

Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments

Songnan Li, *Student Member, IEEE*, Fan Zhang, *Member, IEEE*, Lin Ma, *Student Member, IEEE*, and King Ngan Ngan, *Fellow, IEEE*

Abstract—In the research field of image processing, mean squared error (MSE) and peak signal-to-noise ratio (PSNR) are extensively adopted as the objective visual quality metrics, mainly because of their simplicity for calculation and optimization. However, it has been well recognized that these pixel-based difference measures correlate poorly with the human perception. Inspired by existing works [1]–[3], in this paper we propose a novel algorithm which separately evaluates detail losses and additive impairments for image quality assessment. The detail loss refers to the loss of useful visual information which affects the content visibility, and the additive impairment represents the redundant visual information whose appearance in the test image will distract viewer’s attention from the useful contents causing unpleasant viewing experience. To separate detail losses and additive impairments, a wavelet-domain decoupling algorithm is developed which can be used for a host of distortion types. Two HVS characteristics, i.e., the contrast sensitivity function and the contrast masking effect, are taken into account to approximate the HVS sensitivities. We propose two simple quality measures to correlate detail losses and additive impairments with visual quality, respectively. Based on the findings in [3] that observers judge low-quality images in terms of the ability to interpret the content, the outputs of the two quality measures are adaptively combined to yield the overall quality index. By conducting experiments based on five subjectively-rated image databases, we demonstrate that the proposed metric has a better or similar performance in matching subjective ratings when compared with the state-of-the-art image quality metrics.

Index Terms—Contrast masking, contrast sensitivity function, decoupling algorithm, human visual system.

I. INTRODUCTION

RESEARCH works on visual quality assessment aim at deriving an objective visual quality metric which is consistent with the human perception. A successful objective visual quality metric can release humans from laborious works, such as visual quality monitoring in communication, visual system performance evaluation and vision-related tests in manufacturing environment, etc. Furthermore, in many

Manuscript received October 05, 2010; revised February 15, 2011 and April 21, 2011; accepted April 25, 2011. Date of publication May 10, 2011; date of current version September 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. James E. Fowler.

The authors are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: snli@ee.cuhk.edu.hk; fzhang@ee.cuhk.edu.hk; lma@ee.cuhk.edu.hk; knngan@ee.cuhk.edu.hk).

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/TMM.2011.2152382

image and video processing applications, such as compression, watermarking, restoration, color contrast enhancement, etc., it can be used online to boost algorithm performance, or to reduce computational complexity, e.g., in computer graphic illumination rendering [4].

However, there are many challenges in designing an accurate objective quality metric. Firstly, visual signals have various contents leading to different distortion masking levels. Secondly, before entering the eyes, the visual signals passed through numerous processing stages, e.g., recording, compression, transmission, post-processing, display, etc., each of which will introduce different types of artifacts. The large number of distortion types and various viewing environments make the design of a general-purpose visual quality metric difficult, if not impossible. Thirdly, after entering the eyes, the visual signals are decomposed and interpreted by the Human Visual System (HVS). Physiological knowledge on the HVS is limited to its front end. A complete understanding of the HVS and its mathematical modeling will not be realized in the near future due to the great complexity. To make matters worse, visual quality judgement is viewer dependent, related to unpredictable factors like the viewer’s interests, expectations, quality experience, etc. Therefore, it is impossible for an objective quality metric to match individual subjective ratings unconditionally. Instead, objective metric score is expected to be consistent with an average subjective score derived from dozens of subjects. The viewing conditions are standardized and the distortion types used to validate the metrics are those frequently encountered in the real practical applications.

From the viewpoint of inputs, visual quality metrics can be classified into media-layer, bitstream-layer, and packet-layer metrics. Bitstream-layer and packet-layer metrics are beyond the scope of this paper. For media-layer metrics, plenty of design approaches have been investigated in the literature. For example, pixel-based metrics, such as MSE/SNR/PSNR, etc., correlate visual quality with pixel value differences. They are easy to calculate and optimize, but do not correlate well with the human perception. Metrics, like those in [5]–[8], etc., utilize the prior knowledge on the distortion characteristics. For example, blockiness often occurs along the block boundaries; blurring affects edges most conspicuously; ringing appears prominently in the smooth regions around the edges, etc. A few metrics, as described in [9] and [10], explore the statistical characteristics of natural images to predict visual quality. Natural images possess unique statistical characteristics which will be disturbed by distortions. Deviations from the expected natural statistics are used to quantify the distortion intensity. A large number of

quality metrics are based on models of the low-level processing of the HVS. For example, vision researches suggest that in the early visual pathway, there exist several octave spacing radial frequency channels, each of which is further tuned by orientations. A host of spatial transforms are employed by quality metrics to simulate this HVS mechanism, e.g., the cortex transform used in Winkler's Perceptual Distortion Metric [11], the steerable pyramid transform used in the Teo and Heeger model [12], wavelet transform used in Lai's metric [13] and the discrete cosine transform used in Watson's DCTune [14], etc. Human perception is sensitive to luminance contrast rather than the luminance intensity. Based on this well-known fact, Peli [15] defines a local band limited contrast measure for complex images which assigns a local contrast at every point of an image and at every frequency channel. Similar contrast calculations have been adopted by Daly's Visible Different Predictor [16] and Lubin's model [17]. Contrast sensitivity function (CSF) and spatial masking are also frequently used to model the HVS processing. They are utilized by the proposed method and will be elaborated in Section III. Low-level HVS models are often criticized, e.g., they may be unsuitable for supra-threshold contrast measure and show a lack of geometric meaning. However, a recent work [18] analytically and experimentally proves that the traditional divisive normalization-based models actually are not subject to the above criticism.

High-level HVS processing mechanism (corresponding to the late visual pathway, e.g., after the primary visual cortex) still remains mysterious. Therefore, many recent IQMs simply make use of common knowledge or assumption about the high-level HVS characteristics to guide quality prediction. For example, structural information is critical for cognitive understanding; therefore, the authors of [19] make the assumption that the structure distortion is a good representative of visual quality variation. They proposed a metric which distinguishes structure distortions from luminance and contrast distortions. This assumption has been well accepted and applied in visual quality assessment [20]–[22]. Recently, authors of [23] proposed to use singular vectors out of singular value decomposition (SVD) for quantifying the structural information. Another well-known assumption made by the authors of the visual information fidelity criterion (VIF) [24] is that the HVS correlates visual quality with the mutual information between the reference and test images. The mutual information resembles the amount of useful information that can be extracted from the test image. Although VIF seems to be quite different from SSIM in terms of the fundamental assumption, down to the implementation, the two IQMs share similarities, as analyzed in [25]. In [2], the authors advocate such an assumption that the HVS adapts the quality prediction strategy to the distortion intensity. More precisely, for high-quality images, the HVS attempts to look for distortions in the presence of the image, while for low-quality images, the HVS attempts to look for image content in the presence of the distortions. Two quality measures were proposed with one more suitable for measuring high-quality images and the other for low-quality images.

In this paper, we propose a full-reference image quality metric. Instead of treating the distortions in the test image homogeneously, they are decoupled into two groups: additive

impairments and detail losses, which we believe correlate with visual quality in different ways. The proposed method is inspired by existing works [1]–[3]. In [1], the authors proposed to separate distortions into linear frequency distortions and additive noise degradations, which in essence can be interpreted as the two terms, i.e., detail losses and additive impairments, used in this paper. However, the decoupling algorithm in [1] was specifically designed for halftoning artifacts. In this paper, a general-purpose decoupling algorithm has been developed which is able to handle a host of distortion types. Furthermore, two quality measures were proposed in [1], but how to combine their outputs to yield an overall quality measure is left as an open problem. In this paper, we propose two simple quality measures, i.e., the detail loss measure and the additive impairment measure, and also develop an adaptive method to combine them together. The adaptive combination method is based on the findings in [3] that observers intend to judge low-quality images in terms of the content visibility. As mentioned above, this is similar to the assumption made in [2] that for low-quality images, the HVS attempts to look for image content in the presence of the distortions. In comparison to [2], our method explicitly extracts the detail loss to more accurately quantify the visibility of the image content, while their method achieves this by analyzing changes in local statistics (standard deviation, skewness, and kurtosis) of the log-Gabor subband coefficients.

The rest of this paper is organized as follow. The motivation of the proposed metric is further explained in Section II. Section III presents detailed implementation of the proposed metric. Experimental results are provided and discussed in Section IV. Section V contains the concluding remarks.

II. MOTIVATION

Full-reference image quality metrics take both the test image t and the original image o as inputs. As shown in Fig. 1(a), differences between t and o can be treated as the distortions. Instead of taking the pixel differences, HVS-model based metrics simulate the HVS responses to t and o and measure the response differences to evaluate visual quality. The original image can serve as the masking signal to adjust the distortions, considering the fact that distortions will be less visible in texture areas than in smooth areas. However, artifacts usually can make the test image less textured compared to the original. Therefore, utilizing the original image as the masker is sometimes problematic, especially for low-quality images where the contrasts of the textures or edges have been significantly reduced. To solve this problem, several visual quality metrics [17], [26] employ mutual masking, in which only areas that are highly textured in both t and o produce a significant masking effect.

