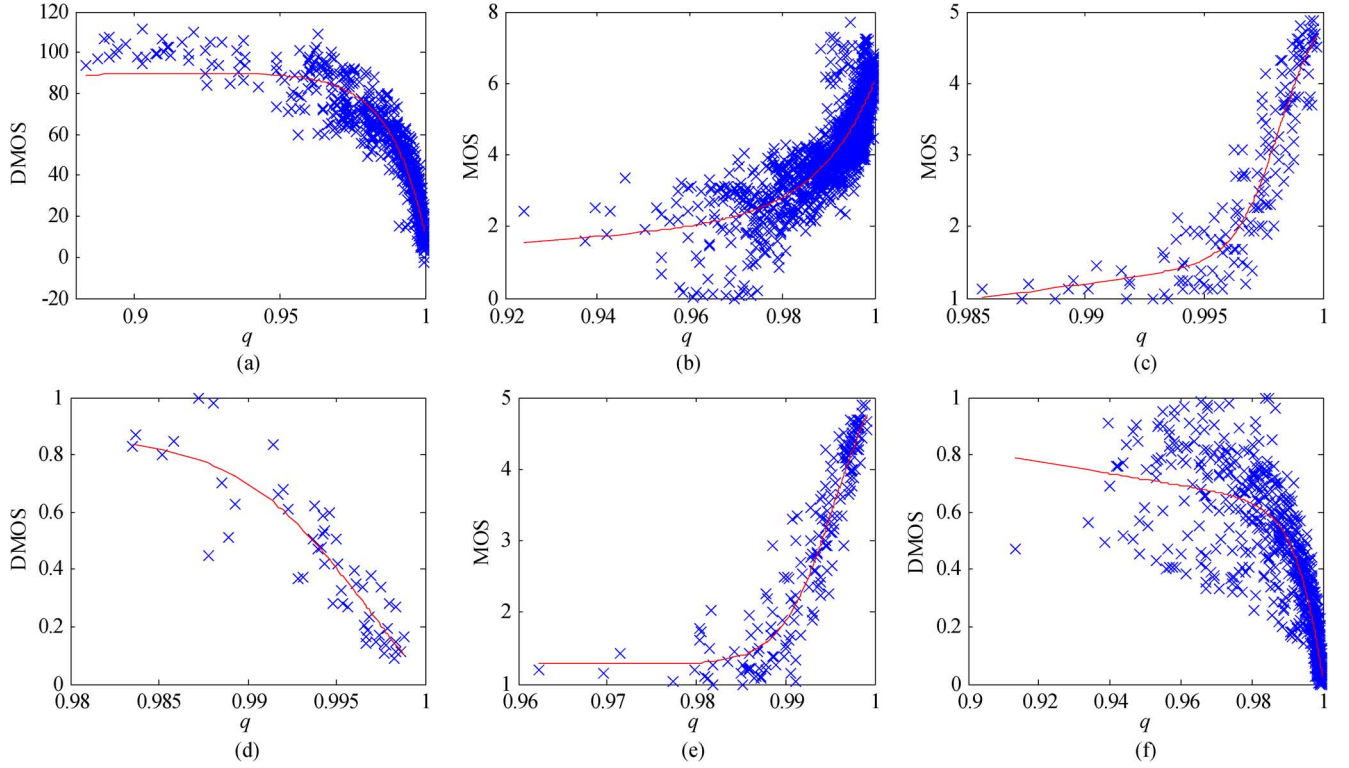
Fig. 6.   Scatter plots of subjective scores versus scores from the proposed scheme q on IQA databases: (a) LIVE; (b) TID; (c) Toyama; (d) A57; (e) IVC; and (f) CSIQ.

their corresponding subjective ratings are used as the ground truths to be compared against the IQA scheme outputs.

A five-parameter logistic mapping between the objective outputs s_o and the subjective scores is employed following the Video Quality Experts Group Phase-I/II tests and validation method [33] to remove any nonlinearity due to the subjective rating process and to facilitate the comparison of the schemes in a common analysis space. The used logistic function has the following form:

$$s_m = f(s_o)$$
$$= \beta_1 \cdot \left(\frac{1}{2} - \frac{1}{1 + \exp\left(\beta_2 \cdot (s_o - \beta_3)\right)} \right) + \beta_4 \cdot s_o + \beta_5 \tag{13}$$

where $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ are the parameters to be fitted by minimizing the sum of squared differences between the mapped values s_m and subjective scores. The subjective scores are the mean opinion score (MOS) or DMOS between the reference and distorted images that have been provided by the IQA databases.

We used four evaluation criteria to compare the performance of the IQA schemes, i.e., 1) Spearman rank-order correlation coefficient (SROCC), 2) Kendall rank-order correlation coefficient (KROCC), 3) Pearson linear correlation coefficient (CC), and 4) root-mean-squared error (RMSE), between the objective scores after nonlinear regression (i.e., s_m) and the subject scores. The mathematical definition of these four criteria can be found in [32]. Among these four criteria, SROCC and KROCC are employed to assess prediction monotonicity [33]. Note that these operate only on the rank of the data points (i.e., they are

independent of any monotonic nonlinear mapping) and ignore the relative distance between data points. CC and RMSE are used to evaluate prediction accuracy [33]. A better objective IQA measure has a higher CC, SROCC, and KROCC, but lower RMSE values; for a perfect match between the mapped objective scores (i.e., s_m) and the subjective scores, CC = SROCC = KROCC = 1 and RMSE = 0. More details, comparison, and discussion of these performance measures can be found in [34] and [35].

