Li, Q.: Information content weighting for perceptual image quality assessment. IEEE Image Proc. 20(5), 1185-1198

Article *in* IEEE Transactions on Image Processing · November 2010

DOI: 10.1109/TIP.2010.2092435 · Source: PubMed

CITATIONS
748

READS
388

2 authors, including:

Zhou Wang
University of Waterloo
224 PUBLICATIONS **43,511** CITATIONS

SEE PROFILE

Information Content Weighting for Perceptual Image Quality Assessment

Zhou Wang, *Member, IEEE*, and Qiang Li, *Member, IEEE*

***Abstract*—Many state-of-the-art perceptual image quality assessment (IQA) algorithms share a common two-stage structure: local quality/distortion measurement followed by pooling. While significant progress has been made in measuring local image quality/distortion, the pooling stage is often done in ad-hoc ways, lacking theoretical principles and reliable computational models. This paper aims to test the hypothesis that when viewing natural images, the optimal perceptual weights for pooling should be proportional to local information content, which can be estimated in units of bit using advanced statistical models of natural images. Our extensive studies based upon six publicly-available subject-rated image databases concluded with three useful findings. First, information content weighting leads to consistent improvement in the performance of IQA algorithms. Second, surprisingly, with information content weighting, even the widely criticized peak signal-to-noise-ratio can be converted to a competitive perceptual quality measure when compared with state-of-the-art algorithms. Third, the best overall performance is achieved by combining information content weighting with multiscale structural similarity measures.**

***Index Terms*—Gaussian scale mixture (GSM), image quality assessment (IQA), pooling, information content measure, peak signal-to-noise-ratio (PSNR), structural similarity (SSIM), statistical image modeling.**

I. Introduction

In RECENT years, there has been an increasing interest in developing objective image quality assessment (IQA) methods that can automatically predict human behaviors in evaluating image quality [1]–[3]. Such perceptual IQA measures have broad applications in the evaluation, control, design and optimization of image acquisition, communication, processing and display systems. Depending upon the availability of a "perfect quality" reference image, they may be classified into full-reference (FR, where the reference image is fully accessible when evaluating the distorted image), reduced-reference (RR, where only partial information about the reference image is available) and no-reference (NR, where no access to the reference image is allowed) algorithms [3].

Many state-of-the-art IQA measures (especially FR algorithms) adopted a common two-stage structure, as illustrated in Fig. 1. In the first stage, image quality/distortion is evaluated locally, where the locality may be defined in space, scale (or spatial frequency) and orientation. For example, spatial domain methods such as the mean squared error (MSE) and the structural similarity (SSIM) index [4], [5] compute pixel- or patch-wise distortion/quality measures in space, while block-discrete cosine transform [6] and wavelet-based [7]–[11] approaches define localized quality/distortion measures across scale, space and orientation. Such localized measurement approaches are consistent with our current understanding about the human visual system (HVS), where it has been found that the responses of many neurons in the primary visual cortex are highly tuned to the stimuli that are "narrow-band" in frequency, space and orientation [12]. The local measurement process typically results in a quality/distortion map defined either in the spatial domain or in the transform domain (e.g., wavelet subbands). A spatial domain example is shown in Fig. 2. To assess the quality of a JPEG compressed image (b) given a reference image (a), two local quality/distortion measures, absolute error and the SSIM index, were computed, resulting an absolute error map (c) and an SSIM map (d). Careful inspection shows that the SSIM index better reflects the spatial variations of perceived image quality. For example, the blockiness in the sky is clearly indicated in Fig. 2(d) but not in Fig. 2(c). To convert such quality/distortion maps into a single quality score, a pooling algorithm is employed in the second stage of the IQA algorithm.

In the literature, significant progress has been made in the design of the first stage, i.e., local quality measurement [1]–[3], but much less is understood about the pooling stage. The potential of spatial pooling has been demonstrated by experimenting with different pooling strategies [13] or optimizing spatially varying weights to maximize the correlation between objective and subjective image quality ratings [14]. A common hypothesis underlying nearly all existing schemes is that the pooling strategy should be correlated with human visual fixation or visual region-of-interest detection. This is supported by a number of interesting recent studies [14]–[16], where it has been shown that sizable performance gain can be obtained by combining objective local quality measures with subjective human fixation or region-of-interest detection data. In practice, however, the subjective data is not available, and the pooling stage is often done in simplistic or ad-hoc ways, lacking theoretical principles as the basis for the development of reliable computational models.

Manuscript received January 21, 2010; revised June 07, 2010 and September 06, 2010; accepted November 04, 2010. Date of publication November 15, 2010; date of current version April 15, 2011. This work was supported in part by Natural Sciences and Engineering Research Council of Canada in the forms of Discovery, Strategic and Collaborative Research and Development (CRD) Grants, and in part by an Ontario Early Researcher Award. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alex C. Kot.

Z. Wang is with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada (e-mail: zhouwang@ieee.org).

Q. Li is with Media Excel Inc., Austin, TX, 78759 USA.

Digital Object Identifier 10.1109/TIP.2010.2092435

The existing pooling approaches can be roughly categorized in the following ways.

- *Minkowski pooling*
  Let q_i be the local quality/distortion value at the ith location in the quality/distortion map. The Minkowski summation is given by

$$\mathrm{Q} = \frac{1}{N}\sum_{i}^{N} q_i^p \tag{1}$$

  where N is the total number of samples in the map, and p is the Minkowski exponent. To give a specific example, let q_i represent the absolute error as in Fig. 2(c), then (1) is directly related to the l_p norm (subject to a monotonic nonlinearity). As special cases, $p = 1$ corresponds to the mean absolute error (MAE), and $p = 2$ to the MSE. As p increases, more emphasis is shifted to the high distortion regions. Intuitively, this makes sense because when most distortions in an image is concentrated in a small region of an image, humans tend to pay more attentions to this low quality region and give an overall quality score lower than direct average of the quality map [13]. In the extreme case $p = \infty$, it converges to $\max_i\{p_i\}$, i.e., the measure is completely determined by the highest distortion point. In practice, the value of p typically ranges from 1 to 4 [5]–[10]. In [13], it was shown that Minkowski pooling can help improve the performance of IQA algorithms, but the best p value depends upon the underlying local metric q_i and there is no simple method to derive it.
- *Local quality/distortion-based pooling*
  The intuitive idea that more emphasis should be put at high distortion regions can be implemented in a more straightforward way by local qulaity/distoriton-based pooling. This can be done by using a nonuniform weighting approach, where the weight may be determined by an error visibility detection map [17]. It may also be computed using the local quality/distortion measure itself [13], such that the overall quality/distortion measure is given by

$$\mathrm{Q} = \frac{\sum_{i=1}^{N} w(q_i) q_i}{\sum_{i=1}^{N} w(q_i)} \tag{2}$$

  where the weighting function $w(\cdot)$ is monotonically increasing when q_i is a distortion measure (i.e., larger value indicates higher distortion), and monotonically decreasing when q_i is a quality measure (i.e., larger value indicates higher quality). Another method to assign more weights to low quality regions is to sort all q_i values and use a small percentile of them that correspond to the lowest quality regions. For example, in [18] and [19], the worst 5% or 6% distortion values were employed in computing the overall quality scores. Local quality/distortion-based pooling has been shown to be effective in improving IQA performance, as reported in [13], [19], though the implementations are often heuristic (for example, in the selection of the weighting function $w(\cdot)$ and the percentile), without theoretical guiding principles.

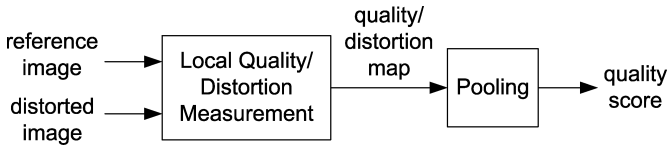


Fig. 1. Two-stage structure of IQA systems.