Instead of treating the distortions indiscriminately, in this paper, we propose a novel image quality metric which separates the distortions into two parts, namely detail losses and additive impairments. As shown in Fig. 1(b), detail losses and additive impairments can be decoupled from the difference image. As its name implies, detail losses refer to the loss of the useful visual information in the test image. Most distortion types cause detail losses especially when their intensities are strong. For example in Fig. 1, the test image suffers from detail losses caused by JPEG compression. On the other hand, additive impairments

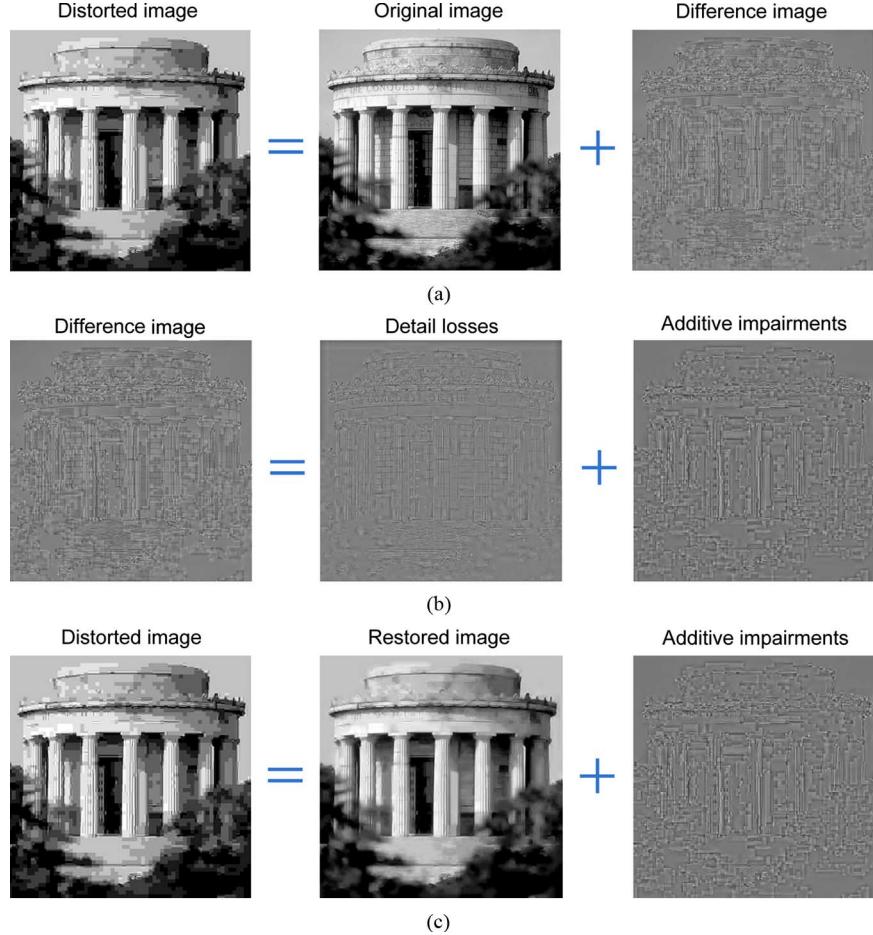


Fig. 1. Example of (a) separating the test image into the original image and the difference image (b) separating the difference image into the detail loss image and the additive impairment image (c) separating the test image into the restored image and the additive impairment image.

refer to the redundant visual information which does not exist in the original but appears in the test image only. For instance in Fig. 1, the additive impairments in the JPEG coded image appear as blocky artifacts. Various additive patterns associated with different distortion types will be illustrated in Section III. In the proposed metric, we separate the test image into a restored image and an additive impairment image, as shown in Fig. 1(c). The restored image exhibits the same amount of detail losses as the test image but is additive impairment free. It can be used as the masker to mask the additive impairments, providing a good solution to the above-mentioned masking problem. Detail losses can be generated by subtracting the restored image from the original image.

Besides the merits relating to visual masking, the necessity of decoupling detail losses and additive impairments also comes from the assumption that they correlate with the visual quality in different ways. The detail loss influences the content visibility of the test image. The upper bound of its amount is determined by the original image. Additive impairments on the other hand are less related to the original, but their appearances will distract our attentions from the original contents causing unpleasant viewing experience. In the proposed metric, we simulate the HVS responses to the restored and the original image respectively, calculate their Minkowski summations, e.g., s_r is the Minkowski sum for the restored image and s_o for the original image and use the ratio of s_r to s_o as the quality score q_1 ,

to quantify the influence of the detail loss on perceived visual quality. Different from the majority of image processing algorithms, contrast enhancement does not suppress details but enhances them. Our metric can handle this exception and it will be explained in Section III-A. We also simulate the HVS responses to the additive impairments and calculate the Minkowski summation, which is then normalized by the number of pixels to generate the additive-impairments-related quality score q_2 .

To yield the overall quality index, we propose an adaptive method to combine q_1 and q_2 together. As mentioned in Section I, the underlying assumption for this adaptive combination is adopted from [2] and [3], that is, as the image quality degrades, the HVS becomes more inclined to predict quality according to the content visibility. Since it is the detail loss score q_1 that determines the content visibility, according to this assumption, it should play a more important role in the final quality index as the distortion intensity increases. Detailed information will be given in Section III-C.

III. PROPOSED METHOD

The proposed image quality metric works with luminance only. Color inputs will be converted to gray scale before further analysis. The current algorithm is extended from our previous work [27]. Its framework is illustrated in Fig. 2, which consists of three major sections: 1) the decoupling algorithm, 2) the HVS

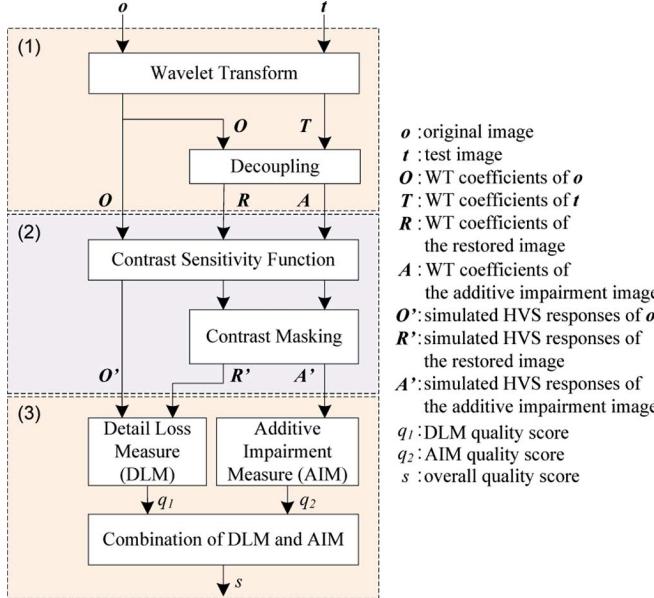


Fig. 2. Framework of the proposed full-reference image quality metric.

processing simulation, and 3) the calculation and adaptive combination of the two quality measures. Details on each section will be given below.

A. Decoupling Additive Impairments and Useful Image Contents

As illustrated in Fig. 1(c), we intend to decouple the restored image \mathbf{r} and the additive impairment image \mathbf{a} from the test image. The restored image is supposed to exhibit the same amount of detail losses as the test image but is additive impairment free, while the additive impairment image contains no original image content but only additive noises. In the literatures, a vast number of image restoration algorithms exist for various applications, e.g., de-noising, de-blocking, de-blurring, super-resolution, etc. In our metric, we do not apply the existing algorithms to decouple \mathbf{r} and \mathbf{a} , because 1) they were designed for specific applications and thus difficult to be generalized; 2) instead of aiming at an image that exhibits the same amount of detail losses as the test image, the objective of these algorithms is to recover the original image as perfectly as possible by compensating the detail losses; and 3) the problem formulation typically does not have a closed-form solution but requires iterative optimization, which makes these algorithms time consuming. In the proposed full-reference image quality metric, the availability of the original image enables us to design a general-purpose and time-efficient algorithm to recover the useful contents of the test image.

1) *Problem Formulation:* Each local patch \mathbf{r}_i (e.g., 8×8 block) of the restored image is to satisfy the following conditions:

$$\begin{cases} \nabla \mathbf{r}_i = k_i \times \nabla \mathbf{o}_i, & k_i \in [0, 1] \\ \bar{\mathbf{r}}_i = \bar{\mathbf{t}}_i \end{cases} \quad (1)$$

where i is the local position index and the gradient of \mathbf{r}_i , i.e., $\nabla \mathbf{r}_i$, is assumed to be shrinking compared to that of the original

image patch \mathbf{o}_i , to take into account the influence of the detail loss. The mean luminance of \mathbf{r}_i , i.e., $\bar{\mathbf{r}}_i$, is equal to the mean luminance of the test image patch $\bar{\mathbf{t}}_i$, since intuitively the original mean luminance cannot be recovered from the test image. The formulation of \mathbf{r}_i is given by the solution of (1):

$$\mathbf{r}_i = k_i \times \mathbf{o}_i + (\bar{\mathbf{t}}_i - k_i \times \bar{\mathbf{o}}_i), \quad k_i \in [0, 1]. \quad (2)$$

To make the restored image exhibit equivalent detail losses as the test image, we maximize the similarity between \mathbf{r}_i and \mathbf{t}_i , by setting the scale factor k_i :

$$\max_{k_i \in [0, 1]} \text{similarity}(\mathbf{r}_i, \mathbf{t}_i). \quad (3)$$

In the proposed decoupling algorithm, the similarity between \mathbf{r}_i and \mathbf{t}_i is measured by the sum of squared differences to facilitate its optimization. Thus, (3) is implemented as

$$\begin{aligned} \min_{k_i \in [0, 1]} & \| \mathbf{r}_i - \mathbf{t}_i \|^2 \\ &= \min_{k_i \in [0, 1]} \| k_i \times \mathbf{o}_i + (\bar{\mathbf{t}}_i - k_i \times \bar{\mathbf{o}}_i) - \mathbf{t}_i \|^2 \\ &= \min_{k_i \in [0, 1]} \| k_i \times (\mathbf{o}_i - \bar{\mathbf{o}}_i) - (\mathbf{t}_i - \bar{\mathbf{t}}_i) \|^2 \end{aligned} \quad (4)$$

where $\|\cdot\|$ denotes the l_2 norm. From (4), we can get the closed-form solution for the scale factor k_i :

$$k_i = \text{clip}\left(\frac{\langle (\mathbf{o}_i - \bar{\mathbf{o}}_i) \cdot (\mathbf{t}_i - \bar{\mathbf{t}}_i) \rangle}{\|\mathbf{o}_i - \bar{\mathbf{o}}_i\|^2}, 0, 1\right) \quad (5)$$

where $\langle \cdot \rangle$ denotes the inner product and $\text{clip}(x, 0, 1)$ is equivalent to $\min(\max(x, 0), 1)$. By applying (2) and (5), the restored image \mathbf{r} can be constructed patch by patch.

2) *Wavelet Domain Solution:* Spatial domain solution given by (2) and (5) has its drawbacks. To illustrate, let us first decompose \mathbf{r}_i by

$$\mathbf{r}_i = \mathbf{r}_i^0 + \sum_{s=1}^S \mathbf{r}_i^s \quad (6)$$

where \mathbf{r}_i^s ($s = 0, 1, \dots, S$)¹ indicates the component reconstructed by the wavelet coefficients of the s th subband. By using (2) and (5), the same scale factor k_i will be assigned to each \mathbf{r}_i^s . Actually, subbands are distorted unevenly in most cases. Therefore, it will make more sense to choose a specific k_i^s for each \mathbf{r}_i^s .