B. Overall Accuracy and Monotonicity Evaluation

Fig. 6 shows the scatter plots of the proposed IQA scheme on the six databases. The values of SROCC, KROCC, CC, and RMSE are listed in Table II, and Table III gives the 95% confidential interval (CI) of the CC values in Table II. In Table II, we can see that the proposed scheme performs consistently well across all the databases. Although the performances of the proposed scheme and multiscale SSIM and MAD are similar across databases, the proposed scheme has a clear advantage in terms of computational effort and time required (execution time will be compared in Section V-D); moreover, the proposed scheme can provide a quality map, whereas multiscale SSIM and MAD can only provide an overall quality score. With regard to IW-SSIM, it is a good IQA scheme because it basically weighs SSIM (which itself is a reasonably effective quality measure) values using weights derived via image information content [32]. Similar weighting can be employed for the proposed scheme to further improve its performance as the next step of the work since the weighting strategy is general to different schemes. Benchmarked with the

TABLE II
PERFORMANCE COMPARISON FOR IQA SCHEMES ON SIX DATABASES

database	criteria	PSNR	SSIM	MS-SSIM	VSNR	VIF	PSNR-HVS-M	Proposed	MAD	IW-SSIM
LIVE	SROCC	0.8756	0.9479	0.9513	0.9280	0.9632	0.9295	0.9554	0.9669	0.9567
	KROCC	0.6865	0.7963	0.8044	0.7625	0.8270	0.7659	0.8131	0.8421	0.8175
	CC	0.8723	0.9449	0.9409	0.9237	0.9598	0.9251	0.9437	0.9674	0.9522
	RMSE	13.3597	8.9454	9.2593	10.4694	7.6670	10.3722	9.0376	6.9235	8.347
TID	SROCC	0.5794	0.7749	0.8542	0.7049	0.7496	0.5951	0.8554	0.8340	0.8559
	KROCC	0.4210	0.5768	0.6568	0.5345	0.5863	0.4692	0.6651	0.6445	0.6636
	CC	0.5726	0.7710	0.8451	0.6823	0.8090	0.5784	0.8462	0.8306	0.8579
	RMSE	1.1003	0.8546	0.7173	0.9810	0.7888	1.0947	0.7151	0.7474	0.6895
Toyama	SROCC	0.6132	0.8794	0.8874	0.8608	0.9077	0.8480	0.9233	0.9362	0.9202
	KROCC	0.4443	0.6939	0.7029	0.6745	0.7315	0.6568	0.7541	0.7823	0.7537
	CC	0.6353	0.8876	0.8922	0.8704	0.9138	0.8580	0.9287	0.9405	0.9248
	RMSE	0.9664	0.5764	0.5652	0.6160	0.5084	0.6428	0.4640	0.4251	0.4761
A57	SROCC	0.6189	0.8066	0.8414	0.9355	0.6223	0.8962	0.9002	0.9023	0.8709
	KROCC	0.4309	0.6058	0.6478	0.8031	0.4589	0.7261	0.7205	0.7233	0.6842
	CC	0.6347	0.8014	0.8575	0.9497	0.6157	0.8749	0.8976	0.9059	0.9034
	RMSE	0.1899	0.1470	0.1264	0.0769	0.1937	0.1190	0.1084	0.1041	0.105
IVC	SROCC	0.6884	0.9018	0.8980	0.7993	0.8964	0.8832	0.9294	0.9146	0.9125
	KROCC	0.5218	0.7223	0.7203	0.6053	0.7158	0.6935	0.7626	0.7406	0.7339
	CC	0.7196	0.9119	0.9108	0.8034	0.9028	0.8905	0.9399	0.9210	0.9231
	RMSE	0.8460	0.4999	0.5029	0.7255	0.5239	0.5544	0.4160	0.4747	0.4686
CSIQ	SROCC	0.8005	0.8756	0.9133	0.8104	0.9195	0.8179	0.9126	0.9467	0.9213
	KROCC	0.5984	0.6907	0.7393	0.6237	0.7537	0.6430	0.7403	0.7970	0.7529
	CC	0.7998	0.8612	0.8990	0.7993	0.9277	0.8137	0.8979	0.9502	0.9144
	RMSE	0.1576	0.1334	0.1150	0.1578	0.0980	0.1526	0.1156	0.0818	0.1063