- *Saliency-based pooling*
  Here we use "saliency" as a general term that represents low-level local image features that are of perceptual significance (as opposed to high-level components such as human faces). The motivation behind saliency-based pooling approaches is that visual attention is attracted to distinctive saliency features and, thus, more importance should be given to the associated regions in the image. A saliency map $\{w_i\}$, created by computing saliency at each image location, can be used as a visual attention predictor, as well as a weighting function for IQA pooling as follows:

$$\mathrm{Q} = \frac{\sum_{i}^{N} w_i q_i}{\sum_{i}^{N} w_i}. \tag{3}$$

  Given an infinite number of possible saliency features, the question is what saliency should be used to create w_i. This can range from simple features such as local variance [13] or contrast [20] to sophisticated computational models based upon automatic point of gaze predictions from low-level vision features [19], [21]–[24]. It has also been found that motion information is another useful feature to use in the pooling stage of video quality assessment algorithms [25]–[27].
- *Object-based pooling*
  Different from low-level vision based saliency approaches, object-based pooling methods resort to high-level cognitive vision based image understanding algorithms that help detect and/or segment significant regions from the image. A similar weighting approach as in (3) may be employed, just that the weight map w_i is generated from object detection or segmentation algorithms. More weights can be assigned to segmented foreground objects [28] or on human faces [26], [29]–[31]. Although object-based weighting has demonstrated improved performance for specific scenarios (e.g., when the image contains distinguishable human faces), they may not be easily applied to general situations where it may not always be an easy task to find distinctive objects that attract visual attention.

In summary, all of the previous pooling strategies are well motivated and have achieved certain levels of success. Combinations of different strategies have also shown to be a useful approach [19], [25], [26], [31]. However, the existing pooling algorithms tend to be ad-hoc, and model parameters are often set by experimenting with subject-rated image databases. What are lacking are not heuristic tricks but general theoretical principles that are not only qualitative sensible but also quantitative manageable, so that reliable computational models for pooling can be derived.

In this research, we look at the IQA pooling problem from an information theoretic point of view. The general belief is that the HVS is an optimal information extractor, as widely

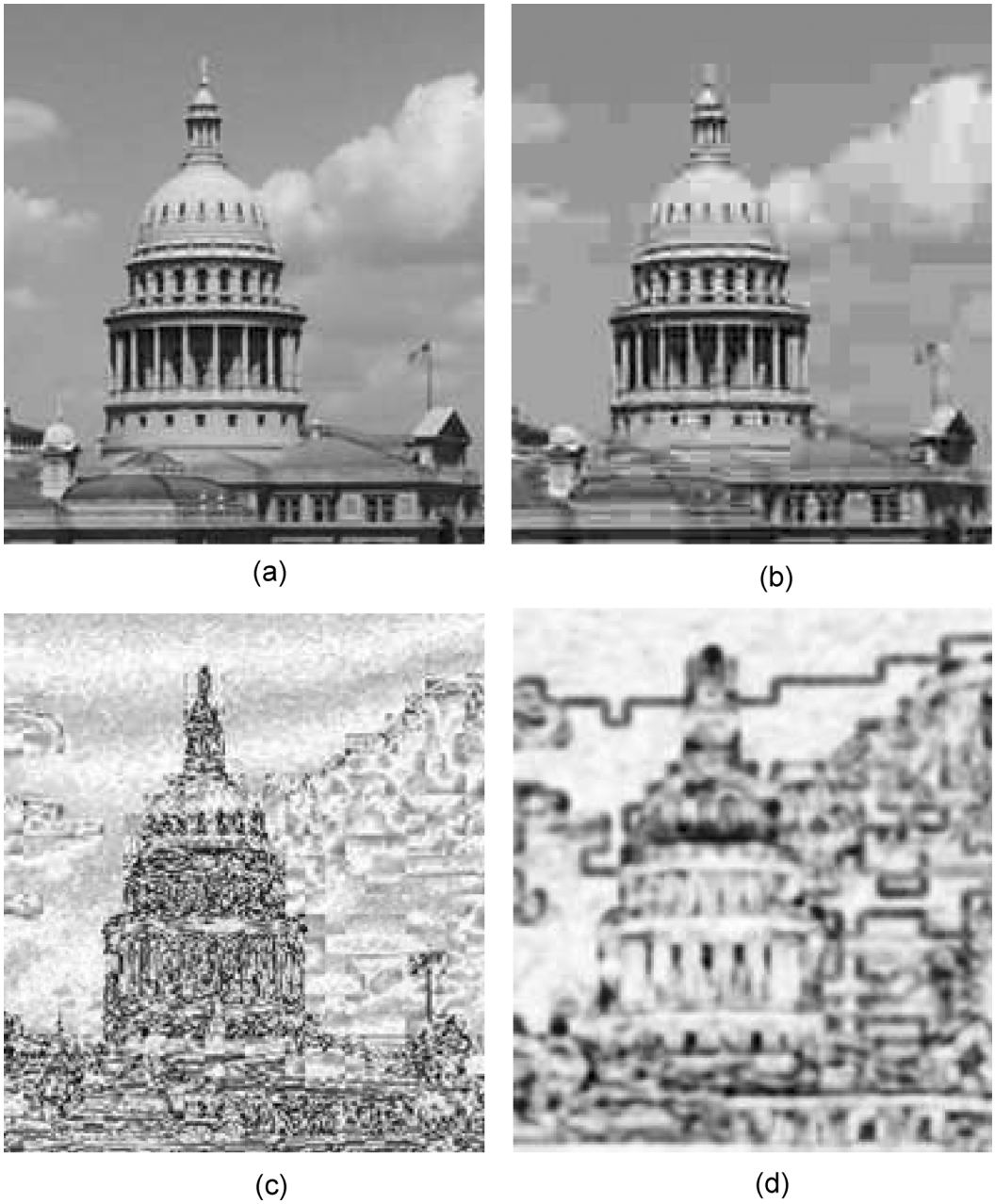


Fig. 2. (a) Original image. (b) Distorted image (by JPEG compression). (c) Absolute error map—brighter indicates better quality (smaller absolute difference). (d) SSIM index map—brighter indicates better quality (larger SSIM value).

hypothesized in computational vision science [32]. To achieve such optimality, the image components that contain more information content would attract more visual attention [33]. Using statistical information theory, the local information content can be quantified in units of bit, provided that a statistical image model is available. The local information content measure can then be employed for IQA weighting. In essence, our approach is saliency-based, but the resulting weighting function also has interesting connections with quality/distortion-based pooling method, which we will discuss later in Section II. Information theoretic methods are by no means new for IQA. In fact, our work is inspired by the success of the visual information fidelity (VIF) method [34], though VIF was not originally proposed for pooling purpose. In [27], based upon statistical models of Bayesian motion perception [35], motion information content and perceptual uncertainty were computed for video quality assessment. In our preliminary work [13], simple local information-based weighting demonstrated promising results for improving IQA performance. In this paper, we build our information content weighting method upon advanced statistical image models and combine it with multiscale IQA methods. This results in superior performance in our extensive tests using six independent databases, which in turn, provides strong support of our general hypothesis.

II. Information Content Weighting

The computation of image information content relies on good statistical image models. In [13], a rather crude spatial

domain local Gaussian model is assumed for spatial pooling of IQA. Inspired by several recent successful approaches in image denoising [36] and IQA [34], [37], [38], here we adopt the Gaussian scale mixture (GSM) model for natural images. As in many other image models, to reduce the high dimensionality of natural images, a Markov assumption is made that the probability density of a pixel (or a transform coefficient) is fully determined by the pixels (coefficients) within a spatial (and/or scale) neighborhood. The remaining task is, thus, the statistical modeling of groups of neighboring pixels (or coefficients). GSM has found to be a powerful model for this purpose [39], where the neighborhood is typically composed of a set of neighboring coefficients in a multiresolution image transform domain. It has been shown that the GSM framework can be easily adapted to account for the marginal statistics of multiresolution transform coefficients of natural images, where the density exhibits strong non-Gaussianity, with sharp peak at zero and heavy tails [32]. Meanwhile, GSM is also effective in describing the amplitude-dependency between neighboring coefficients [39].

Let R be a length-K column vector that contains a group of K neighboring transform coefficients (e.g., wavelet or Laplacian pyramid transform [40] coefficients). We model it as a GSM, which can be expressed as a product of two independent components

$$R = s\,U \tag{4}$$

where U is a zero-mean Gaussian vector with covariance matrix $\mathbf{C}_U$, and s is called a mixing multiplier. The general form of GSM allows s to be a random variable that has a certain distribution in a continuous scale. To simplify the computation, we assume that s only takes a fixed value at each location (but varies over space and scale). The benefit of this simplification is that when s is fixed and given, R is simply a zero-mean Gaussian vector with covariance

$$\mathbf{C}_{R|\,s} = s^2\mathbf{C}_U. \tag{5}$$

An important concept that we learned from the information theoretical IQA approaches [34], [37] is that the information contained in an image is not equated with the amount of information perceived by the visual system. The mutual information between the images before and after the visual perceptual channel provides a more useful measure. Following this idea, we propose a model to compute perceptual information content, which is illustrated in Fig. 3. First, the reference signal R passes through a distortion channel, resulting in a distorted signal D

$$D = gR + V = gsU + V \tag{6}$$

where the distortion is modeled based upon a gain factor g followed by additive independent Gaussian noise contamination V with covariance $\mathbf{C}_V = \sigma_v^2\mathbf{I}$ (where $\mathbf{I}$ represents the identity matrix). Although this model seems to be over simplistic in capturing all potential types of distortions such as blocking and ringing artifacts that often appear in compressed images, it was claimed to achieve a reasonable balance in terms of the level of perceptual annoyance across distortion types [34]. This was demonstrated empirically in [34] using an image synthesis approach, where images under different types of distortions were compared with synthesized distortion images using the local attenuation/noise model. Although the real and synthesized distorted images look different in terms of the types of artifacts, the synthesized images reproduced more reasonably balanced perceptual annoyance than an additive noise-only distortion model [34]. Stronger and more theoretical justifications of this distortion model are still yet to be discovered.