The same decomposition as described in (6) can be applied to the original image and the test image to derive \mathbf{o}_i^s and \mathbf{t}_i^s , respectively. Since the mean luminance of both \mathbf{t}_i^s and \mathbf{o}_i^s ($s \neq 0$) is equal to zero, similar to (2), we have

$$\mathbf{r}_i^s = k_i^s \times \mathbf{o}_i^s, \quad k_i^s \in [0, 1]. \quad (7)$$

Given an orthonormal discrete wavelet transform (DWT), the following equations hold:

$$\begin{aligned} \min_{k_i^s \in [0, 1]} & \| \mathbf{r}_i^s - \mathbf{t}_i^s \|^2 \\ &= \min_{k_i^s \in [0, 1]} \| \text{DWT}[\mathbf{r}_i^s - \mathbf{t}_i^s] \|^2 \\ &= \min_{k_i^s \in [0, 1]} \| \text{DWT}[k_i^s \times \mathbf{o}_i^s - \mathbf{t}_i^s] \|^2 \\ &= \min_{k_i^s \in [0, 1]} \| k_i^s \times \text{DWT}[\mathbf{o}_i^s] - \text{DWT}[\mathbf{t}_i^s] \|^2 \\ &= \min_{k_i^s \in [0, 1]} \| k_i^s \times \mathbf{O}_i^s - \mathbf{T}_i^s \|^2 \end{aligned} \quad (8)$$

¹ $s = 0$ indicates the approximation subband.

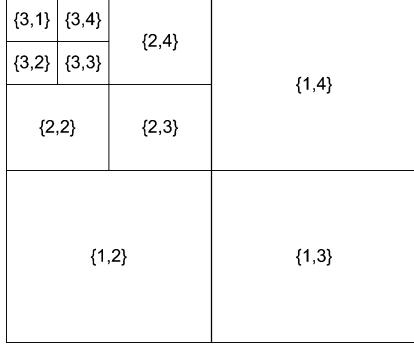


Fig. 3. Subband indexing. Each subband is indexed by a level and an orientation $\{\lambda, \theta\}$. $\theta = 2, 3$, and 4 denote the vertical, diagonal, and horizontal subbands, respectively.

where \mathbf{O}_i^s and \mathbf{T}_i^s denote the DWT coefficients of \mathbf{o}_i^s and \mathbf{t}_i^s , respectively. From (8), we can get the closed-form solution for the scale factor k_i^s :

$$k_i^s = \text{clip}\left(\frac{\langle \mathbf{O}_i^s \cdot \mathbf{T}_i^s \rangle}{\|\mathbf{O}_i^s\|^2}, 0, 1\right). \quad (9)$$

Simplification can be made that instead of using a vector of DWT coefficients, \mathbf{O}_i^s or \mathbf{T}_i^s can be represented by a single DWT coefficient. In this way, (9) is simplified to the division of two scalar values. In the following discussion, $\{\lambda, \theta\}$ is used to index each wavelet subband, as illustrated in Fig. 3 and $\{i, j\}$ is used to index the DWT coefficient position. A four-level $db2$ DWT is applied to both the original and the test images, generating the DWT coefficients $O(\lambda, \theta, i, j)$ and $T(\lambda, \theta, i, j)$, respectively. Based on the above-mentioned simplification, scale factors of the high frequency subbands are given by

$$k(\lambda, \theta, i, j) = \text{clip}\left(\frac{T(\lambda, \theta, i, j)}{O(\lambda, \theta, i, j) + 10^{-30}}, 0, 1\right) \quad (10)$$

where the constant 10^{-30} is to avoid dividing by zero. To satisfy $\tilde{\mathbf{r}}_i = \tilde{\mathbf{t}}_i$ as required by (1), the approximation subband of the restored image is made to equalize that of the test image. Eventually, the DWT coefficients of the restored image can be obtained by

$$R(\lambda, \theta, i, j) = \begin{cases} T(\lambda, \theta, i, j), & \theta = 1 \\ k(\lambda, \theta, i, j) \times O(\lambda, \theta, i, j), & \text{otherwise} \end{cases} \quad (11)$$

where $\theta = 1$ indicates the approximation subband. Since DWT is a linear operator and the additive impairment image is given by $\mathbf{a} = \mathbf{t} - \mathbf{r}$ [as illustrated in Fig. 1(c)], DWT coefficients of \mathbf{a} can be calculated by

$$A(\lambda, \theta, i, j) = T(\lambda, \theta, i, j) - R(\lambda, \theta, i, j). \quad (12)$$

3) Special Case—Contrast Enhancement: Contrast enhancement will improve rather than degrade the image quality, as long as the image contrast is not too high to look natural. If the value of the scale factor k is not bounded, the above decoupling algorithm can handle this special case. However, the constraint on the value of k , i.e., $k \in [0, 1]$, is necessary, due to the use of the sum of squared differences (SSD) to measure the similarity

between the restored image and the test image, as given in (4) and (8). Although it is easy to optimize, the SSD simply measures the pixel differences to approximate rather than accurately quantify the structural similarity. Therefore for many artifacts, e.g., white noises, impulse noises, etc., $k > 1$ may result in a smaller SSD, compared with $k \in [0, 1]$. However, a smaller SSD does not always correspond to a higher structural similarity. In fact, the prior knowledge tells us that except for contrast enhancement, the majority of the distortion types should not lead to $k > 1$. Therefore, instead of relaxing the constraint on the value of k , we develop another method to distinguish contrast enhancement, as elaborated below.

Each DWT coefficient pair of the original image, $O(\lambda, \theta = 2, i, j)$ and $O(\lambda, \theta = 4, i, j)$, is represented by a point in the angular space, whose angle can be calculated by

$$\Psi_o(\lambda, i, j) = \arctan\left(\frac{O(\lambda, \theta = 2, i, j)}{O(\lambda, \theta = 4, i, j) + 10^{-30}}\right) + \pi \times u(-O(\lambda, \theta = 4, i, j)) \quad (13)$$

where $u(\cdot)$ is the unit step function. $\Psi_t(\lambda, i, j)$ of the test image can be obtained in a similar way. The absolute angle difference in degrees is given by

$$\Psi_{diff}(\lambda, i, j) = |\Psi_o(\lambda, i, j) - \Psi_t(\lambda, i, j)| \times \frac{180}{\pi}. \quad (14)$$

Different from the other distortion types, contrast enhancement will result in very small Ψ_{diff} . For example as illustrated in Fig. 4, each blue circle represents a DWT coefficient pair of the original image, $O(\lambda, \theta = 2, i, j)$ and $O(\lambda, \theta = 4, i, j)$, with the corresponding Ψ_o ranging from 44.5° to 45° . The red and green circles represent the same DWT coefficient pairs of the test images, $T(\lambda, \theta = 2, i, j)$ and $T(\lambda, \theta = 4, i, j)$, with the former belonging to a white noise distorted image and the latter to a contrast enhanced image. Apparently, contrast enhancement hardly changes their angles, while white noise does dramatically. We found in more experiments that this observation holds for other angle ranges besides 44.5° to 45° and for other distortion types besides white noises. Moreover, the DWT coefficient pair $O(\lambda, \theta = 2, i, j)$ and $O(\lambda, \theta = 4, i, j)$ used in (13) can be replaced by other pairs indexed by $\{\lambda, \theta = 2, i, j\}$ and $\{\lambda, \theta = 3, i, j\}$, or $\{\lambda, \theta = 3, i, j\}$ and $\{\lambda, \theta = 4, i, j\}$, leading to similar experimental results.

Based on the above analysis, we calculate Ψ_{diff} . If it is smaller than a threshold, contrast change is supposed to occur, in which case no additive impairment exists. This is equivalent to replacing (11) by

$$R(\lambda, \theta, i, j) = \begin{cases} T(\lambda, \theta, i, j), & \theta = 1 \text{ or } \Psi_{diff}(\lambda, i, j) < 1^\circ \\ k(\lambda, \theta, i, j) \times O(\lambda, \theta, i, j), & \text{otherwise} \end{cases} \quad (15)$$

where 1° is experimentally chosen as the threshold, so that contrast enhancement can be largely identified and also no perceptible additive impairment will be observed in the restored image.

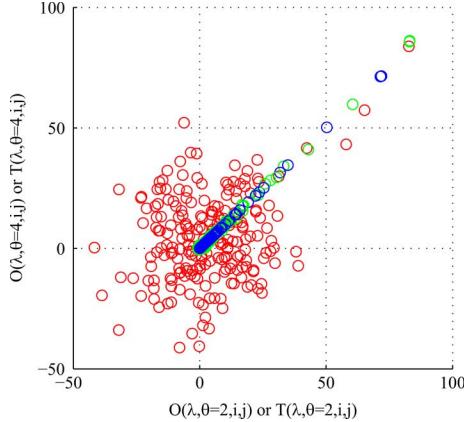


Fig. 4. Blue circles represent $db2$ DWT coefficient pairs $[O(\lambda, \theta = 2, i, j)$ and $O(\lambda, \theta = 4, i, j)]$ of the original image (*I16.bmp* in the image database TID), with $\Psi_o(\lambda, i, j)$ ranging from 44.5° to 45° . The red and green circles represent the corresponding $db2$ DWT coefficient pairs $[T(\lambda, \theta = 2, i, j)$ and $T(\lambda, \theta = 4, i, j)]$ of the test images. The red circles belong to a white noise distorted image (top left corner image of Fig. 5). The green circles belong to a contrast enhanced image (*I16_17_3.bmp* of TID).

4) Visualization of the Decoupling Results: Fig. 5 illustrates the results of the proposed decoupling algorithm, as formulated by (10), (12), and (15), on five typical distortion types: Gaussian white noise, JPEG coding, JPEG2000 coding, Gaussian blur, and wireless transmission error. All test images are from the image database TID [28], [29]. The restored images and the additive impairment images are generated by inverse-transforming their respective DWT coefficients. To better visualize the additive patterns with 8-bit representation, pixel values of each additive impairment image are linearly projected into the range $0 \sim 255$. Generally, the decoupling algorithm succeeds in separating the additive impairments from the useful contents. The additive impairment images show only the additive noises and the restored images contain similar detail losses as their respective distorted images. It should be noted that the contrast of the additive impairment image for Gaussian blur, which originally is quite low, has been enhanced by the linear projection.