TABLE III
CI FOR IQA SCHEMES ON SIX DATABASES

Scheme	LIVE	TID	Toyama	A57	IVC	CSIQ
PSNR	[0.8544, 0.8881]	[0.5398, 0.6037]	[0.5354, 0.7177]	[0.4420, 0.7713]	[0.6420, 0.7827]	[0.7744, 0.8226]
SSIM	[0.9368, 0.9519]	[0.7510, 0.7896]	[0.8505, 0.9159]	[0.6795, 0.8803]	[0.8840, 0.9334]	[0.8430, 0.8775]
MS-SSIM	[0.9322, 0.9484]	[0.8310, 0.8582]	[0.8565, 0.9194]	[0.7655, 0.9151]	[0.8826, 0.9326]	[0.8854, 0.9110]
VSNR	[0.9126, 0.9334]	[0.6560, 0.7069]	[0.8282, 0.9029]	[0.9145, 0.9706]	[0.7455, 0.8492]	[0.7738, 0.8221]
VIF	[0.9539, 0.9650]	[0.7919, 0.8248]	[0.8848, 0.9357]	[0.4166, 0.7584]	[0.8721, 0.9264]	[0.9178, 0.9364]
PSNR_HVS_M	[0.9143, 0.9346]	[0.5458, 0.6092]	[0.8121, 0.8934]	[0.7928, 0.9258]	[0.8562, 0.9169]	[0.7899, 0.8351]
Proposed	[0.9355, 0.9509]	[0.8321, 0.8592]	[0.9045, 0.9470]	[0.8290, 0.9395]	[0.9204, 0.9547]	[0.8841, 0.9100]
MAD	[0.9625, 0.9716]	[0.8152, 0.8447]	[0.9202, 0.9558]	[0.8426, 0.9446]	[0.8957, 0.9403]	[0.9433, 0.9563]
IW-SSIM	[0.9452, 0.9583]	[0.8448, 0.8700]	[0.8993, 0.9440]	[0.8385, 0.9430]	[0.8985, 0.9419]	[0.9028, 0.9247]

TABLE IV
AVERAGE PERFORMANCE OVER SIX DATABASES

database	criteria	PSNR	SSIM	MS-SSIM	VSNR	VIF	PSNR-HVS-M	Proposed	MAD	IW-SSIM
Direct Average	SROCC	0.6960	0.8644	0.8909	0.8398	0.8431	0.8283	0.9127	0.9168	0.9063
	KROCC	0.5171	0.6810	0.7119	0.6673	0.6789	0.6591	0.7426	0.7550	0.7343
	CC	0.7057	0.8630	0.8909	0.8381	0.8548	0.8234	0.9090	0.9193	0.9126
Weighted Average	SROCC	0.6994	0.8455	0.8914	0.7905	0.8456	0.7458	0.8967	0.8971	0.8978
	KROCC	0.5232	0.6615	0.7116	0.6160	0.6860	0.5941	0.7228	0.7328	0.7240
	CC	0.6982	0.8405	0.8830	0.7777	0.8742	0.7369	0.8874	0.8970	0.8974

remaining schemes, Table III shows that the proposed scheme significantly outperforms (i.e., the lower bound of CI for the proposed scheme is larger than the upper bound of CI for the scheme in comparison) on at least one database (e.g., LIVE for PSNR, VSNR, and PSNR-HVS-M; TID for SSIM and VIF).

To provide an overall indication of the comparative performance of the different schemes, Table IV gives the average SROCC, KROCC, and CC results over six databases, where the average values are computed in two cases, in the same way as done in [32]. In the first case, the correlation scores are directly averaged, whereas in the second case, different weights are as-

signed to the databases depending on the number of distorted images in each database (refer to Section V-A for such numbers). In Table IV, we can see that the proposed scheme performs the best or close to the best on average no matter what kind of averaging is used and what the evaluation criterion is.

C. Robustness Evaluation

The robustness of the proposed scheme on different databases is also demonstrated in Table II. Note that the SSIM performs well on IVC database but not on A57 database; VIF is not good on A57 although it is the best for LIVE. Likewise, VSNR is the

TABLE V
SROCC COMPARISONS FOR INDIVIDUAL DISTORTION TYPES

Distortion type	PSNR	SSIM	MS-SSIM	VSNR	VIF	PSNR-HVS-M	Proposed	MAD	IW-SSIM
WN	**0.9070**	0.8107	0.8086	0.7734	0.8799	**0.9110**	0.8577	0.8388	0.8028
WN for color	**0.8995**	0.8029	0.8054	0.7798	**0.8785**	0.8479	0.8091	0.8258	0.8015
Spatially Correlated noise	**0.9170**	0.8145	0.8209	0.7673	0.8703	**0.9262**	0.8907	0.8678	0.7909
Masked noise	**0.8515**	0.7795	0.8107	0.7348	**0.8698**	0.7555	0.7409	0.7336	0.8068
High frequency noise	**0.9270**	0.8729	0.8694	0.8822	0.9075	**0.9136**	0.8936	0.8864	0.8732
Impulse noise	**0.8724**	0.6732	0.6907	0.6438	0.8331	**0.8339**	0.7229	0.6499	0.6579
Quantization noise	0.8696	0.8531	0.8589	0.8255	0.7956	**0.9007**	**0.8752**	0.8160	0.8182
GB	0.8697	0.9544	0.9563	0.9323	0.9546	0.9219	**0.9589**	0.9197	**0.9580**
Image denoising	0.9416	0.9530	**0.9582**	0.9290	0.9189	0.9572	**0.9724**	0.9434	0.9463
JPG	0.8717	0.9252	0.9322	0.9181	0.9170	**0.9338**	**0.9392**	0.9275	0.9181
J2K	0.8132	0.9625	0.9700	0.9517	0.9713	**0.9762**	**0.9759**	0.9707	0.9749
JPG transmission errors	0.7516	0.8678	**0.8681**	0.8052	0.8582	0.8105	**0.8835**	0.8661	0.8560
J2K transmission errors	0.8309	0.8577	0.8606	0.7914	0.8510	**0.9321**	**0.8925**	0.8394	0.8313
Non eccentricity noise	0.5815	0.7107	0.7377	0.5723	**0.7608**	0.6812	0.7372	**0.8287**	0.7719
Block-wise distortion	0.6193	**0.8462**	0.7546	0.1952	0.8320	0.6238	**0.8862**	0.7970	0.7889
Mean shift	0.6957	**0.7231**	**0.7338**	0.3885	0.5132	0.6932	0.7170	0.5161	0.6757
Contrast change	0.5859	0.5246	0.6381	0.4250	**0.8190**	0.5815	**0.6737**	0.2723	0.6273