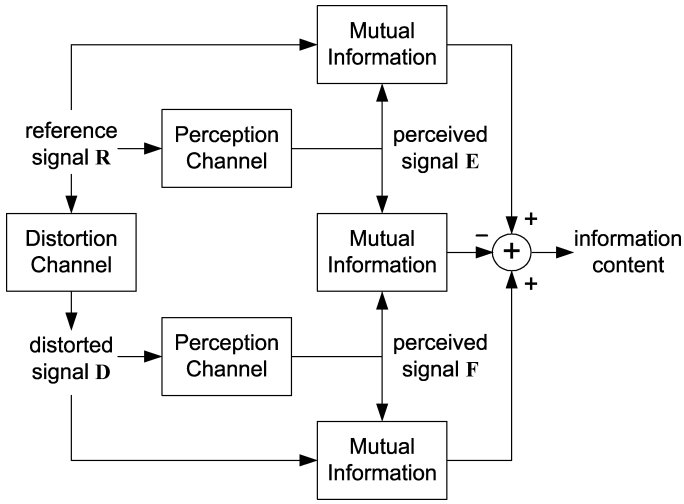


Fig. 3. Diagram for computing information content.

Next, both the reference and distorted signals pass through a perceptual visual noise channel

$$E = R + N_1 = sU + N_1 \tag{7}$$

$$F = D + N_2 = gsU + V + N_2 \tag{8}$$

where N_1 and N_2 are assumed to be independent white Gaussian noise with diagonal covariance $\mathbf{C}_{N_1} = \mathbf{C}_{N_2} = \sigma_n^2\mathbf{I}$. This simple one-parameter (σ_n) visual distortion model aims to capture the lumped uncertainty of the visual system [34]. Similar to (5), we can then compute the covariance matrices of D, E, and F as

$$\mathbf{C}_D = g^2s^2\mathbf{C}_U + \sigma_v^2\mathbf{I} \tag{9}$$

$$\mathbf{C}_E = s^2\mathbf{C}_U + \sigma_n^2\mathbf{I} \tag{10}$$

$$\mathbf{C}_F = g^2s^2\mathbf{C}_U + \sigma_v^2\mathbf{I} + \sigma_n^2\mathbf{I}. \tag{11}$$

Since all the computation in the rest of this section assumes a fixed and known multiplier s, for notational convenience, we drop the conditional notation "$|s$" in all the derivations.

Based upon the approach given in [34], at each location, the information of the original and distorted images perceived by the visual system can be computed by the mutual information $I(R;E)$ and $I(D;F)$, respectively. Here we move one step further to estimate the total perceptual information content from both images. More specifically, we compute the sum of $I(R;E)$ and $I(D;F)$ minus the common information shared between E and F. This results in a total information content weight measure given by

$$w = I(R;E) + I(D;F) - I(E;F). \tag{12}$$

To compute (12), it is useful to be aware that R, D, E and F are all Gaussian for given fixed s. As a result, the mutual information evaluations, $I(R;E)$, $I(D;F)$ and $I(E;F)$, can be calculated based upon the determinants of the covariances [41] by

$$I(R;E) = \frac{1}{2}\log_2\left[\frac{|\mathbf{C}_R||\mathbf{C}_E|}{|\mathbf{C}_{(R,E)}|}\right] \tag{13}$$

$$I(D;F) = \frac{1}{2}\log_2\left[\frac{|\mathbf{C}_D||\mathbf{C}_F|}{|\mathbf{C}_{(D,F)}|}\right] \tag{14}$$

$$I(E;F) = \frac{1}{2}\log_2\left[\frac{|\mathbf{C}_E||\mathbf{C}_F|}{|\mathbf{C}_{(E,F)}|}\right] \tag{15}$$

where

$$\mathbf{C}_{(R,E)} = \begin{bmatrix}\mathbf{C}_R & \mathbf{C}_{RE}\\ \mathbf{C}_{ER} & \mathbf{C}_E\end{bmatrix} \tag{16}$$

$$\mathbf{C}_{(D,F)} = \begin{bmatrix}\mathbf{C}_D & \mathbf{C}_{DF}\\ \mathbf{C}_{FD} & \mathbf{C}_F\end{bmatrix} \tag{17}$$

$$\mathbf{C}_{(E,F)} = \begin{bmatrix}\mathbf{C}_E & \mathbf{C}_{EF}\\ \mathbf{C}_{FE} & \mathbf{C}_F\end{bmatrix}. \tag{18}$$

Equation (16) can be simplified based upon the fact that

$$\mathbf{C}_{RE} = \mathbf{C}_{ER} = \mathcal{E}\{RE^T\} = s^2\mathbf{C}_U = \mathbf{C}_R \tag{19}$$

where $\mathcal{E}$ is the expectation operator and we have used the fact that U and N_1 are independent. This leads to

$$|\mathbf{C}_{(R,E)}| = \left|\begin{bmatrix}\mathbf{C}_R & \mathbf{C}_R\\ \mathbf{C}_R & \mathbf{C}_E\end{bmatrix}\right| = |\sigma_n^2\mathbf{C}_R|. \tag{20}$$

Similarly, we can derive

$$\mathbf{C}_{DF} = \mathbf{C}_{FD} = g^2s^2\mathbf{C}_U + \sigma_V^2\mathbf{I} = \mathbf{C}_D \tag{21}$$

$$\mathbf{C}_{EF} = \mathbf{C}_{FE} = g\,s^2\mathbf{C}_U \tag{22}$$

and

$$|\mathbf{C}_{(D,F)}| = \left|\begin{bmatrix}\mathbf{C}_D & \mathbf{C}_D\\ \mathbf{C}_D & \mathbf{C}_F\end{bmatrix}\right| = |\sigma_n^2\mathbf{C}_D|. \tag{23}$$

Combining (12), (13), (14), (15), (20), and (23), we can simplify our information content weight computation to the following expression:

$$w = \frac{1}{2}\log_2\left[\frac{|\mathbf{C}_{(E,F)}|}{\sigma_n^{4K}}\right]. \tag{24}$$

Plug (22), (10), and (11) into (18), we have

$$|\mathbf{C}_{(E,F)}| = |\left[\left(\sigma_v^2+\sigma_n^2\right)s^2 + \sigma_n^2g^2s^2\right]\mathbf{C}_U + \sigma_n^2\left(\sigma_v^2+\sigma_n^2\right)\mathbf{I}|. \tag{25}$$

To compute the determinant of $\mathbf{C}_{(E,F)}$, it is useful to apply an eigenvalue decomposition to the covariance matrix $\mathbf{C}_U = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where $\mathbf{Q}$ is an orthogonal matrix, and $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues λ_k for $k = 1,\ldots,K$ along its diagonal entries. Equation (25) can then be expressed as

$$|\mathbf{C}_{(E,F)}| = |\mathbf{Q}\left\{\left[\sigma_v^2+(1+g^2)\sigma_n^2\right]s^2\mathbf{\Lambda} + \sigma_n^2\left(\sigma_v^2+\sigma_n^2\right)\mathbf{I}\right\}\mathbf{Q}^T|. \tag{26}$$

Since $\mathbf{Q}$ is orthogonal and the expression between the two $\mathbf{Q}$ matrices in (26) is a diagonal matrix, the determinant of $\mathbf{C}_{(E,F)}$ can be easily computed as

$$|\mathbf{C}_{(E,F)}| = \prod_{k=1}^{K}\left\{\left[\sigma_v^2+(1+g^2)\sigma_n^2\right]s^2\lambda_k + \sigma_n^2\left(\sigma_v^2+\sigma_n^2\right)\right\}. \tag{27}$$

Plug this into (24) and simplify the expression, we obtain

$$w = \frac{1}{2}\sum_{k=1}^{K}\log_2\left\{1+\frac{\sigma_v^2}{\sigma_n^2}+\left(\frac{\sigma_v^2}{\sigma_n^4}+\frac{1+g^2}{\sigma_n^2}\right)s^2\lambda_k\right\}. \tag{28}$$

Although the derivation mentioned here is completely based upon evaluations of local information content, the resulting weight function (28) shows some interesting connections with local distortion/quality-weighted pooling method described in Section I. In particular, based upon the distortion model (6), the variations from R to D are characterized by the gain factor g and the random distortion σ_v^2. Since g is a scale factor along the signal direction, it does not cause structural changes of the signal. Therefore, the structural distortions are essentially captured by σ_v^2. Note that the weight function (28) increases monotonically with σ_v^2. This implies that more weights are given to the regions with larger distortions, which is in line with the philosophy behind quality/distortion-weighted pooling.