B. Simulating the HVS Processing

HVS models typically include sequential processing modules, such as color processing, luminance adaptation, frequency and orientation channel decomposition, contrast sensitivity function, contrast masking, etc., to simulate the low-level HVS processing of the visual inputs. Although anatomy and neuroscience provide us with detailed physiological knowledge about the front end of the HVS (optics, retina, LGN, primary visual cortex, etc.), a thorough understanding of the latter stages of the visual pathway is still unavailable. Consequently, most HVS models take into account only the low-level perception factors evaluated in psychophysical experiments. For the proposed metric, incorporating such a low-level HVS model may serve as a complement to the assumption of the overarching principle, i.e., detail losses and additive impairments should be evaluated separately. Instead of featuring an accurate, usually time-consuming HVS model, the proposed metric efficiently utilizes two HVS characteristics, namely contrast sensitivity

function (CSF) and contrast masking, which were found to be most effective for image quality assessment [30].

1) Contrast Sensitivity Function: Contrast sensitivity is the reciprocal of the contrast threshold, i.e., the minimum contrast value for an observer to detect a stimulus. The contrast thresholds are derived from psychovisual experiments using simple stimuli, like sine-wave gratings or Gabor patches. In these experiments, the stimulus is presented to an observer with its contrast gradually increased. The contrast threshold is determined when the observer is just able to detect the stimulus. The HVS's contrast sensitivity was found to depend on the characteristics of the visual stimulus, e.g., its spatial frequency, orientation, etc. For still images, the CSF peaks at the middle frequency and drops with both increasing and decreasing frequencies. Ngan *et al.* [31] and Nill [32] showed that the generalized CSF model can be expressed as

$$H(\omega) = (a + b\omega)\exp(-c\omega) \quad (16)$$

where a , b , and c are constants and ω is the spatial frequency in cycles per degree (cpd) of visual angle. The proposed metric employs this model with the same parameters setting as in [31] ($a = 0.31$, $b = 0.69$, and $c = 0.29$) which leads to the CSF peaking at 3 cpds subtended at the normal viewing distance of four times the picture height. According to [34] and [33, Section 2.3.1], the nominal spatial frequency of each DWT level λ can be given by

$$F(\lambda) = \frac{\pi \times f_s \times d}{180 \times h \times 2^\lambda} \quad (17)$$

where d is the viewing distance, h is the picture height, and f_s is the cycles per picture height. We set the ratio of d to h as a constant of 4 in the following experiments. As in [2], to take into account the oblique effect, i.e., the HVS is more sensitive to the horizontal and vertical channels than the diagonal channels, the ω in (16) is replaced by $F(\lambda, \theta) = F(\lambda)/[0.15p(\theta) + 0.85]$, where $p(\theta) = 1$ for the vertical and horizontal DWT subbands and $p(\theta) = -1$ for the diagonal DWT subband. As illustrated in Fig. 2, the CSF is applied to the original image and the two decoupled images. It is implemented by multiplying each DWT subband with its corresponding CSF value derived from (16) and (17).

2) Contrast Masking: Contrast masking (CM) refers to the visibility threshold elevation of a target signal caused by the presence of a masker signal. In psychovisual experiments for evaluating CM, the target signal was superposed onto the masker signal and the visibility threshold of the target was recorded for a variety of maskers. It was found that generally higher target visibility threshold (stronger masking effect) will be caused by higher masker contrast and closer similarity between the target and the masker in spatial frequency, orientation, phase, etc. In [34], it was discussed that contrast masking should be computed over a broad range of orientations but over only a limited range of space and spatial frequencies. Therefore, we simply use (18) to calculate the CM thresholds:

$$\text{MT}_\lambda = b(\lambda) \times \sum_{\theta=1}^3 (|\mathbf{M}_{\lambda, \theta}| \otimes \mathbf{w}) \quad (18)$$

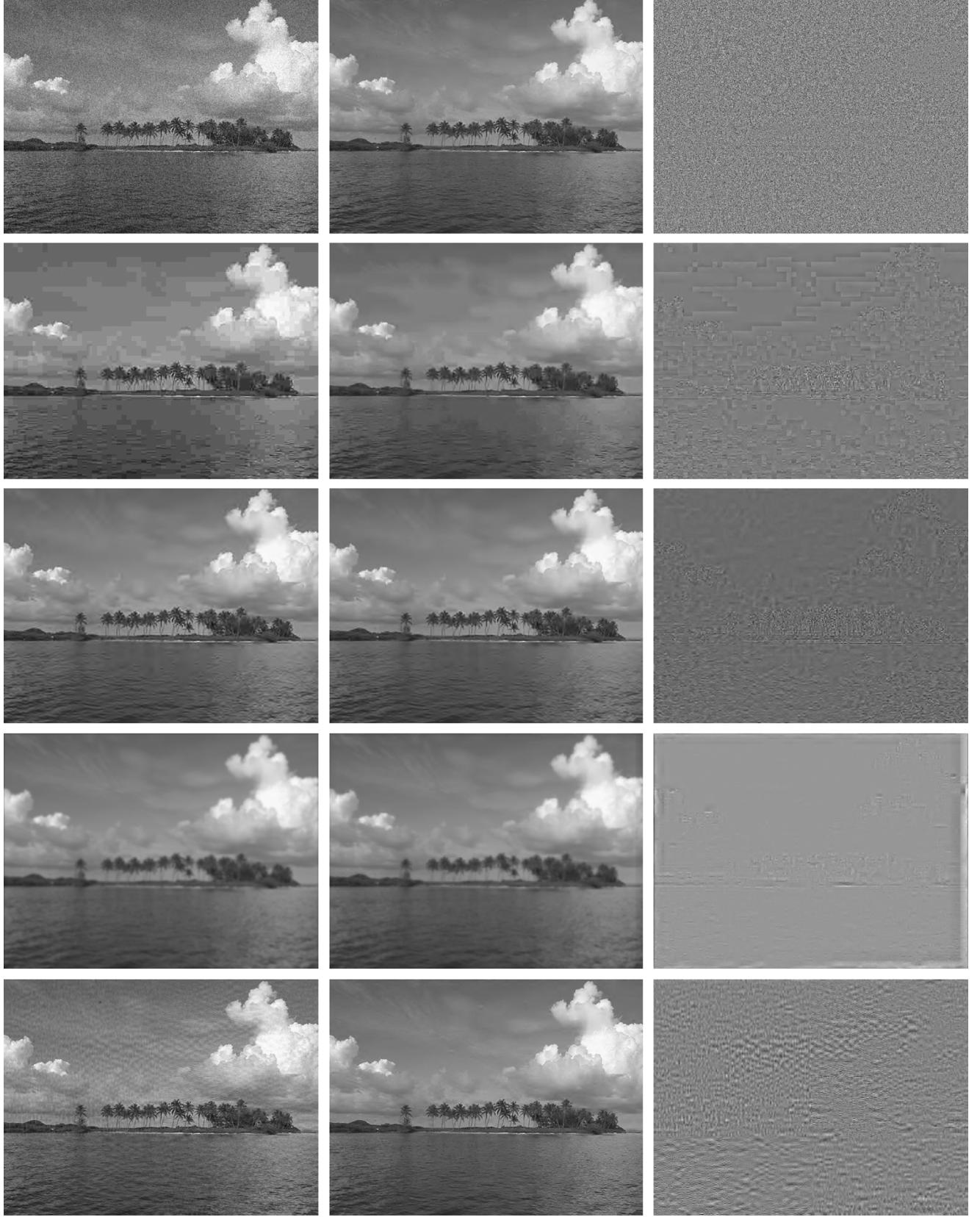


Fig. 5. Results of the proposed decoupling algorithm on five typical distortion types, from top row to bottom row: Gaussian white noise, JPEG coding, JPEG2000 coding, Gaussian blur, and wireless transmission error. From left to right: the test images, the restored images, and the additive impairment images.

where \mathbf{w} is a 3×3 weighting matrix as shown in Fig. 6, $|\mathbf{M}_{\lambda,\theta}|$ is the absolute DWT subband of the masker signal, operator \otimes

indicates convolution, and \mathbf{MT}_λ is the CM threshold map for each of the three DWT subbands in scale λ . The $b(\lambda)$ can be used

1/30	1/30	1/30
1/30	1/15	1/30
1/30	1/30	1/30

Fig. 6. Weighting matrix \mathbf{w} .

to alter the slope of the masking function. After taking into account the CSF, $b(\lambda)$ can be chosen as a constant for all frequency channels [1], [16]. Therefore, as in [26], we set $b(\lambda) = 1$ for all the DWT subbands. In the proposed metric, spatial masking is implemented after the CSF weighting, as shown in Fig. 2. Since the restored and additive impairment images are superposed, one's presence will affect the visibility of the other. Therefore, both images serve as the masker to mask the other. Specifically, we take the absolute value of their CSF-weighted DWT coefficients, subtract from them the corresponding CM thresholds measured by (18), and clip the resultant negative values to 0. On the other hand, for the original image, we directly take the absolute value of its CSF-weighted DWT coefficients as the approximated HVS responses, without considering the CM effect.

C. Two Quality Measures and Their Combination

As mentioned in Section II, the detail loss measure (DLM) and the additive impairment measure (AIM) are given by (19) and (20):

$$q_1 = \left\{ \sum_{\lambda=1}^4 \sum_{\theta=2}^4 \left[\sum_{i,j \in \text{center}} R'(\lambda, \theta, i, j)^{\beta_s} \right]^{\beta_f / \beta_s} \right\}^{1/\beta_f} \quad (19)$$

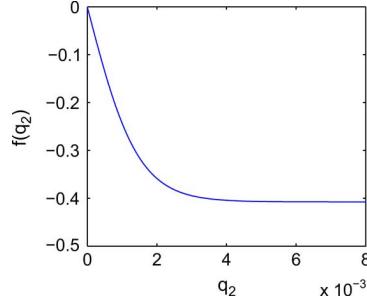
$$q_2 = \left\{ \sum_{\lambda=1}^4 \sum_{\theta=2}^4 \left[\sum_{i,j \in \text{center}} O'(\lambda, \theta, i, j)^{\beta_s} \right]^{\beta_f / \beta_s} \right\}^{1/\beta_f} \quad (20)$$

$$N = \left\{ \sum_{\lambda=1}^4 \sum_{\theta=2}^4 \left[\sum_{i,j \in \text{center}} A'(\lambda, \theta, i, j)^{\beta_s} \right]^{\beta_f / \beta_s} \right\}^{1/\beta_f} \quad (20)$$

where O' , R' , and A' represent the approximated HVS responses of the restored image, the original image, and the additive impairment image, respectively; N is the number of pixels; and $(i, j) \in \text{center}$ indicates that only the central region of each subband is used in the spatial pooling, which helps to overcome the edge effect of the wavelet transform and also serves as a simple region of interest (ROI) model [35]. In (19) and (20), the spatial and subband pooling of the HVS responses are implemented by the Minkowski summation, the general form of which is

$$C = \left(\sum_i c_i^\beta \right)^{1/\beta} \quad (21)$$

where c_i could be element of O' , R' , or A' . Following the parameterization method in [36], we choose the spatial pooling exponent β_s and the frequency pooling exponent β_f by optimizing the metric's performance on a training set. The param-

Fig. 7. Equation (23) with $\alpha_1 = -0.815$ and $\alpha_2 = 1375$.

eterization process is discussed in Section IV-C. It should be noted that the approximation subband is not used in our metric. This is an extreme case of assigning a tiny weight to the approximation subband. By doing so, we exclude the influence of the luminance shift on perceived visual quality, which is well known to be trivial.