TABLE VI
EXECUTION TIME (IN SECONDS PER IMAGE) FOR DIFFERENT SCHEMES

Schemes	PSNR	SSIM	MS-SSIM	VSNR	VIF	PSNR-HVS-M	Proposed	MAD	IW-SSIM
Time	0.0037	0.0454	0.1673	0.4452	3.4829	2.5586	0.0873	34.3821	0.5762

best for A57 but is relatively poor on TID database. Overall, MAD, IW-SSIM, and the proposed scheme give more consistent and stable performances across all the six databases in comparison with the other schemes. Among the three schemes, MAD performs slightly less well on TID; IW-SSIM performs slightly less well on A57, whereas the proposed one performs slightly less well on CSIQ.

To further examine the robustness of the IQA schemes, the performance on each distortion type in TID database is shown in Table V, where the two best IQA schemes have been highlighted in boldface for each distortion type. The TID database is used since it is the largest database and contains the most distortion types, which are listed in the first column of Table V. We include only the SROCC values since other performance criteria lead to similar conclusions. From the table, we can see that, for noise-contaminated images, traditional PSNR and its variant (i.e., PSNR-HVS-M) are still the best IQA schemes; the proposed scheme, the SSIM, multiscale SSIM, and IW-SSIM have similar performance. For other distortion types, the proposed scheme performs quite well (i.e., the best or the second best). Such results are expected since, for the cases where noise would not severely damage the edge or create false edge, the proposed scheme emphasizing edge information maintains the performance; when the distortion causes error both around edges and nonedges (e.g., blockwise distortion), the performance of the proposed scheme is good. It is also noted that the MAD has some difficulty to predict the quality of the images with distortions caused by mean value shift or contrast change.

D. Efficiency Evaluation

To compare the efficiency (i.e., computational complexity) of different schemes, we measured the average execution time required for per image (512 × 512 in size) in A57 database on a PC with 2.40-GHz Intel Core 2 CPU and 2 GB of RAM.

Table VI shows the required time in seconds per image, with all the codes being implemented with MATLAB. The codes for the SSIM, multiscale SSIM, VSNR, VIF, PSNR-HVS-M, MAD, and IW-SSIM are from [23], [36], [36]–[38], [46], and [47], respectively. It is shown in Table VI that the proposed scheme takes more time than the PSNR and the SSIM only, and it is faster than the multiscale SSIM since no multiscale image decomposition is involved in the proposed scheme. MAD and IW-SSIM also take much longer processing time than the proposed scheme. To be more precise, the proposed method takes only about 0.25% and 15.2% of the time taken by MAD and IW-SSIM, respectively.

E. Impact of the Parameter Values

The impact of the parameter values is shown with the plots of SROCC as a function of the parameter values. In the proposed scheme, there are two parameters, i.e., K' and p, where K' is for the masking effect (its valuation has been determined in Section III-C) and p is used to integrate contrast–structure and luminance distortions (as discussed in Section IV-B).