To finish the computation in (28), we need to estimate a set of parameters, including $\mathbf{C}_U$, s^2, g and σ_v^2. As in [36], we estimate $\mathbf{C}_U$ using

$$\hat{\mathbf{C}}_U = \frac{1}{N}\sum_{i=1}^{N}R_iR_i^T \tag{29}$$

where N is the number of evaluation windows in the subband, and R_i is the ith neighborhood coefficient vector. This needs to be computed only once for each subband. The multiplier s is spatially varying and can be estimated using a maximum likelihood estimator [39]

$$\hat{s}^2 = \frac{1}{K}R^T\mathbf{C}_U^{-1}R. \tag{30}$$

Finally, the distortion parameters g and σ_v^2 can be obtained by least square regression that optimizes

$$\hat{g} = \arg\min_g \|D - gR\|_2^2. \tag{31}$$

Take derivative of the squared error function with respective to g and let it equal zero, we have

$$\hat{g} = \frac{R^TD}{R^TR}. \tag{32}$$

Substitute this into (6), we can estimate σ_v^2 using V^TV/K, which leads to

$$\hat{\sigma}_v^2 = \frac{1}{K}(D^TD - \hat{g}R^TD). \tag{33}$$

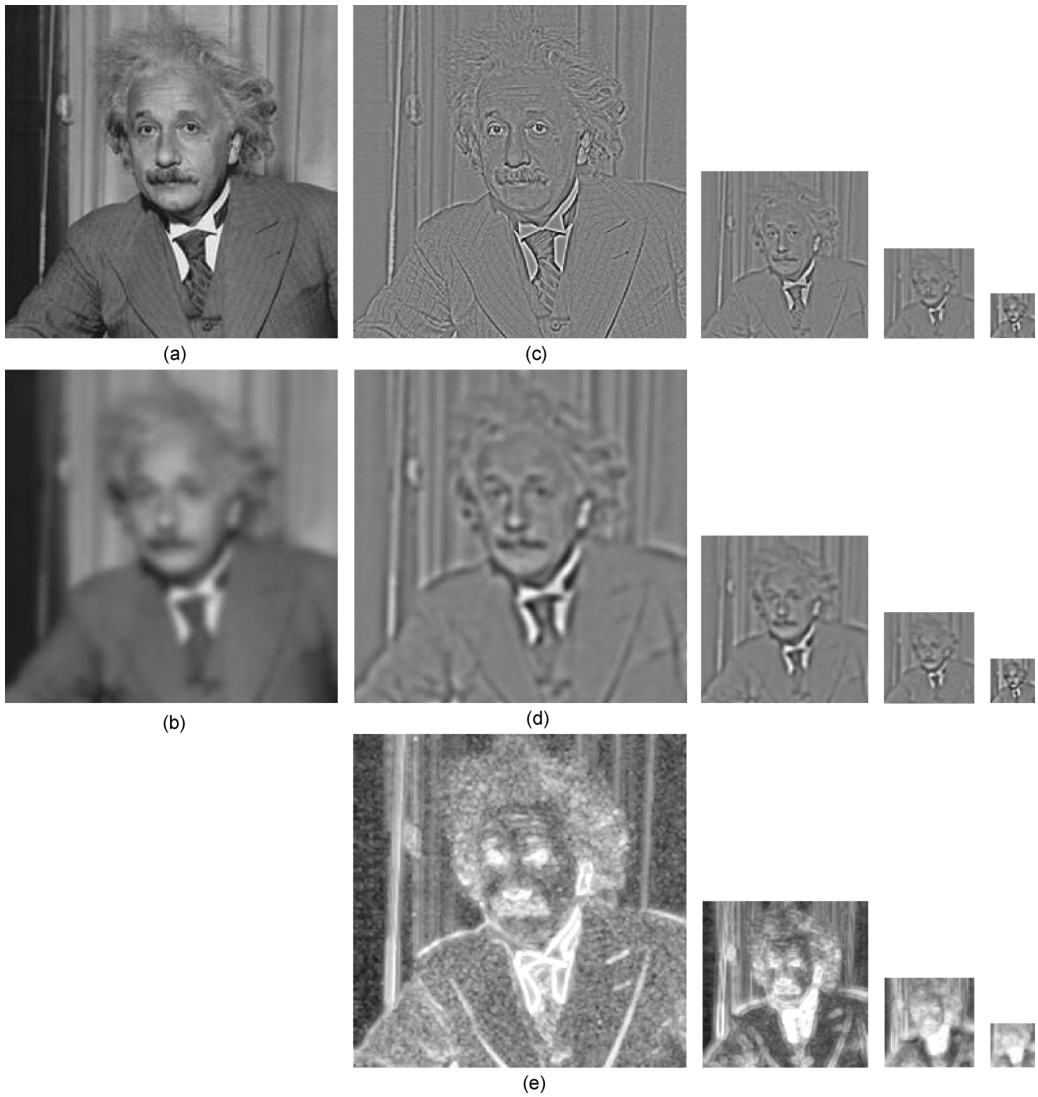


Fig. 4. Computation of local information content maps. (a),(b) Original and distorted images. (c),(d) Corresponding Laplacian pyramid subbands at four scales (enhanced for visualization). (e) Corresponding information content maps computed at four scales (enhanced for visualization). Brighter indicates larger information content.

When computing information content weights for real-world images, we first apply a five-scale Laplacian pyramid decomposition [40] to the original and distorted images, resectively. We then compute information content weight using a sliding window that runs across each subband, where at each location, the window includes 3×3 spatial neighborhood coefficients together with one parent coefficient (as a result, $K = 10$). This process results in an information content weight map for each scale. An example of the "Einstein" image is given in Fig. 4. By visually inspecting the reference and distorted images, we observe that the information content is distributed unevenly over space. For example, compared with the background, the eye regions and some sharp edge areas in the images are perceptually more informative. As expected, these observations are well represented by the information content maps, where brighter indicates more information content and, thus, higher visual importance in IQA.

III. IQA Algorithms

A. *Information Content Weighted PSNR*

Let x_i and y_i be the ith pixel in the original image $\mathbf{x}$ and the distorted image $\mathbf{y}$, respectively. The MSE and PSNR between the two images are given by

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{34}$$

$$\text{PSNR} = 10\log_{10}\left(\frac{L^2}{\text{MSE}}\right) \tag{35}$$

where N is the total number of pixels in the image and L is the maximum dynamic range. For 8 b/pixel gray-scale images, $L = 255$.

Here we define an information content weighted MSE (IW-MSE) and an information content weighted PSNR (IW-PSNR)

measures by incorporating the Laplacican pyramid transform [40] domain information content weights computed as in (28). Let $x_{j,i}$ and $y_{j,i}$ be the ith transform coefficients at the jth scale, and $w_{j,i}$ be the information content weight computed at the corresponding location, then we define IW-MSE as

$$\text{IW–MSE} = \prod_{j=1}^{M} \left[\frac{\sum_i w_{j,i}(x_{j,i} - y_{j,i})^2}{\sum_i w_{j,i}} \right]^{\beta_j} \tag{36}$$

where M is the number of scales, β_j is the weight given to the jth scale, and the weights are defined in similar ways as in the multiscale SSIM approach [42], which will be discussed in more detail in Section III-B. Analogous to MSE-PSNR conversion, IW-MSE can be converted to IW-PSNR by

$$\text{IW–PSNR} = 10 \log_{10} \left(\frac{L^2}{\text{IW–MSE}} \right). \tag{37}$$

B. Information Content Weighted MultiScale SSIM

The basic spatial domain SSIM algorithm [5] is based upon separated comparisons of local luminance, contrast and structure between an original and a distorted images. Given two local image patches $\mathbf{x}$ and $\mathbf{y}$ extracted from the original and distorted images, respectively, the luminance, contrast and structural similarities between them are evaluated as

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{38}$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{39}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{40}$$

respectively. Here, μ_x, σ_x and σ_{xy} represent the mean, standard deviation and cross-correlation evaluations, respectively. $C_1 = (K_1L)^2, C_2 = (K_2L)^2, C_3 = C_2/2$ are small constants that have been found to be useful in characterizing the saturation effects of the visual system at low luminance and contrast regions and stabilizing the performance of the measure when the denominators are close to zero. The local SSIM index is defined as the product of the three components, which gives

$$\text{SSIM}_{\text{local}} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{41}$$

When this local measurement is applied to an entire image using a sliding window approach, an SSIM quality map is created, as exemplified by Fig. 2(d). The overall SSIM value of the whole image is simply the average of the SSIM map.