DLM and AIM are combined by

$$s = q_1 + f(q_2) \quad (22)$$

where the function $f(\cdot)$ is given by

$$f(x) = \alpha_1 \times \left(0.5 - \frac{1}{1 + \exp(\alpha_2 \times x)} \right). \quad (23)$$

The parameters α_1 and α_2 of (23) are determined by training as discussed in Section IV-C. Fig. 7 shows the training result with $\alpha_1 = -0.815$ and $\alpha_2 = 1375$ which is a monotonously decreasing function with a gradually vanishing slope. The same Δq_2 will result in smaller change of (23) for low-quality images (large q_2) than for high-quality images (small q_2). In other words, q_2 will play a less and less important role as the visual quality deteriorates, while at the same time, the significance of q_1 increases. As shown in Fig. 7, AIM does not further influence the metric output s when its measure q_2 exceeds certain value. For instance, if there are two tested images, both of which have severe additive impairments quantified by $q_2 > 6 \times 10^{-3}$, then the one with the larger q_1 (clearer content visibility) will be predicted by our metric to be of higher visual quality.

IV. EXPERIMENTS

A. Subjective Image Databases

In this section, we present the experimental results of the proposed image quality metric on five subjectively-rated image databases, including LIVE [37], CSIQ [38], TID [28], [29], IVC [39], and TOY [40]. Each of the subjective image databases consists of hundreds of or even thousands of distorted images contaminated by a variety of distortion types, from commonly encountered ones, such as white noises, coding artifacts, etc., to unconventional ones, such as the non-eccentricity pattern and the block pattern in TID. Table I lists major characteristics of each database. The subjectively-rated database provides each of its distorted images a subjective score, e.g., mean opinion score (MOS) or differential mean opinion score (DMOS). These subjective scores were derived from subjective viewing tests in which a lot of human observers participated and provided their

TABLE I
MAJOR CHARACTERISTICS OF THE FIVE SUBJECTIVELY-RATED IMAGE DATABASES

	Distorted Images	Reference Images	Distortion Types	Typical Image Size
TID	1700	25	17	512 × 384
CSIQ	866	30	6	512 × 512
LIVE	779	29	5	768 × 512
IVC	185	10	5	512 × 512
TOY	168	14	2	768 × 512

opinions on the visual quality of each distorted image. The subjective viewing tests followed the international standards [41]–[43] which provide detailed instructions on the viewing conditions, the observer selection procedure, the assessment procedure, etc., to guarantee the reliability of the obtained data. Therefore to evaluate predictive performance, these subjective ratings can be used as the ground truths to be compared against the metric outputs.

B. Performance Measures

Following the Video Quality Expert Group's work [44], each metric score x is mapped to $Q(x)$ first, to obtain a linear relationship between $Q(x)$ and the subjective scores. Its effect is illustrated in Fig. 8. The employed nonlinear mapping function is given by (24):

$$Q(x) = \beta_1 \times \left(0.5 - \frac{1}{1 + \exp(\beta_2 \times (x - \beta_3))} \right) + \beta_4 \times x + \beta_5. \quad (24)$$

The fitting parameters $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ are determined by minimizing the sum of squared differences between the mapped objective scores $Q(x)$ and the subjective ratings. To evaluate the predictive performance, we calculate four common performance measures, which are the linear correlation coefficient (LCC), the root mean squared error (RMSE), the Spearman rank-order correlation coefficients (SROCC), and the outlier ratio (OR). The mapped scores $Q(x)$ and the subjective ratings serve as their inputs. LCC between two data sets, X and Y , is defined as

$$LCC(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X}_i)(Y_i - \bar{Y}_i)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_i)^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}}. \quad (25)$$

RMSE between X and Y is given by

$$RMSE(X, Y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2}. \quad (26)$$

SROCC assesses how well the metric predicts the ordering of the distorted images and can be defined as the LCC of the ranks of X and Y . The OR is defined as

$$OR = \frac{n_{outlier}}{n} \quad (27)$$

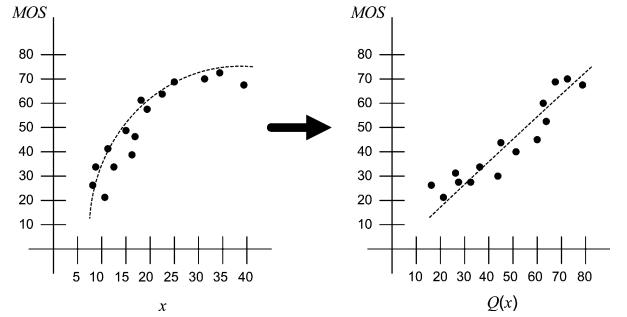


Fig. 8. Illustration of the nonlinear mapping effect.

where $n_{outlier}$ is the number of predictions outside two standard deviations² of the subjective scores and n is the total number of predictions. A comparison between these performance measures can be found in [45].

C. Parameterization

In the proposed metric, there are four parameters to be determined, i.e., the spatial pooling exponent β_s , the frequency pooling exponent β_f , and α_1 and α_2 in (23). β_s and β_f are chosen from integers $\{1, 2, 3, 4\}$ which include the typical values for pooling exponents. Larger value of the pooling exponents makes the quality prediction depend more on the severe distorted regions/subbands. In the literature, two strategies are often used in the parameterization process. One is to choose the parameters depending on how well the resulting model fits the physiological or psychophysical data, e.g., in [11]. The other strategy is to train the parameters to optimize performance in terms of predicting subjective ratings, e.g., [36]. We adopt the second strategy. More precisely, another subjective image database, known as A57 [46], is used to train the four parameters. It consists of 54 test images generated from three 512×512 sized original images. It covers six distortion types including white noises, JPEG compression, JPEG2000 compression, blur, DCQ quantization, and LH subbands quantization. The training objective is to maximize the predictive performance of the proposed metric on database A57. The four parameters are automatically chosen by a global optimization algorithm [47]. The parameterization result is $\beta_s = 3$, $\beta_f = 1$, $\alpha_1 = -0.815$, and $\alpha_2 = 1375$. The database A57 is only used in this parameter training process. In the following experiments, the proposed metric is tested on the other five subjective image databases as introduced in Section IV-A.

D. Overall Performance

Fig. 9 shows the scatter plots of the proposed image quality metric on five subjective image databases. In all the graphs, each point represents a test image. The vertical axis denotes the subjective ratings of the perceived distortions and the horizontal axis denotes the nonlinearly mapped metric outputs.

In Table II, the proposed algorithm is compared against eight full-reference image quality metrics (IQM), i.e., PSNR, WSNR [48], NQM [1], VSNR [49], MSSIM [50], Q [23], IWSSIM [51], and VIF [24], all of which work with luminance

²The standard deviation indicates the variation of individual subjective ratings around the mean value.

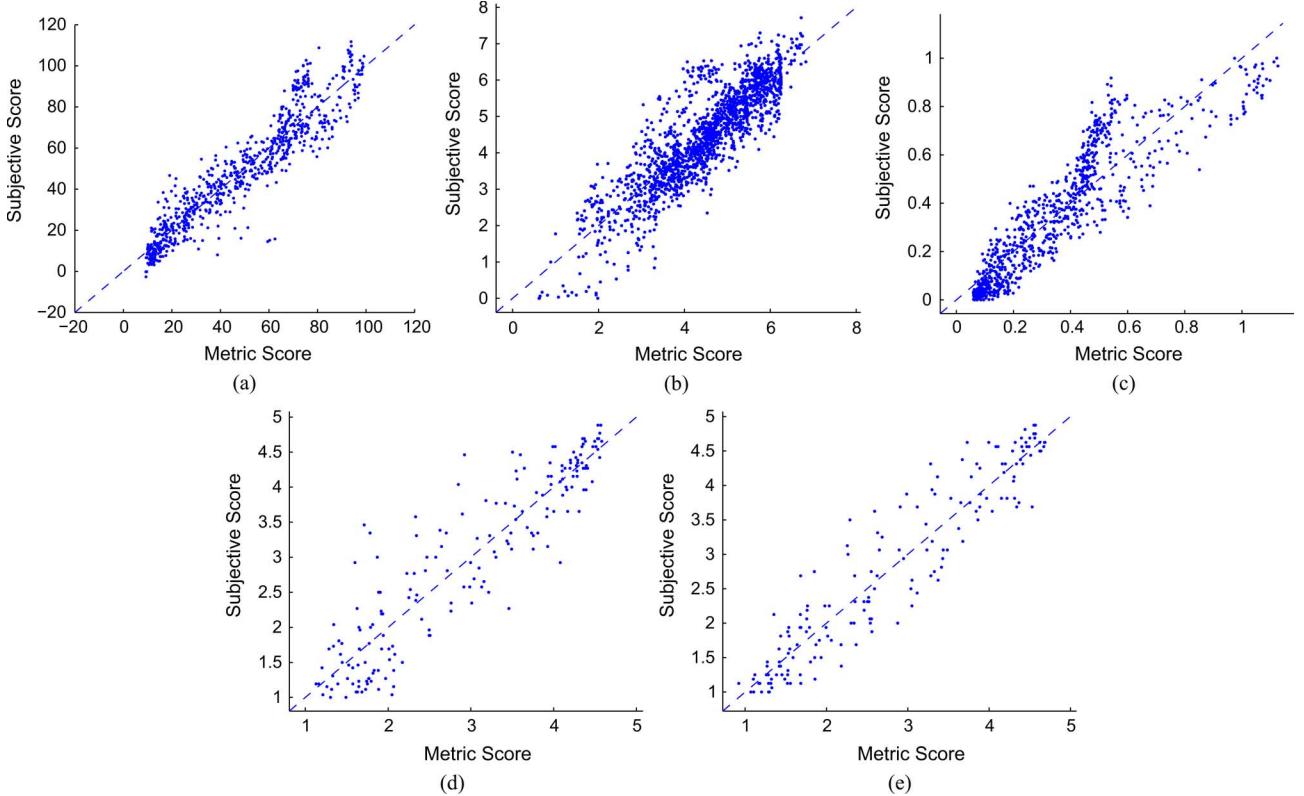


Fig. 9. Scatter plots of the proposed image quality metric on five subjective image databases (after the nonlinear mapping). (a) LIVE. (b) TID. (c) CSIQ. (d) IVC. (e) TOY.

only.³ LCC, RMSE, SROCC, and OR are calculated for performance comparison. According to their definitions that are briefly introduced in Section IV-B, larger LCC and SROCC indicate stronger correlation between the metric outputs and the subjective ratings, while smaller RMSE and OR indicate less prediction errors. OR can be calculated for three image databases only, LIVE, CSIQ, and TOY, since the others do not provide the standard deviation data. In general, the proposed IQM demonstrates quite good and stable performance. It is the best-performer or tied for the best-performer on most of the databases. LIVE image database contains five commonly encountered distortion types. It covers a much wide range of distortion intensity, which facilitates quality prediction. On the other hand, TID contains 17 distortion categories, a relatively small distortion intensity range, and images with both natural scenes and artificial contents. This explains why most tested IQMs achieve their best performance on LIVE and worst performance on TID.