To confirm that the chosen K' is a reasonable choice, Fig. 7 plots the SROCC as a function of K' for all the databases. As can be observed, when the value of K' is large (e.g., $K' \geq 200$), the value for the SROCC is nearly constant no matter what the value of K' is used, the difference of SROCC values for different values of K' is smaller than 0.01, and this is valid for every database. Although the optimal value of K' to achieve the highest SROCC value is different for different databases (i.e., 15 for Toyama; 40 for LIVE, IVC, and CSIQ; and 1000 for TID and A57), there are two common trends for different databases, as shown in these plots: a) SROCC drops fast when the value of K' close to 0, and this is expected since, for a small K', the contrast masking effect for the case the masking signal is small, is not properly reflected; and (b) when $K' > 200$, the SROCC is
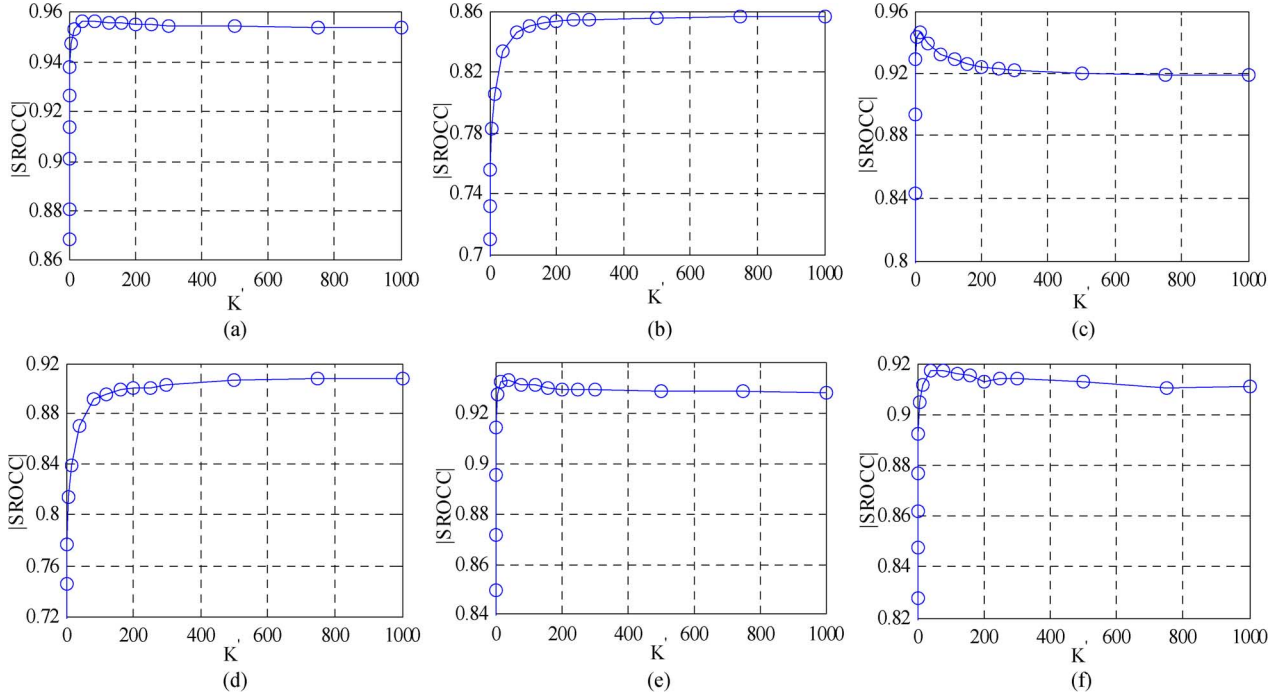
Fig. 7. Plot of $|SROCC|$ as a function of K' for (a) LIVE, (b) TID, (c) Toyama, (d) A57, (e) IVC, and (f) CSIQ databases.
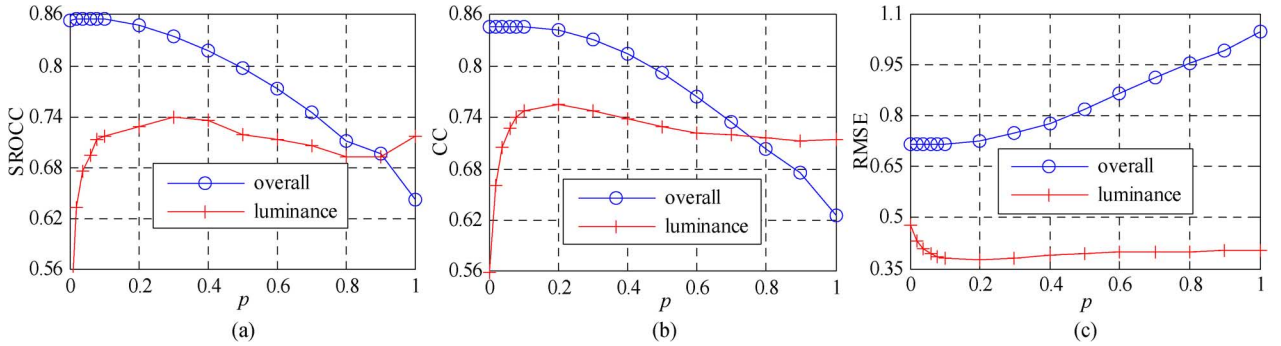

Fig. 8. Plot of (a) SROCC, (b) CC, and (c) RMSE as a function of P for the proposed integration approach and for the TID data set [25].

very close to (i.e., no smaller than 97.7% of) the highest possible value when the optimal value for K' is used. Therefore, the choice of $K' = 200$ is reasonable for our purpose.