It has been found that the performance of the previous single-scale SSIM algorithm depends upon the scale it is applied to [42] and [43]. In [42], a multiscale SSIM (MS-SSIM) approach was proposed that incorporates SSIM evaluations at different scales. Psychovisual experiments were carried out to find the relative weights between scales. Interestingly, the measured weight function peaks at middle-resolution scales and drops at both low- and high-resolution scales, consistent with the contrast sensitivity function extensively studied in the vision literature [12]. Let $\mathbf{x}_{j,i}$ and $\mathbf{y}_{j,i}$ be the ith local image patches (extracted from the ith evaluation window) at the jth scale, and let N_j be the number of evaluation windows in the scale, then the jth scale SSIM evaluation is computed as

$$\text{SSIM}_j = \frac{1}{N_j} \sum_i c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) \tag{42}$$

for $j = 1, \ldots, M-1$, and

$$\text{SSIM}_j = \frac{1}{N_j} \sum_i l(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) \tag{43}$$

for $j = M$. The overall MS-SSIM measure is defined as

$$\text{MS–SSIM} = \prod_{j=1}^{M} (\text{SSIM}_j)^{\beta_j} \tag{44}$$

where the β_j values were obtained through psychophysical measurement [42].

By combining information content weighting with multiscale SSIM, we define an information content weighted SSIM measure (IW-SSIM). Let $w_{j,i}$ be the information content weight computed at the ith spatial location in the jth scale using (28), the jth scale IW-SSIM measure is defined as

$$\text{IW–SSIM}_j = \frac{\sum_i w_{j,i}\, c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i})}{\sum w_{j,i}} \tag{45}$$

for $j = 1, \ldots, M-1$, and

$$\text{IW–SSIM}_j = \frac{1}{N_j} \sum_i l(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) \tag{46}$$

for $j = M$. The final overall IW-SSIM measure is then computed as

$$\text{IW–SSIM} = \prod_{j=1}^{M} (\text{IW–SSIM}_j)^{\beta_j} \tag{47}$$

using the same set of scale weights β_j's as in MS-SSIM.

The proposed IW-PSNR and IW-SSIM algorithms do not involve any training process or any new parameters for tuning. All parameters are inherited from previous publications. These include $K_1 = 0.01$ and $K_2 = 0.03$ from [5]; $\sigma_n = 0.4$ from [34]; $M = 5$ from [42]; and the fine-to-coarse scale weights $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\} = \{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ from [42].

C. Interpretation of VIF Based Upon Information Content Weighting

Based upon the interpretation in its original publication, the VIF algorithm [34] does not seem to fit into the two-stage framework shown in Fig. 1, because the information content is summed over the entire image space before the fidelity ratio is computed

$$\text{VIF} = \frac{\sum_i I(R_i; F_i \,|\, s_i)}{\sum_i I(R_i; E_i \,|\, s_i)}. \tag{48}$$

Here we show that with some simple transformations, VIF indeed can be nicely interpreted using the same two-stage framework. Specifically, we can write

$$\text{VIF} = \frac{\sum_i w_i \text{VIF}_i}{\sum_i w_i} \tag{49}$$

where we have defined a *local* VIF measure (which follows the same philosophy as the general VIF concept [34])

$$\text{VIF}_i = \frac{I(R_i; F_i \,|\, s_i)}{I(R_i; E_i \,|\, s_i)} \tag{50}$$

and a weighting function

$$w_i = I(R_i; E_i \,|\, s_i). \tag{51}$$

Interestingly, this weight definition is essentially an information content measure, although different from what we use in our approach [as in (12)].

IV. Validation and Comparison

We validate the proposed IW-PSNR and IW-SSIM measures and compare them with 13 other algorithms.

- PSNR, which has a wide usage in the image processing literature. It also provides useful baseline comparisons.
- SSIM [5], MS-SSIM [42], visual signal-to-noise ratio (VSNR) [44], VIF [34], PSNR-HVS-M [45], and most apparent distortion (MAD) [17], which are state-of-the-art algorithms that have demonstrated competitive performance. They are also available online [43], [46]–[49] that facilitate repeatable experimental verifications.
- Distortion-weighted PSNR (DW-PSNR) and distortion weighted SSIM (DW-SSIM), which were implemented by ourselves to provide direct comparisons between quality/distortion- and information content-weighted approaches. Specifically, the weighting approach of (2) is adopted, where the function $w(\cdot)$ is defined as $w(x) = |x|^8$ and $w(x) = 1/|x|$ for DW-PSNR and DW-SSIM, respectively, which maximize the performance of DW-based weighting approaches according to the empirical results presented in [13].
- Contrast weighted PSNR (CTW-PSNR) and contrast weighted SSIM (CTW-SSIM), where we replaced information content weighting with a local contrast-based weighting approach to facilitate a straightforward comparison of the two pooling approaches. In particular, the scale-dependent contrast measure proposed in [20] was adopted.
- Saliency weighted PSNR (SW-PSNR) and saliency weighted SSIM (SW-SSIM), where saliency maps computed using the model proposed in [21] (using the SaliencyToolbox presented in [50], [51]) were employed to create the local weighting function. This helps make direct comparisons between the weighting approaches based upon information content measures and widely accepted saliency measures designed to predict human fixations.

To the best of our knowledge, there are currently six publicly-available subject-rated image databases that are widely recognized in the IQA research community. We include all of them in our algorithm validation and comparisons. Since the construction of our algorithms does not require training or parameter tuning, all image databases are used for testing only. These databases include those shown in the following.

- The LIVE database [46] was developed at The University of Texas at Austin. It contains seven data sets of 982 subject-rated images, including 779 distorted images created from 29 original images with five types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression (2 sets); b) JPEG compression (2 sets); c) White noise contamination (1 set); d) Gaussian blur (1 set); and e) fast fading channel distortion of JPEG2000 compressed bitstream (1 set). The subjective test was carried out with each data set individually. A cross-comparison set that mixes images from all distortion types is then used to help align the subject scores across data sets. The subjective scores of all images are then adjusted accordingly. The alignment process is rather crude. However, the aligned subjective scores (all data) are still very useful references, which are particularly important for testing general-purpose IQA algorithms, for which cross-distortion comparisons are highly desirable.
- The Cornell-A57 database [52] was created at Cornell University. It contains 54 distorted images with six types of distortions including a) quantization of the LH subbands of a 5-level discrete wavelet transform, where the subbands were quantized via uniform scalar quantization with step sizes chosen such that the root mean-squared (RMS) contrast of the distortions was equal; b) additive Gaussian white noise; c) baseline JPEG compression; d) JPEG2000 compression without visual frequency weighting; e) JPEG2000 compression with the dynamic contrast-based quantization algorithm, which applies greater quantization to the fine spatial scales relative to the coarse scales in an attempt to preserve global precedence; and f) blurring by using a Gaussian filter.
- The IVC database [53], [54] was developed at Ecole Polytechnique de l'Universite de Nantes. It includes 185 distorted images generated from ten original images. There are four types of distortions that are a) JPEG compression; b) JPEG2000 compression; c) Local adaptive resolution (LAR) coding; and d) Blurring.
- The Toyama-MICT database [55] was created at Toyama University. It contains 196 images, including 168 distorted images generated by JPEG and JPEG2000 compression.
- The Tampere Image Database 2008 (TID2008) [56], [57] was developed with a joint international effort between Finland, Italy, and Ukraine. It includes 1700 distorted images generated from 25 reference images with 17 distortion types at four distortion levels. The types of distortions include: a) Additive Gaussian noise; b) Additive noise in color components is more intensive than additive noise in the luminance component; c) Spatially correlated noise; d) Masked noise; e) High frequency noise; f) Impulse noise; g) Quantization noise; h) Gaussian blur; i) Image denoising; j) JPEG compression; k) JPEG2000 compression; l) JPEG transmission errors; m) JPEG2000 transmission errors; n) Non eccentricity pattern noise;