E. Statistical Significance

To assess the statistical significance of the performance difference between two metrics, F-test is conducted on the prediction residuals between the metric outputs (after nonlinear mapping) and the subjective ratings. The residuals are supposed to be Gaussian. Smaller residual variance implies more accurate prediction. Let F denote the ratio between the residual variances

of two different metrics (with the larger variance as the numerator). If F is larger than $F_{critical}$ which is calculated based on the number of residuals and a given confidence level, then the difference between the two metrics is considered to be significant at the specified confidence level. Table III lists the residual variance of each metric on five subjective image databases. Notably due to the differences in employed subjective scales, the residual variance varies a lot across different image databases. As in [45], we use a simple criterion to measure the Gaussianity of the prediction residuals: if the residuals have a kurtosis between 2 and 4, they are taken to be Gaussian. The results of the Gaussian test are given in Table III, with (1) indicating Gaussian and (0) indicating non-Gaussian. The $F_{critical}$ with 95% confidence is also shown in Table III for each database. In Table IV, the proposed metric is compared with the other metrics regarding the statistical significance. In each entry, the symbol “1”, “0”, or “-” means that on the image databases indicated by the first column of the table, the proposed metric is statistically (with 95% confidence) better, worse, or indistinguishable, respectively, when compared with its competitors indicated by the first row. As shown in Table IV, the proposed metric outperforms most of its competitors statistically. Although its performance on database LIVE is worse in comparison to the state-of-the-art metrics VIF and IWSSIM, it demonstrates better performance on other two image databases. Notably the prediction residual of the proposed metric is less Gaussian than those of the VIF and the IWSSIM on several databases. As analyzed in [2], this implies that the residual variance of the proposed metric has been inflated by a few outliers.

³MATLAB function `rgb2gray` is used in the experiment to convert color images to gray scale.

TABLE II
OVERALL PERFORMANCE AND COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN NINE OBJECTIVE IQMs. THE BOLDFACE ENTRIES INDICATE THE STATISTICALLY BEST PERFORMERS AND THEIR EQUIVALENTS AS ELABORATED IN SECTION IV-E

		PSNR	WSNR	NQM	VSNR	Q	MSSIM	VIF	IWSSIM	Proposed
LIVE	LCC	0.872	0.914	0.911	0.923	0.924	0.942	0.960	0.951	0.936
	RMSE	13.36	11.06	11.21	10.50	10.40	9.140	7.649	8.425	9.627
	SROCC	0.856	0.915	0.908	0.927	0.925	0.944	0.964	0.957	0.946
	OR	0.688	0.577	0.636	0.599	0.539	0.561	0.535	0.531	0.539
TID	LCC	0.562	0.564	0.614	0.675	0.816	0.841	0.808	0.858	0.869
	RMSE	1.110	1.108	1.059	1.018	0.776	0.726	0.789	0.690	0.662
	SROCC	0.553	0.581	0.624	0.652	0.817	0.852	0.749	0.856	0.861
CSIQ	LCC	0.799	0.773	0.741	0.804	0.900	0.897	0.927	0.906	0.928
	RMSE	0.158	0.166	0.176	0.159	0.114	0.112	0.098	0.111	0.098
	SROCC	0.804	0.783	0.743	0.796	0.881	0.913	0.919	0.921	0.933
	OR	0.341	0.326	0.368	0.311	0.258	0.248	0.226	0.246	0.218
IVC	LCC	0.704	0.862	0.849	0.803	0.839	0.881	0.903	0.923	0.913
	RMSE	0.866	0.618	0.642	0.726	0.662	0.561	0.524	0.469	0.496
	SROCC	0.679	0.859	0.834	0.798	0.831	0.884	0.896	0.913	0.903
TOY	LCC	0.642	0.800	0.893	0.870	0.891	0.892	0.914	0.925	0.942
	RMSE	0.959	0.750	0.562	0.617	0.568	0.564	0.508	0.476	0.421
	SROCC	0.613	0.801	0.891	0.861	0.869	0.886	0.908	0.920	0.937
	OR	0.214	0.136	0.071	0.095	0.077	0.101	0.047	0.048	0.035
Speed (sec./image)	0.01	0.19	0.95	0.11	5.41	0.27	1.90	0.97	0.34	

TABLE III
RESIDUAL VARIANCES OF NINE METRICS ON FIVE SUBJECTIVE IMAGE DATABASES. SYMBOL “1” IN THE BRACKET MEANS THE DISTRIBUTION OF THE RESIDUAL IS GAUSSIAN. SYMBOL “0” MEANS NON-GAUSSIAN

	LIVE $F_{critical} = 1.125$	TID $F_{critical} = 1.082$	CSIQ $F_{critical} = 1.118$	IVC $F_{critical} = 1.275$	TOY $F_{critical} = 1.291$
PSNR	181.0 (1)	1.232 (1)	0.0249 (1)	0.719 (1)	0.924 (1)
WSNR	122.5 (0)	1.228 (0)	0.0277 (1)	0.383 (1)	0.566 (1)
NQM	125.9 (0)	1.122 (0)	0.0309 (1)	0.414 (1)	0.317 (1)
VSNR	110.5 (1)	0.967 (0)	0.0323 (1)	0.530 (1)	0.382 (0)
Q	108.3 (0)	0.601 (0)	0.0130 (0)	0.441 (1)	0.324 (1)
MSSIM	83.64 (1)	0.527 (1)	0.0137 (1)	0.307 (1)	0.320 (1)
VIF	58.54 (1)	0.623 (1)	0.0096 (1)	0.276 (1)	0.260 (1)
IWSSIM	71.06 (0)	0.476 (0)	0.0123 (1)	0.221 (1)	0.228 (1)
Proposed	92.78 (0)	0.439 (0)	0.0095 (1)	0.247 (0)	0.178 (1)

TABLE IV
PERFORMANCE COMPARISON REGARDING STATISTICAL SIGNIFICANCE. IN EACH ENTRY, THE SYMBOL “1”, “0”, OR “-” MEANS THAT ON THE IMAGE DATABASE, THE PROPOSED METRIC IS STATISTICALLY (WITH 95% CONFIDENCE) BETTER, WORSE, OR INDISTINGUISHABLE IN COMPARISON TO ITS COMPETITOR

	PSNR	WSNR	NQM	VSNR	Q	MSSIM	VIF	IWSSIM
LIVE	1	1	1	1	1	-	0	0
TID	1	1	1	1	1	1	1	
CSIQ	1	1	1	1	1	1	-	1
IVC	1	1	1	1	1	-	-	-
TOY	1	1	1	1	1	1	1	-

F. Performance on Individual Distortion Types

Good (bad) overall performance does not necessarily mean good (bad) performance for individual distortion types. For example, it is possible that a metric which achieves excellent performance for individual distortion types fails badly regarding overall performance: it may tolerate distortion type A more than B, leading to higher quality prediction for type-A images than type-B images. Therefore, in this subsection, we compare the metric performance on individual distortion types. More precisely, in Table V we show the predictive performance of nine IQMs on five common distortion types, i.e., JPEG compression, JPEG2000 compression, additive Gaussian white noise (AGWN), blur, and contrast change. For easier comparison, only the SROCC values are listed. SROCC is chosen because it is suitable for measuring a small number of data points and its value will not be affected by an unsuccessful

monotonic nonlinear mapping. In Table V, the bold entries denote the best performer in terms of SROCC for each distortion type on each database. The underlined entries denote the statistically best performer (with 95% confidence) and the italicized entries indicate metrics that are statistically indistinguishable from the underlined ones. In general, both the proposed metric and the IWSSIM achieve the highest SROCC on six image sets. By comparing the SROCC difference, it can be observed that the proposed metric is tied for the best performer (in terms of SROCC) on most of the tested image sets. Furthermore, the proposed algorithm is the statistically best-performing metric on nine image sets and is statistically equivalent to the best performer on all but two of the rest of the image sets.