To show the impact of p for (12), we simulated various values of p on TID database [25] first since it includes luminance distortion (mean shift distortion, with 25 reference images and 4 distortion levels for each reference image). In Fig. 8, we show the absolute values of the SROCC, CC, and RMSE plotted as a function of p for the overall performance (i.e., the images with all the 17 distortion types) and the performance on luminance distortion (i.e., the images with mean shift distortion). We found that $p = 0.1$ to yield good tradeoff between performance on luminance distortion and for the overall cases (i.e., the SROCC value for $p = 0.1$ is over 99% of the highest SROCC value, among all the possible SROCC values when different values of p are used). In the figure, we can see that the overall performance decreases with the increase in p (i.e., increase in the contribution of e in q) since e is less capable of measuring the contrast and structural distortions compared to g. On the other

hand, p cannot be zero since that will mean the luminance distortion is completely ignored. Therefore, we use $p = 0.1$ since it gives good performance both for luminance distortion and for the overall case.

In addition, Fig. 9 shows the SROCC values as a function of p for the other databases. In the figure, we observe the following: a) for four databases, SROCC drops with the increase in p; this is expected since contrast–structural distortion is the dominant factor in IQA, and according to (11) and (12), the larger the value of p, the larger the impact of the luminance distortion measurement; (b) when the value of p is in the range of [0, 0.2], the values of the SROCC are nearly constant no matter what the value of p is used (the difference of SROCC values for different values of p is smaller than 0.01), and the values for SROCC are very close to (i.e., no smaller than 99% of) the highest possible value. Therefore, a value of p in the range of [0, 0.2] is a good database-independent choice for the overall performance. However, if the value of p is too small (i.e., $p < 0.1$), the performance on luminance distortion would degrade too much, as
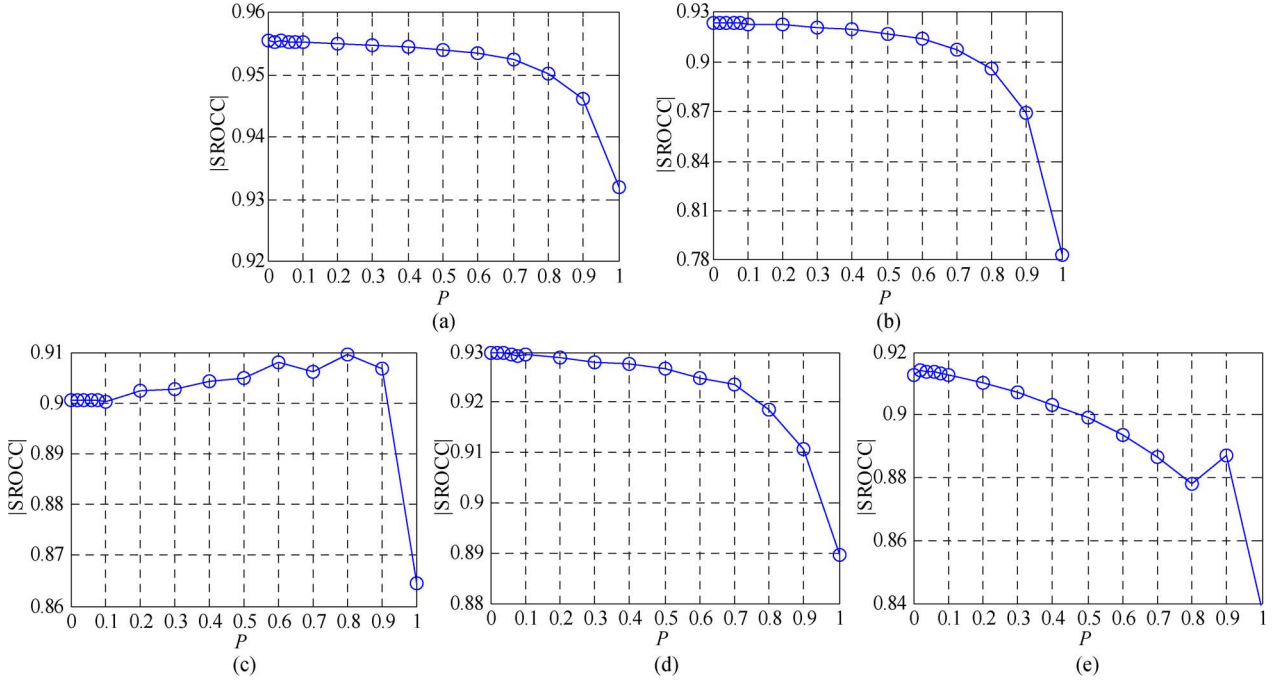
Fig. 9.   Plot of $|SROCC|$ as a function of p for (a) LIVE, (b) Toyama, (c) A57, (d) IVC, and (e) CSIQ databases.

TABLE VII
SROCC COMPARISONS FOR EACH COMPONENT OF THE PROPOSED SCHEME

Database	q	g	g_m	g_g	$\Delta(g\_m)$	$\Delta(g\_g)$	Database	q	g	g_m	g_g	$\Delta(g\_m)$	$\Delta(g\_g)$
LIVE	0.9554	0.9555	0.8689	0.9057	0.0866	0.0498	A57	0.9002	0.9004	0.5624	0.7742	0.3380	0.1262
TID	0.8554	0.8532	0.6592	0.6231	0.1940	0.2301	IVC	0.9294	0.9297	0.8002	0.7725	0.1296	0.1572
Toyama	0.9233	0.9241	0.6799	0.7135	0.2442	0.2106	CSIQ	0.9126	0.9126	0.8153	0.7616	0.0973	0.1510

shown and discussed in Fig. 8. Therefore, the choice of $p = 0.1$ is reasonable.