TABLE I
PERFORMANCE COMPARISONS OF 15 IQA ALGORITHMS ON SIX PUBLICLY AVAILABLE IMAGE DATABASES

LIVE Database (779 images) [46]						Cornell A57 Database (54 images) [52]					
Model	PLCC	MAE	RMS	SRCC	KRCC	Model	PLCC	MAE	RMS	SRCC	KRCC
PSNR	0.8723	10.51	13.36	0.8756	0.6865	PSNR	0.6347	0.1607	0.1899	0.6189	0.4309
SSIM [5]	0.9449	6.933	8.946	0.9479	0.7963	SSIM [5]	0.8017	0.1209	0.1469	0.8066	0.6058
MS-SSIM [42]	0.9489	6.698	8.619	0.9513	0.8044	MS-SSIM [42]	0.8603	0.1007	0.1253	0.8414	0.6478
VSNR [44]	0.9229	8.089	10.52	0.9271	0.7610	VSNR [44]	**0.9146**	**0.0809**	**0.0994**	**0.9355**	**0.8031**
VIF [34]	**0.9598**	**6.148**	**7.667**	**0.9632**	**0.8270**	VIF [34]	0.6157	0.1397	0.1937	0.6223	0.4589
PSNR-HVS-M [45]	0.9251	7.966	10.37	0.9295	0.7659	PSNR-HVS-M [45]	0.8748	0.0923	0.1190	**0.8962**	**0.7261**
MAD [17]	0.9394	7.293	9.368	0.9438	0.7920	MAD [17]	0.8816	0.0942	0.1160	0.8645	0.6702
DW-PSNR	0.9166	7.946	10.92	0.9188	0.7583	DW-PSNR	0.8699	0.1037	0.1212	0.8371	0.6436
CTW-PSNR	0.9349	7.204	9.695	0.9364	0.7855	CTW-PSNR	0.8889	0.0942	0.1126	0.8816	0.6995
SW-PSNR	0.9274	7.777	10.22	0.9277	0.7676	SW-PSNR	0.8955	0.0943	0.1094	0.8690	0.6884
IW-PSNR	0.9329	7.332	9.838	0.9328	0.7800	IW-PSNR	0.8974	**0.0879**	0.1084	0.8759	0.6967
DW-SSIM	**0.9556**	**6.212**	**8.047**	**0.9570**	**0.8197**	DW-SSIM	0.8758	0.0976	0.1186	0.8431	0.6436
CTW-SSIM	0.9413	7.312	9.220	0.9477	0.7964	CTW-SSIM	0.8809	0.0943	0.1163	0.8624	0.6786
SW-SSIM	0.9401	7.314	9.317	0.9438	0.7887	SW-SSIM	0.8918	0.0911	0.1112	0.8851	0.6926
IW-SSIM	0.9522	6.470	8.347	0.9567	0.8175	IW-SSIM	**0.9034**	0.0892	**0.1054**	0.8709	0.6842

IVC Database (185 images) [53], [54]						Toyama-MICT Database (168 images) [55]					
Model	PLCC	MAE	RMS	SRCC	KRCC	Model	PLCC	MAE	RMS	SRCC	KRCC
PSNR	0.6719	0.7191	0.9023	0.6884	0.5218	PSNR	0.6329	0.7817	0.9689	0.6132	0.4443
SSIM [5]	**0.9119**	**0.3777**	**0.4999**	0.9018	0.7223	SSIM [5]	0.8887	0.4386	0.5738	0.8794	0.6939
MS-SSIM [42]	0.9108	0.3813	0.5029	0.8980	0.7203	MS-SSIM [42]	0.8927	0.4328	0.5640	0.8874	0.7029
VSNR [44]	0.7904	0.5860	0.7463	0.7993	0.6053	VSNR [44]	0.8705	0.4654	0.6159	0.8608	0.6745
VIF [34]	0.9028	0.4104	0.5239	0.8964	0.7158	VIF [34]	0.9138	0.4038	0.5084	0.9077	0.7315
PSNR-HVS-M [45]	0.8788	0.4614	0.5815	0.8832	0.6935	PSNR-HVS-M [45]	0.8406	0.5541	0.6777	0.8480	0.6568
MAD [17]	0.8741	0.4728	0.5918	**0.9150**	**0.7406**	MAD [17]	0.9116	0.3951	0.5145	0.9086	0.7354
DW-PSNR	0.8431	0.5307	0.6552	0.8513	0.6543	DW-PSNR	0.7824	0.6132	0.7793	0.7824	0.5915
CTW-PSNR	0.8854	0.4657	0.5663	0.8925	0.7042	CTW-PSNR	0.8449	0.5474	0.6694	0.8556	0.6614
SW-PSNR	0.8760	0.4753	0.5875	0.8813	0.6938	SW-PSNR	0.8042	0.5945	0.7437	0.8140	0.6247
IW-PSNR	0.8963	0.4403	0.5403	0.8998	0.7165	IW-PSNR	0.8380	0.5567	0.6829	0.8475	0.6508
DW-SSIM	0.9090	0.3965	0.5078	0.8990	0.7156	DW-SSIM	**0.9278**	**0.3647**	**0.4426**	**0.9301**	**0.7704**
CTW-SSIM	0.8872	0.4368	0.5621	0.8826	0.6908	CTW-SSIM	0.9116	0.3972	0.5145	0.9047	0.7291
SW-SSIM	0.8768	0.4385	0.5858	0.8681	0.6826	SW-SSIM	0.8975	0.4384	0.5518	0.8885	0.7046
IW-SSIM	**0.9231**	**0.3694**	**0.4686**	**0.9125**	**0.7339**	IW-SSIM	**0.9248**	**0.3677**	**0.4761**	**0.9202**	**0.7537**

TID2008 Database (1700 images) [56], [57]						CSIQ Database (866 images) [58]					
Model	PLCC	MAE	RMS	SRCC	KRCC	Model	PLCC	MAE	RMS	SRCC	KRCC
PSNR	0.5223	0.8683	1.1435	0.5531	0.4027	PSNR	0.7512	0.1366	0.1733	0.8058	0.6084
SSIM [5]	0.7732	0.6546	0.8511	0.7749	0.5768	SSIM [5]	0.8612	0.0992	0.1334	0.8756	0.6907
MS-SSIM [42]	**0.8451**	**0.5578**	**0.7173**	**0.8542**	**0.6568**	MS-SSIM [42]	0.8991	0.0870	0.1149	0.9133	0.7393
VSNR [44]	0.6820	0.6908	0.9815	0.7046	0.5340	VSNR [44]	0.7355	0.1335	0.1779	0.8109	0.6248
VIF [34]	0.8090	0.5990	0.7888	0.7496	0.5863	VIF [34]	**0.9277**	**0.0743**	**0.0980**	0.9195	**0.7537**
PSNR-HVS-M [45]	0.5519	0.8036	1.1190	0.5612	0.4509	PSNR-HVS-M [45]	0.7725	0.1290	0.1667	0.8222	0.6529
MAD [17]	0.7480	0.6641	0.8907	0.7708	0.5734	MAD [17]	0.8202	0.1258	0.1502	0.8988	0.7272
DW-PSNR	0.5852	0.8102	1.0882	0.5891	0.4394	DW-PSNR	0.8053	0.1176	0.1556	0.8239	0.6594
CTW-PSNR	0.6478	0.7338	1.0223	0.6593	0.5083	CTW-PSNR	0.7774	0.1272	0.1651	0.8066	0.6432
SW-PSNR	0.6292	0.7585	1.0430	0.6423	0.4862	SW-PSNR	0.7782	0.1277	0.1649	0.8088	0.6419
IW-PSNR	0.6664	0.7177	1.0006	0.6823	0.5255	IW-PSNR	0.8024	0.1217	0.1567	0.8311	0.6592
DW-SSIM	0.8040	0.5911	0.7979	0.8168	0.6217	DW-SSIM	**0.9359**	**0.0721**	**0.0925**	**0.9340**	**0.7704**
CTW-SSIM	0.8244	0.5678	0.7595	0.8062	0.6178	CTW-SSIM	0.8917	0.0898	0.1188	0.8985	0.7222
SW-SSIM	0.8223	0.5746	0.7637	0.8267	0.6309	SW-SSIM	0.8979	0.0884	0.1156	0.9053	0.7281
IW-SSIM	**0.8579**	**0.5276**	**0.6895**	**0.8559**	**0.6636**	IW-SSIM	0.9144	0.0801	0.1063	**0.9213**	0.7529

o) Local block-wise distortions of different intensity; p) Mean shift (intensity shift); and q) Contrast change.