Many factors may contribute to the inconsistent predictive performances across the different databases. For instance, most of the IQMs in our experiment including the proposed one perform quite well for white noise of all databases except TID. Further looking into this exception, it was found out that white noise of TID spans a relatively small quality variation (PSNR variance 11.4) compared with the other databases (e.g., PSNR variance 96.6 for LIVE), which will increase prediction difficulties. The same explanation can be applied to the varying metric performance on distortion type JPEG, for which LIVE and CSIQ exhibit larger quality variation and thus facilitate quality prediction. Another cross-database performance variance can be found in the distortion type Contrast Change: all IQMs in the

TABLE V
PERFORMANCE COMPARISON ON INDIVIDUAL DISTORTION TYPES, USING SROCC AS THE PERFORMANCE MEASURE. THE BOLD ENTRIES DENOTE THE BEST PERFORMER IN TERMS OF SROCC. THE UNDERLINED ENTRIES DENOTE THE STATISTICALLY BEST PERFORMER (WITH 95% CONFIDENCE). THE ITALICIZED ENTRIES INDICATE METRICS THAT ARE STATISTICALLY INDISTINGUISHABLE FROM THE UNDERLINED ONES

		PSNR	WSNR	NQM	VSNR	Q	MSSIM	VIF	IWSSIM	Proposed	Images
Blur	LIVE	0.782	0.908	0.839	0.941	0.868	0.958	0.972	0.972	0.965	145
	TID	0.869	0.932	0.884	0.933	0.882	<u>0.960</u>	<u>0.954</u>	0.964	0.957	100
	CSIQ	0.929	0.964	0.958	0.944	0.942	0.971	<u>0.974</u>	0.978	0.972	150
	IVC	0.805	0.881	0.956	0.968	0.867	0.944	0.972	0.962	0.944	20
AWGN	LIVE	0.985	0.973	0.986	0.978	0.978	0.973	<u>0.985</u>	0.967	0.982	145
	TID	0.907	0.871	0.767	0.772	0.844	0.809	0.879	0.786	0.876	100
	CSIQ	0.936	0.885	0.938	0.924	0.928	0.947	<u>0.957</u>	0.938	0.958	150
Contrast change	TID	0.585	0.572	0.619	0.423	0.521	0.639	<u>0.818</u>	0.630	0.852	100
	CSIQ	0.862	0.822	0.947	0.869	0.901	<u>0.951</u>	0.934	0.954	<u>0.951</u>	116
JPEG	LIVE	0.880	0.957	0.964	0.965	0.967	0.979	0.984	0.980	0.979	175
	TID	0.871	0.932	0.907	0.917	0.913	0.934	<u>0.917</u>	0.918	<u>0.932</u>	100
	CSIQ	0.887	0.956	0.952	0.903	0.957	0.962	0.970	0.966	<u>0.967</u>	150
	IVC	0.738	0.878	0.853	0.777	0.819	0.917	<u>0.931</u>	0.946	<u>0.925</u>	50
	TOY	0.286	0.745	0.889	0.797	0.872	0.834	<u>0.906</u>	0.920	<u>0.917</u>	84
JPEG2000	LIVE	0.895	0.920	0.943	0.955	0.954	0.965	<u>0.969</u>	0.965	0.971	169
	TID	0.813	0.956	0.953	0.951	0.941	0.973	<u>0.970</u>	0.972	0.967	100
	CSIQ	0.936	0.969	0.963	0.947	0.958	0.968	0.967	0.968	0.975	150
	IVC	0.850	0.917	0.859	0.868	0.909	0.928	<u>0.935</u>	0.949	0.952	50
	TOY	0.860	0.883	0.907	0.925	0.905	0.945	<u>0.955</u>	0.954	0.961	84

test demonstrate much better performance on Csiq than Tid. The main reason is that Csiq contains contrast reduction only, while Tid covers both contrast reduction and enhancement, the latter cannot be handled well by most IQMs. The proposed IQM demonstrates best performance for Contrast Change on Tid, mainly because it deliberately takes contrast enhancement (CE) into consideration, as elaborated in Section III-A, without which the SROCC will drop from 0.85 to around 0.41. Since our metric cannot perfectly separate CE from additive impairments and does not distinguish moderate CE from excessive CE, its performance on CE is just moderately better than the second best performer.

G. Hypotheses Validation

By using the abundant subjective data, in this subsection we try to justify three hypotheses made by us in Section II: 1) using the restored and additive impairment images to mask each other is a better way to implement spatial masking in comparison to the traditional method which uses the reference image as the masker; 2) decoupling detail losses and additive impairments do benefit the perceptual quality prediction; and 3) detail losses and additive impairments correlate with visual quality in different ways so that it is better to measure them using different methods.

To verify the first assumption, we modify the spatial masking process of the proposed metric. Since the restored and additive impairment images are superposed to form the test image, one's presence will affect the visibility of the other. Therefore, in the proposed metric, both images serve as the masker to modulate the intensity of the other. As discussed in Section II, traditional masking uses reference image to mask the distortions. We argued that it might be inferior to the proposed method since the test image often becomes less textured than the reference and it is impossible to simulate how additive impairments diminish content visibility. To prove this, in our metric we replace the proposed masking with the traditional one by using the reference to mask the additive impairments and by leaving the restored image unchanged. All the other components are kept the

TABLE VI
OVERALL PERFORMANCE COMPARISON (IN TERMS OF SROCC) BETWEEN THE PROPOSED METRIC, M1 AND M2. M1 IS A MODIFIED VERSION OF THE PROPOSED METRIC USING TRADITIONAL SPATIAL MASKING SCHEME. M2 IS A MODIFIED VERSION OF THE PROPOSED METRIC WITHOUT DECOUPLING ADDITIVE IMPAIRMENTS AND DETAIL LOSSES

	LIVE	TID	CSIQ	IVC	TOY
Proposed	0.946	0.861	0.933	0.903	0.937
M1	0.939	0.839	0.904	0.896	0.927
M2	0.902	0.710	0.871	0.859	0.834

same and the parameters are trained on database A57. The resulting metric is named as M1 and its performance is compared with the proposed metric in Table VI. It can be observed that although the performance improvement of the proposed metric over M1 is marginal on three databases, the proposed masking indeed outperforms the traditional one consistently.

The advantage of the proposed decoupling strategy comes from several aspects. Besides the new masking strategy mentioned above, the decoupling also makes the measure of additive impairments and detail losses using specific ways possible, the necessity of which will be further discussed in the following paragraph. Furthermore, the high-level HVS assumption as discussed in Sections I and II can be implemented when combining the two quality measures together. By discarding all these merits, we develop another metric, known as M2, which treats the image differences homogeneously and utilizes the traditional strategy for spatial masking. For fair comparison, M2 also employs the contrast sensitivity function described in Section III-B. The classical Minkowski summation is applied for the spatial and frequency pooling, with the two pooling exponents trained on database A57. The performance of M2 is shown in Table VI. We can see that without the benefits from the proposed decoupling, the predictive performance dramatically degrades.

From Table V, it can be observed that PSNR achieves outstanding predictive performance on additive noise (AWGN) distorted images: it is either the best performer or statistically equivalent to the best performer. However, PSNR performs

TABLE VII

PERFORMANCE COMPARISON (IN TERMS OF SROCC) BETWEEN M2 AND q_1 ON DISTORTION TYPE BLUR. M2 UTILIZES THE CLASSICAL MINKOWSKI SUMMATION TO INTEGRATE DISTORTIONS WITHOUT CONSIDERING THE ORIGINAL CONTENT. ON THE OTHER HAND, q_1 AS GIVEN BY (19) NORMALIZES THE INTEGRATED DISTORTION INTENSITY WITH THE ORIGINAL CONTENT

	Blur			
	LIVE	TID	CSIQ	IVC
M2	0.843	0.898	0.939	0.871
DLM	0.966	0.953	0.974	0.930

badly for images distorted by JPEG or JPEG2000 compression, blur, etc., which majorly suffer from detail loss. This observation inspires us to hypothesize that additive impairment and detail loss may be associated with perceptual quality in different ways and thus should be measured by two distinct methods. Specifically, since additive impairments are relatively independent of the original image content, we assume that visual quality with respect to additive impairment can be predicted by analyzing their intensities without considering the original content. On the other hand, visual quality with respect to detail losses is supposed to be determined by the percentage of visual information losses/remains. This assumption is consistent with VIF in which the useful information communicated by the test image is normalized by the total information of the original image. Therefore, in (20) and (19), the integrated distortion intensity is normalized by the pixel number and the original image content, respectively. It should be noted that DLM [given by (19)] is an approximate calculation of the percentage of visual information remains. Its low complexity makes the proposed metric time efficient. As a primary validation of the above assumption, we compare the predictive performances of DLM and M2 on images distorted by Gaussian blur. Blurred images mainly suffer from detail losses; hence, it is suitable to use blurred images to test detail loss measures. As mentioned above, M2 which can be considered as an improved version of PSNR that utilizes the classical Minkowski summation to predict visual quality. In comparison to DLM, it only analyzes the distortion intensity but does not consider the original content. Table VII shows the comparison results. DLM outperforms M2 by a large margin which indicates the usefulness of taking into account the original content in detail loss measure. The development of more accurate additive impairment measure and detail loss measure needs further investigation. We advocate in this paper to consider their differences for visual quality assessment.

H. Complexity Analysis

The execution speed of each IQM is listed in the bottom row of Table II. We ran the speed test on a PC with a dual-core 3-GHz CPU and a 4-Gbyte memory. The speed is in the unit of second per image and is an averaged value calculated on the image database A57. The proposed algorithm is implemented using MATLAB. IWSSM is downloaded from the author's website [52]. Q is provided by its author. Both of them are implemented using MATLAB. The other IQMs are from the visual quality assessment package MeTriX MuX version 1.1 [53]. As claimed in [53], MeTriX MuX is a MATLAB package that implements

wrapper code to provide a common interface for various IQMs and almost every IQM code is in its original form, i.e., programmed by the IQM's authors. It can be seen from Table II that the proposed metric has a moderate computational complexity compared with the others.

The huge complexity of VIF mainly comes from the highly overcomplete steerable pyramid decomposition. IWSSM applies a five-scale Laplacian pyramid decomposition and computes an information content weight map for each scale. Q performs singular value decomposition for both the original and the test image which is time consuming. Both VSNR and MSSIM employ 9/7 biorthogonal wavelet transform and the faster speed of VSNR is partially attributed to its .dll implementation of the wavelet transform. Both WSNR and NQM utilize fast Fourier transform implemented by MATLAB function. However, WSNR simply weights the Fourier domain differences with CSF, while NQM takes into account more HVS characteristics such as contrast calculation, CSF, contrast masking, etc., which makes NQM much more time consuming. The proposed IQM uses *db2* orthonormal wavelet transform. Actually, the performance of our metric is not sensitive to the choice of the wavelet transform. Using *db2* wavelet is mainly out of concern for algorithm complexity: higher order wavelets are not used because they will increase complexity without further enhancing performance. In fact, *Haar* (*db1*) wavelet, which is even faster, can perform equally well given a slightly modification on the parameter setting. However, by using *Haar* wavelet transform, the decoupled images will appear to be severe blocky, especially for low-quality test images. To avoid the misconception that such blockiness is an intrinsic flaw of the proposed decoupling algorithm, *db2* wavelet transform is used instead of *Haar*.

V. CONCLUSION

An effective algorithm is proposed which explicitly separates detail losses and additive impairments for image quality assessment. For impairments separation, a wavelet domain decoupling algorithm is developed which works effectively for most distortion types and is able to distinguish contrast enhancements from additive impairments. Two important HVS characteristics, i.e., CSF and contrast masking, are incorporated into the metric to better simulate HVS responses to the visual inputs. We propose two simple quality measures DLM and AIM, which are responsible for correlating detail losses and additive impairments with visual quality, respectively. The outputs of DLM and AIM are adaptively combined to yield the overall quality measure, in such a way that DLM will play a more important role in quality prediction for low-quality images than for high-quality images. By experiments based on five subjectively-rated image databases, we demonstrate the effectiveness of the proposed image quality metric in matching subjective ratings. For the future work, we will take into account the chrominance distortions and extend the proposed algorithm to video quality assessment. Moreover, we will investigate the influence of distortion distribution on the perceived visual quality, e.g., the differences between detail losses happening to the ROI and non-ROI and between local additive impairments and global ones, etc.