F. Impact of Each Component of the Scheme

To discuss the impact of each component in the proposed scheme toward the performance, the SROCC values for different components are given in Table VII, where q is the proposed overall scheme as in (11); g is the proposed gradient similarity scheme as in (9); g_m is the initial form of the proposed gradient similarity scheme defined in (5a), without the modification in masking as in Section III-C; g_g is the proposed scheme without gradient component [i.e., the gradient information is substituted with pixel gray level information as the input in (9)]; and $\Delta(g\_m)$ and $\Delta(g\_g)$ are the differences of g_m and g_g with g, respectively.

From the table, we can see that q and g have similar performance on each database, and this is expected since g accounts for the major effect in the evaluation. The results of $\Delta(g\_g)$ have shown the impact of the gradient component, and in the table, we can see that the incorporation of gradient component has improved about 0.05 in the SROCC value for LIVE database and 0.1–0.2 for the other databases; the possible reason for the smaller improvement on LIVE database is that this database contains a wider range of distortion strength and many IQA schemes can achieve good performance on it. It is also shown in $\Delta(g\_m)$ that the modification for masking can achieve some performance improvement on every database; the performance improvements on LIVE and CSIQ databases are the smallest

(around 0.09) since the considered masking effect has more impact on near-threshold distortion compared with the impact on suprathreshold distortion, and the two databases have fewer images (in percentage) with near-threshold distortion; for A57 database, the performance improvement is the largest (larger than 0.3) since majority of the images in A57 database have lower distortion than the lowest quality images from most other databases.

VI. CONCLUSION

Since the HVS is more sensitive to the edge regions than the nonedge ones, edge information should be carefully and explicitly incorporated in designing of an IQA scheme. We have proposed a new IQA scheme based on the concept of gradient similarity to alleviate the shortcoming of the existing relevant schemes in this regard. We have demonstrated that the proposed gradient similarity measure can be used to gauge contrast and structural changes. In addition, we have made our scheme match better with the masking effect and visibility threshold. Finally, luminance similarity is devised and incorporated to form a complete quality evaluation scheme.

The effectiveness of the proposed IQA scheme has been demonstrated with six public benchmark IQA databases. Compared with eight other representative and prominent IQA schemes, similar or better performance (i.e., more consistent with the human judgment) is achieved. Robustness across different databases and distortion types is also confirmed. In addition to its robustness and high accuracy, the proposed

scheme also provides the pixel error map and is relatively simple with its mathematical formulation and computational complexity, making it more practical to be used and embedded in various optimization processes in image processing tasks.

ACKNOWLEDGMENT

REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

[5] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proc. 3rd Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, Scottsdale, AZ, Jan. 2007.

[6] A. M. Eskicioglu, A. Gusev, and A. Shnayderman, "An SVD-based gray-scale image quality measure for local and global assessment," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 422–429, Feb. 2006.

[7] D. Rouse and S. S. Hemami, "Understanding and simplifying the structural similarity metric," in *Proc. Int. Conf. Image Process.*, 2008, pp. 325–328.

[8] Methodology for the Subjective Assessment of the Quality of Television Pictures Jun. 2002, Recommendation ITU-R BT.500-11.

[9] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Rate bounds on SSIM index of quantized images," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1624–1639, Sep. 2008.

[10] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[11] Q. Ma, L. Zhang, and B. Wang, "New strategy for image and video quality assessment," *J. Electron. Imaging*, vol. 19, no. 1, pp. 011019:1–011019:14, 2010.

[12] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.

[13] D. Kelly, "Motion and vision I: Stabilized images of stationary gratings," *J. Opt. Soc. Amer.*, vol. 69, no. 9, pp. 1266–1274, 1979.

[14] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imaging*, vol. 19, no. 1, pp. 011006:1–011006:21, 2010.

[15] X. Ran and N. Farvardin, "A perceptually motivated three-component image model-Part I: Description of the model," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 401–415, Apr. 1995.

[16] E. Ong, W. Lin, Z. Lu, S. Yao, and M. Etoh, "Visual distortion assessment with emphasis on spatially transitional regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 4, pp. 559–566, Apr. 2004.

[17] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 573–576.

[18] C. Yang, W. Gao, and L. Po, "Discrete wavelet transform-based structural similarity for image quality assessment," in *Proc. Int. Conf. Image Process.*, 2008, pp. 377–380.

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

[20] G. Chen, C. Yang, and S. Xie, "Gradient-based structural similarity for image quality assessment," in *Proc. Int. Conf. Image Process.*, 2006, pp. 2929–2932.

[21] D. Kim, H. Han, and R. Park, "Gradient information-based image quality metric," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 930–936, May 2010.

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

[23] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik, and L. K. Cormack, Image and Video Quality Assessment Research at LIVE [Online]. Available: http://live.ece.utexas.edu/research/quality/

[24] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin, "Metrics performance comparison for color image database," in *Proc. 4th Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, Scottsdale, AZ, Jan. 2009.