- The Categorical Image Quality (CSIQ) Database [58] was developed at Oklahoma State University. 30 original images were used to create a total of 866 distorted images using six types of distortions at four to five distortion levels. The distortion types include JPEG compression, JPEG2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring.

We use five evaluation metrics to compare the performance of IQA measures. Some of the metrics were included in previous tests carried out by the video quality experts group (VQEG) [59]. Other metrics are adopted from previous publications [56], [60]. These evaluation metrics are shown in the following.

TABLE II
AVERAGE PERFORMANCE OVER SIX DATABASES

Direct Average				Database Size-Weighted Average			
Model	PLCC	SRCC	KRCC	Model	PLCC	SRCC	KRCC
PSNR	0.6811	0.6925	0.5158	PSNR	0.6622	0.6887	0.5172
SSIM [5]	0.8636	0.8644	0.6810	SSIM [5]	0.8416	0.8455	0.6615
MS-SSIM [42]	0.8928	0.8909	0.7119	MS-SSIM [42]	**0.8847**	**0.8914**	**0.7116**
VSNR [44]	0.8193	0.8397	0.6671	VSNR [44]	0.7615	0.7903	0.6157
VIF [34]	0.8548	0.8431	0.6789	VIF [34]	0.8743	0.8456	0.6860
PSNR-HVS-M [45]	0.8073	0.8234	0.6577	PSNR-HVS-M [45]	0.7140	0.7314	0.5881
MAD [17]	0.8625	0.8836	0.7065	MAD [17]	0.8198	0.8509	0.6712
DW-PSNR	0.8004	0.8004	0.6244	DW-PSNR	0.7305	0.7369	0.5767
CTW-PSNR	0.8299	0.8387	0.6670	CTW-PSNR	0.7613	0.7743	0.6163
SW-PSNR	0.8184	0.8238	0.6504	SW-PSNR	0.7497	0.7627	0.5999
IW-PSNR	0.8389	0.8449	0.6715	IW-PSNR	0.7754	0.7896	0.6267
DW-SSIM	**0.9014**	**0.8954**	**0.7210**	DW-SSIM	0.8777	0.8821	0.7080
CTW-SSIM	0.8895	0.8837	0.7058	CTW-SSIM	0.8720	0.8659	0.6884
SW-SSIM	0.8877	0.8863	0.7046	SW-SSIM	0.8712	0.8748	0.6929
IW-SSIM	**0.9126**	**0.9063**	**0.7343**	IW-SSIM	**0.8974**	**0.8978**	**0.7240**

- Pearson Linear correlation coefficient (PLCC) after a nonlinear mapping between the subjective and objective scores. For the ith image in an image database of size N, given its subjective score o_i (mean opinion score (MOS) or difference of MOS (DMOS) between reference and distorted images) and its raw objective score r_i, we first apply a nonlinear function to r_i given by [60]

$$q(r) = a_1 \left\{ \frac{1}{2} - \frac{1}{1+\exp[a_2(r-a_3)]} \right\} + a_4 r + a_5 \quad (52)$$

  where a_1 to a_5 are model parameters found numerically using a nonlinear regression process in MATLAB optimization toolbox to maximize the correlations between subjective and objective scores. The PLCC value can then be computed as

$$\text{PLCC} = \frac{\sum_i (q_i - \bar{q}) * (o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 * \sum_i (o_i - \bar{o})^2}}. \quad (53)$$

- MAE is calculated using the converted objective scores after the nonlinear mapping described previously

$$\text{MAE} = \frac{1}{N} \sum |q_i - o_i|. \quad (54)$$

- RMS error is computed similarly as

$$\text{RMS} = \sqrt{\frac{1}{N} \sum (q_i - o_i)^2}. \quad (55)$$

- Spearman's rank correlation coefficient (SRCC) is defined as:

$$\text{SRCC} = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2-1)} \quad (56)$$

  where d_i is the difference between the ith image's ranks in subjective and objective evaluations. SRCC is a nonparametric rank-based correlation metric, independent of any monotonic nonlinear mapping between subjective and objective scores.
- Kendall's rank correlation coefficient (KRCC) is another nonparametric rank correlation metric given by

$$\text{KRCC} = \frac{N_c - N_d}{\frac{1}{2}N(N-1)} \quad (57)$$

  where N_c and N_d are the numbers of concordant and discordant pairs in the data set, respectively.

Among the previously mentioned metrics, PLCC, MAE, and RMS are adopted to evaluate *prediction accuracy* [59], and SRCC and KRCC are employed to assess *prediction monotonicity* [59]. A better objective IQA measure should have higher PLCC, SRCC, and KRCC while lower MAE and RMS values.

In all of our tests, only the distorted images in the six databases were employed (i.e., reference images are excluded). This avoids several difficulties in computing the evaluation metrics. Specifically, the reference images have infinite PSNR

TABLE III
SPEARMAN RANK ORDER CORRELATION COEFFICIENT COMPARISONS FOR INDIVIDUAL DISTORTION TYPES

Data set	PSNR	IW-PSNR	Increase	Ave. Increase	SSIM	IW-SSIM	increase	Ave. Increase
LIVE - JPEG (1)	0.8779	0.9641	+0.0862		0.9637	0.9645	+0.0008	
LIVE - JPEG (2)	0.7699	0.8456	+0.0757	+0.0798	0.9215	0.9409	+0.0253	+0.0126
CSIQ - JPEG	0.8881	0.9655	+0.0774		0.9546	0.9662	+0.0116	
LIVE - JPEG2000 (1)	0.9264	0.9664	+0.0400		0.9637	0.9751	+0.0114	
LIVE - JPEG2000 (2)	0.8549	0.9555	+0.1006	+0.0607	0.9604	0.9725	+0.0121	+0.0104
CSIQ - JPEG2000	0.9362	0.9777	+0.0415		0.9606	0.9683	+0.0077	
LIVE - Gaussian Blur	0.7823	0.9370	+0.1547	+0.0971	0.9517	0.9719	+0.0202	+0.0188
CSIQ - Gaussian Blur	0.9291	0.9685	+0.0394		0.9609	0.9782	+0.0173	
CSIQ - Contrast Change	0.8621	0.9230	+0.0609	+0.0609	0.7922	0.9539	+0.1617	+0.1617
LIVE - Gaussian Noise	0.9854	0.9806	−0.0048	−0.0035	0.9694	0.9667	−0.0027	+0.0190
CSIQ - Gaussian Noise	0.9363	0.9341	−0.0022		0.8974	0.9380	+0.0406	
CSIQ - 1/f Noise	0.9339	0.9352	+0.0013	+0.0013	0.8922	0.9059	+0.0137	+0.0225
LIVE - JP2 Trans. Error	0.8907	0.8326	−0.0581	−0.0581	0.9556	0.9442	−0.0114	−0.0114

value, making it hard to perform nonlinear regression and compute PLCC, MAE, and MSE values. In addition, since all reference images are assumed to have perfect quality, there are no natural relative ranks between them, resulting in ambiguities when computing SRCC and KRCC metrics.

Table I shows our test results of 15 IQA measures using the six databases. To provide an evaluation of the overall performance of the IQA measures under comparison, Table II gives the average PLCC, SRCC, and KRCC results over six databases, where the average values are computed in two cases. In the first case, the correlation scores are directly averaged, while in the second case, different weights are given to different databases, depending upon their sizes (measured as the numbers of images, i.e., 779 for LIVE, 54 for Cornell A57, 185 for IVC, 168 for Toyama, 1700 for TID2008, and 866 for CSIQ databases, respectively). For each evaluation metric in each test, we highlight the best two results with boldface. We have three major observations based upon the results shown in Tables I and II:

- First, information content weighting leads to consistent improvement in the performance of IQA algorithms for different underlying local quality measures. This can be seen by comparing the performance between {PSNR and IW-PSNR}, or {SSIM, MS-SSIM, and IW-SSIM}. In fact, for every database and every evaluation metric in Tables I and II, IW-based weighting always results in performance improvement. Although not all improvements are significant (which are not surprising as several existing IQA measures have already achieved fairly high performance for the databases being tested), the consistency of improvements is perhaps a stronger indicator of the effect and reliability of information content weighting.
- Second, information content weighting converts the widely criticized PSNR measure into a quite competitive perceptual IQA approach. Indeed, the performance of IW-PSNR is often comparable to many state-of-the-art algorithms. This is quite surprising because both PSNR and IW-PSNR are based upon rather poor local image quality measurement (point-wise absolute error), as demonstrated in Fig. 2. This is probably a more straightforward and stronger demonstration of the power of information content weighting.
- Third, DW-, CTW-, and SW-based pooling all can improve the performance of image quality measures. Some of them achieve superior performance in subtests. For example, DW-SSIM has outstanding performance on CSIQ and Toyama-MICT databases. However, such improvement is not as consistent and reliable as the IW approach. For example, DW-SSIM is not as impressive on IVC and TID2008 databases. The best overall performance is achieved by IW-SSIM, which is a combination of several useful ideas, including local SSIM measurement, multiscale signal analysis and weighting, and information content-based pooling. It is worth mentioning that this is achieved without introducing any new parameter and without involving any training or parameter tuning process.