REFERENCES

- [1] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.
- [2] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 1–21, 2010.
- [3] D. M. Rouse, R. Pepion, S. S. Hemami, and P. L. Callet, "Image utility assessment and a relationship with image quality assessment," *Human Vis. Electron. Imag. XIV*, vol. 7240, pp. 724010–724010-14, 2009.
- [4] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," in *Proc. SIGGRAPH Conf.*, 1999, pp. 73–82.
- [5] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process. Lett.*, vol. 4, no. 11, pp. 317–320, Nov. 1997.
- [6] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process.-Image Commun.*, vol. 19, no. 2, pp. 163–172, 2004.
- [7] E. P. Ong, W. S. Lin, Z. K. Lu, X. K. Yang, S. S. Yao, F. Pan, L. J. Jiang, and F. Moschetti, "A no-reference quality metric for measuring image blur," in *Proc. 7th Int. Symp. Signal Processing and Its Applications*, 2003, vol. 1, pp. 469–472.
- [8] H. Cheng and J. Lubin, "Reference free objective quality metrics for MPEG coded video," *Human Vis. Electron. Imag. X*, vol. 5666, pp. 160–167, 2005.
- [9] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.
- [10] Z. Wang and E. P. Simoncelli, "Local phase coherence and the perception of blur," *Adv. Neural Inf. Process. Syst. 16*, vol. 16, pp. 1435–1442, 2004.
- [11] S. Winkler, "A perceptual distortion metric for digital color video," *Human Vis. Electron. Imag. IV*, vol. 3644, pp. 175–184, 1999.
- [12] P. C. Teo and D. J. Heeger, "Perceptual image distortion," *Human Vis., Vis. Process., Digit. Display V*, vol. 2179, pp. 127–141, 1994.
- [13] Y. K. Lai and C. C. J. Kuo, "A Haar wavelet approach to compressed image quality measurement," *J. Vis. Commun. Image Represent.*, vol. 11, no. 1, pp. 17–40, 2000.
- [14] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," *Soc. Inf. Display Digest of Technical Papers, XXIV*, pp. 946–949, 1993.
- [15] E. Peli, "Contrast in complex images," *J. Optic. Soc. Amer. Optics Image Sci. Vis.*, vol. 7, no. 10, pp. 2032–2040, 1990.
- [16] S. Daly, "The visible differences predictor—an algorithm for the assessment of image fidelity," *Human Vis., Vis. Process., Digit. Display III*, vol. 1666, pp. 2–15, 1992.
- [17] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993.
- [18] V. Laparra, J. Munoz-Mari, and J. Malo, "Divisive normalization image quality metric revisited," *J. Optic. Soc. Amer. Optics Image Sci. Vis.*, vol. 27, no. 4, pp. 852–864, 2010.
- [19] 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.
- [20] G. Zhai, W. Zhang, X. Yang, and Y. Xu, "Image quality assessment metrics based on multi-scale edge presentation," in *Proc. IEEE Workshop Signal Process. Syst. Design Implement.*, 2005, pp. 331–336.
- [21] C. L. Yang, W. R. Gao, and L. M. Po, "Discrete wavelet transform based structural similarity for image quality assessment," in *Proc. ICIP*, 2008, pp. 377–380.
- [22] M. Zhang and X. Mou, "A psychovisual image quality metric based on multi-scale structure similarity," in *Proc. ICIP*, 2008, pp. 381–384.
- [23] M. Narwaria and W. S. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.
- [24] 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.
- [25] K. Seshadrinathan and A. C. Bovik, "Unifying Analysis of Full Reference Image Quality Assessment," in *Proc. IEEE Int. Conf. Image Processing*, 2008, vol. 1–5, pp. 1200–1203.
- [26] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 717–730, May 1999.
- [27] S. Li and K. N. Ngan, "Subtractive impairment, additive impairment and image visual quality," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS2010)*, Paris, France, 2010, pp. 3373–3376.
- [28] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin, "TAMPERE IMAGE DATABASE 2008 TID2008, Version 1.0, 2008. [Online]. Available: <http://www.ponomarenko.info/tid2008.htm>.
- [29] N. Ponomarenko, F. Battisti, and K. Egiazarian, "Metrics performance comparison for color image database," in *Proc. 4th Int. Workshop Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, AZ, Jan. 14–16, 2009, p. 6.
- [30] F. Zhang, S. Li, L. Ma, and K. N. Ngan, "Limitation and challenges of image quality measurement," in *Proc. Visual Communications and Image Processing Conf. (VCIP2010)*, Huang Shan, China, 2010, vol. 7744, pp. 774402.1–774402.8.
- [31] K. N. Ngan, K. S. Leong, and H. Singh, "Adaptive cosine transform coding of images in perceptual domain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 11, pp. 1743–1750, Nov. 1989.
- [32] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. 33, no. 6, pp. 551–557, Jun. 1985.
- [33] Y. Wang, J. Ostermann, and Y. Q. Zhang, *Video Processing and Communications*. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [34] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1164–1175, Aug. 1997.
- [35] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.
- [36] M. Masry, S. S. Hemami, and Y. Sermadevi, "A scalable wavelet-based video distortion metric and applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 260–273, Feb. 2006.
- [37] H. R. Sheikh *et al.*, LIVE Image Quality Assessment Database, Release 2, 2005. [Online]. Available: <http://live.ece.utexas.edu/research/quality>.
- [38] D. M. Chandler, CSIQ Database, 2010. [Online]. Available: <http://vision.okstate.edu/csiq/>.
- [39] P. L. Callet and F. Autrusseau, Subjective Quality Assessment IRCCyN/IVC Database, 2005. [Online]. Available: <http://www.irccyn.ec-nantes.fr/ivcdb/>.
- [40] Y. Horita *et al.*, MICT Image Quality Evaluation Database. [Online]. Available: <http://mict.eng.u-toyama.ac.jp/mict/index2.html>.
- [41] ITU-R Recommendation BT.710-4, Subjective Assessment Methods for Image Quality in High-Definition Television. Geneva, Switzerland, International Telecommunication Union, 1998.
- [42] ITU-R Recommendation BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures. Geneva, Switzerland, International Telecommunication Union, 2002.
- [43] ITU-R Recommendation BT.814-1, Specifications and Alignment Procedures for Setting of Brightness and Contrast of Displays. Geneva, Switzerland, International Telecommunication Union, 1994.
- [44] Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment II, Video Quality Expert Group (VQEG), 2003. [Online]. Available: <http://www.vqeg.org/>.
- [45] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
- [46] D. M. Chandler and S. S. Hemami, A57 Database. [Online]. Available: <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.
- [47] Genetic Algorithm Toolbox. [Online]. Available: <http://www.shef.ac.uk/acse/research/ecrg/gat.html>.
- [48] K. V. T. Mitsu, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in *Proc. ICASSP*, 1993, pp. 301–304.
- [49] 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.
- [50] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. Conf. Record 37th Asilomar Conf. Signals, Systems & Computers*, 2003, vol. 1 and 2, 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. Wang. [Online]. Available: <https://ece.uwaterloo.ca/~z70wang/research/iwssim/>.
- [53] M. Gaubatz, Metrix MUX Visual Quality Assessment Package: MSE, PSNR, SSIM, MSSIM, VSNR, VIF, VIFP, UQI, IFC, NQM, WSNR, SNR. [Online]. Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux.



Songnan Li (S'08) received the B.S. and M.S. degrees from the Department of Computer Science, Harbin Institute of Technology, Harbin, China, in 2004 and 2006, respectively. He is now pursuing the Ph.D. degree in the Department of Electronic Engineering, the Chinese University of Hong Kong.

His research interests include visual quality assessment, video de-interlacing, video compression, algorithm complexity optimization, etc.



King Ngi Ngan (M'79–SM'91–F'00) received the Ph.D. degree in electrical engineering from the Loughborough University, Loughborough, U.K.

He is currently a Chair Professor in the Department of Electronic Engineering, Chinese University of Hong Kong. He was previously a full Professor at the Nanyang Technological University, Singapore, and the University of Western Australia, Perth. He holds honorary and visiting professorships from numerous universities in China, Australia, and South East Asia. He has published extensively including

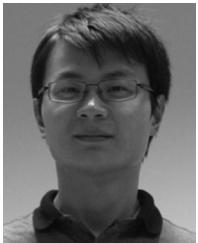
three authored books, five edited volumes, over 300 refereed technical papers, and edited nine special issues in journals. In addition, he holds ten patents in the areas of image/video coding and communications.

Prof. Ngan is an Associate Editor of the *Journal on Visual Communications and Image Representation*, as well as an Area Editor of *EURASIP Journal of Signal Processing: Image Communication* and served as an Associate Editor of the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY* and *Journal of Applied Signal Processing*. He chaired a number of prestigious international conferences on video signal processing and communications and served on the advisory and technical committees of numerous professional organizations. He was a general co-chair of the IEEE International Conference on Image Processing (ICIP) held in Hong Kong in September 2010. He is a Fellow of IET (U.K.) and IEAust (Australia) and an IEEE Distinguished Lecturer in 2006–2007.



Fan Zhang (M'11) received the B.E. and Ph.D. degrees from Huazhong University of Science and Technology, Wuhan, China, in 2002 and 2008, respectively, both in electronic and information engineering.

He was a visiting student at the Nanyang Technological University in 2008 and a postdoctor at the Chinese University of Hong Kong from 2009 to 2010. He has been a research engineer with Technicolor Company, Beijing, China, since 2010. His research interests are quality-of-experience and perceptual watermarking.



Lin Ma (S'09) received the B.S. and M.S. degrees from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, both in computer science. He is now pursuing the Ph.D. degree in the Department of Electronic Engineering at the Chinese University of Hong Kong (CUHK).

He was a Research Intern with Microsoft Research Asia from October 2007 to March 2008. He was a Research Assistant with the Department of Electronic Engineering, CUHK, from November 2008 to July 2009. His research interests include image/video quality assessment, super-resolution, restoration, and compression.

Mr. Ma received the Best Paper Award in the Pacific-Rim Conference on Multimedia (PCM) 2008.