[25] N. Ponomarenko and K. Egiazarian, Tampere Image Database 2008 TID2008 [Online]. Available: http://www.ponomarenko.info/tid2008.htm

[26] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. Parvez Sazzad, MICT Image Quality Evaluation Database [Online]. Available: http://mict.eng.u-toyama.ac.jp/ mict/index2.html

[27] D. M. Chandler and S. S. Hemami, VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html

[28] A. Ninassi, P. Le Callet, and F. Autrusseau, "Pseudo no reference image quality metric using perceptual data hiding," in *Proc. SPIE, Human Vis. Electron. Imaging*, Jan. 2006, vol. 6057, p. 60570G.

[29] A. Ninassi, P. Le Callet, and F. Autrusseau, Subjective Quality Assessment-IVC Database [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb

[30] E. C. Larson and D. M. Chandler, Categorical Image Quality (CSIQ) Database [Online]. Available: http://vision.okstate.edu/csiq

[31] Y. H. Huang, T. S. Ou, P. Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Jul. 2010.

[32] 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.

[33] Video Quality Expert Group (VQEG), Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment II [Online]. Available: http://www.vqeg.org/ 2003

[34] 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.

[35] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[36] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, The SSIM Index for Image Quality Assessment [Online]. Available: http://ece.uwaterloo.ca/~z70wang/research/ssim/

[37] MeTriX MuX Visual Quality Assessment Package [Online]. Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

[38] PSNR-HVS-M Download Page [Online]. Available: http://www.ponomarenko.info/psnrhvsm.htm

[39] D. M. Rouse and S. S. Hemami, "Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM," in *Proc. Western New York Image Process. Workshop*, Rochester, NY, Oct. 2007.

[40] W. Lin, L. Dong, and P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 900–909, Jul. 2005.

[41] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 2002, pp. IV-3313–IV-3316.

[42] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[43] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Process., Image Commun.*, vol. 20, no. 7, pp. 662–680, Aug. 2005.

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

[45] G. Cheng, J. Huang, C. Zhu, Z. Liu, and L. Cheng, "Perceptual image quality assessment using a geometric structural distortion model," in *Proc. Int. Conf. Image Process.*, 2010, pp. 325–328.

[46] E. Larson and D. Chandler, Full-Reference Image Quality Assessment and the Role of Strategy: The Most Apparent Distortion [Online]. Available: http://vision.okstate.edu/mad/

[47] Z. Wang, IW-SSIM: Information Content Weighted Structural Similarity Index for Image Quality Assessment [Online]. Available: https://ece.uwaterloo.ca/~z70wang/research/iwssim/

Anmin Liu received the B.Sc. degree in information science and electronic engineering from Zhejiang University, Hangzhou, China, in 2007. He is currently working toward the Ph.D. degree in the School of Computer Engineering, Nanyang Technological University, Singapore.

His research interests include image/video coding, quality evaluation, and perceptual modeling and processing.

Weisi Lin (M'92–SM'98) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from King's College London, London, U.K., in 1992.

He taught and conducted research at Zhongshan University; Shantou University, China; Bath University, U.K.; National University of Singapore, Singapore; Institute of Microelectronics, Singapore; and Institute for Infocomm Research, Singapore. He has been the Project Leader of 13 major successfully delivered projects in digital multimedia technology development. He also served as the Laboratory Head of Visual Processing Laboratory and the Acting Department Manager of Media Processing Department of the Institute for Infocomm Research . He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He edited one book, authored one book and five book chapters, and has published over 190 refereed papers in international journals and conferences. He believes that good theory is practical, and so, he has kept a balance of academic research and industrial deployment throughout his working life. His areas of expertise include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.

Dr. Lin is a Fellow of the Institution of Engineering Technology and an Honorary Fellow of Singapore Institute of Engineering Technologists. He is also a Chartered Engineer (U.K.). He organized special sessions in IEEE ICME06, IEEE IMAP07, IEEE ISCAS10, PCM09, SPIE VCIP10, APSIPA11, and MobiMedia11. He gave invited/keynote/panelist talks in VPQM06, IEEE ICCCN07, SPIE VCIP10, and IEEE MMTC QoEIG (2011), as well as tutorials in PCM07, PCM09, IEEE ISCAS08, IEEE ICME09, APSIPA10, and IEEE ICIP10 . He currently serves in the editorial boards of IEEE TRANSACTIONS ON MULTIMEDIA, IEEE Signal Processing Letters, and Journal of Visual Communication and Image Representation, as well as in four IEEE Technical committees. He cochairs the IEEE Multimedia Communications Technical Committee Special Interest Group on Quality of Experience. He has been on the Technical Program committees and/or Organizing committees of a number of international conferences.

Manish Narwaria received the B.Tech. degree in electronics and communication engineering from Amrita Vishwa Vidyapeetham University, Coimbatore, India, in 2008. He is currently working toward the Ph.D. degree in the School of Computer Engineering, Nanyang Technological University, Singapore.

His research interests include image/video and speech quality assessment, pattern recognition, and machine learning.