To examine the effects of information content weighting on different image distortion types, we carried out a breakdown test on the individual data sets in LIVE and CSIQ databases. The results are shown in Table III. It can be observed that the IW approach leads to consistent improvement for JPEG compression, JPEG2000 compression and blur distortions, but is not as

TABLE IV
PEARSON LINEAR CORRELATION COEFFICIENT COMPARISONS FOR LOW AND HIGH QUALITY IMAGES

Data set	PSNR	IW-PSNR	Increase	Ave. Increase	SSIM	IW-SSIM	Increase	Ave. Increase
LIVE - High	0.7814	0.7447	−0.0367		0.8592	0.8437	−0.0155	
A57 - High	0.2875	0.5800	+0.2925		0.5818	0.5021	−0.0797	
IVC - High	0.4286	0.7175	+0.2889	+0.1185	0.7431	0.7516	+0.0085	+0.0265
Toyama - High	0.4845	0.6812	+0.1967		0.7955	0.8222	+0.0267	
TID2008 - High	0.1621	0.1742	+0.0121		0.4653	0.4894	+0.0241	
CSIQ - High	0.6675	0.6250	−0.0425		0.4805	0.6755	+0.1950	
LIVE - Low	0.6139	0.8576	+0.2437		0.7909	0.8513	+0.0604	
A57 - Low	0.5671	0.8233	+0.2562		0.6692	0.8548	+0.1856	
IVC - Low	0.3721	0.6800	+0.3079	+0.2246	0.7426	0.7443	+0.0017	+0.0973
Toyama - Low	0.4313	0.6231	+0.1918		0.6255	0.6970	+0.0715	
TID2008 - Low	0.4719	0.5401	+0.0682		0.6295	0.7728	+0.1433	
CSIQ - Low	0.3229	0.6027	+0.2798		0.6110	0.7321	+0.1211	

TABLE V
COMPARISONS OF COMPUTATION TIME (IN SECOND/IMAGE)

Model	PSNR	VSNR	VIF	PSNR-HVS-M	MAD	SSIM	MS-SSIM	DW-SSIM	CTW-SSIM	SW-SSIM	IW-SSIM
Time	0.1084	0.7492	3.0193	2.1189	41.926	0.1247	0.3714	0.5929	0.6026	0.8298	1.6503

reliable when the distortion types are noise contamination or transmission error.

To examine how the proposed IW approach behaves on different levels of image distortions, we conducted breakdown tests on all six databases by evenly dividing each database into low-quality and high-quality halves. The results are shown in Table IV. It appears that on average, improvement is achieved on both low-quality and high-quality images, but the level and consistency of improvement are much more significant at low-quality than high-quality levels.

Finally, to compare the computational complexity of different algorithms, we measured the average computation time required to assess an image of size 512×512 (using a computer with Intel Q6800 processor at 2.93 GHz). Table V reports the measurement results, which are rough estimates only, as no code optimization has been done on our Matlab implementations. It can be seen that IW-SSIM takes more time than PSNR, VSNR and other versions of SSIM, but less time than VIF and MAD. In particular, the savings over VIF might be due to the use of the Laplacian pyramid, rather than the steerable pyramid decompositions, which have higher computational complexity and include more orientation subbands. Since almost all methods under comparison (except for MAD) have quite high speed (less than a few seconds per image), computational complexity may not be a major concern in most real-world applications.

To facilitate future study and comparisons, we have put the Matlab code of the proposed IW-PSNR and IW-SSIM algorithms as well as our evaluation results online at http://www.ece.uwaterloo.ca/~z70wang/research/iwssim/.

V. CONCLUSIONS AND DISCUSSIONS

This paper targets at finding the optimal pooling strategy for the design of IQA algorithms. We propose a multiscale information content weighting approach based upon a GSM model of natural images [39]. We show that this novel weighting method leads to significant and consistent performance improvement of both PSNR- and SSIM-based IQA algorithms. Interestingly, the widely recognized VIF algorithm [34] can also be reinterpreted in the same information content weighting framework. Our extensive tests with six publicly-available independent image databases show that the proposed IW-SSIM algorithm achieves the best overall performance. We believe that our results support the general principle underlying our approach, i.e., the optimal weight for pooling should be directly proportional to local information content measured in units of bit.

The success of the IW-SSIM approach may be understood as a natural consequence of an effective combination of several proven useful approaches in IQA research. These include multiscale image decomposition followed by scale-variant weighting, SSIM-based local quality measurement [5], and information theoretic analysis of visual information content and fidelity [34], [37]. The current method may be extended in many directions. Specifically, the image model currently being

employed is based upon local magnitude statistics only. Advanced models that capture nonlocal characteristics of natural images or phase and orientation regularities may lead to more accurate information content measures. In addition, although the images in five of the six test databases being employed in this paper are color images, only the luminance components of the images were used for IQA. How to make use of the color components, and especially how to evaluate spatio-chromatic information content is still an unresolved problem.

REFERENCES

[1] T. N. Pappas, R. J. Safranek, and J. Chen, , A. Bovik, Ed., "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Proc.*, 2nd ed. New York: Academic, 2005.

[2] , H. R. Wu and K. R. Rao, Eds., *Digital Video Image Quality and Perceptual Coding*. Boca Raton, FL: CRC Press, 2005.

[3] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Rafael, CA: Morgan & Claypool Publishers, Mar. 2006.

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

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[6] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," in *Proc. Soc. Inf. Display Dig. Tech. Papers*, 1993, vol. XXIV, pp. 946–949.

[7] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 1989, pp. 1945–1948.

[8] S. Daly, , A. B. Watson, Ed., "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993, pp. 179–206.

[9] J. Lubin, , A. B. Watson, Ed., "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993, pp. 163–178.

[10] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Process.*, 1994, pp. 982–986.

[11] 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.

[12] B. A. Wandell, Sinauer Associates, Inc., Foundations of Vision 1995.

[13] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, Oct. 2006, pp. 2945–2948.

[14] E. C. Larson and D. M. Chandler, "Unveiling relationships between regions of interest and image fidelity metrics," *Proc. SPIE Vis. Commun. Image Process.*, vol. 6822, pp. 6822A1–16, Jan. 2008.

[15] E. C. Larson, C. T. Vu, and D. M. Chandler, "Can visual fixation patterns improve image fidelity assessment?," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2008, vol. 3, pp. 2572–2575.

[16] U. Engelke, V. X. Nguyen, and H.-J. Zepernick, "Regional attention to structural degradations for perceptual image quality metric design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar.–Apr. 2008, pp. 869–872.

[17] 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, pp. 011006:1–011006:21, Jan.–Mar. 2010.

[18] S. Wolf and M. H. Pinson, "Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system," *Proc. SPIE*, vol. 3845, pp. 266–277, 1999.

[19] 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.

[20] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer.*, vol. 7, pp. 2032–2040, Oct. 1990.

[21] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[22] E. Peli, "Feature detection algorithm based on a visual system model," *Proc. IEEE*, vol. 90, no. 1, pp. 78–93, Jan. 2001.

[23] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barbba, "Does where you gaze on an image affect your preception of quality? applying visual attention to image quality metric," in *Proc. IEEE Int. Conf. Image Process.*, Apr. 2007, vol. 2, pp. 169–172.

[24] U. Rajashekar, A. C. Bovik, and L. K. Cormack, "Gaffe: A gaze-attention fixation finding engine," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 564–573, Apr. 2008.

[25] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process.: Image Commun.*, vol. 19, Special Issue on Objective Video Quality Metrics, pp. 121–132, Feb. 2004.

[26] Z. K. Lu, W. Lin, X. K. Yang, E. P. Ong, and S. S. Yao, "Modeling visual attention's modulatory afterreffects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1928–1942, Nov. 2005.

[27] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," *J. Opt. Soc. Amer. A*, vol. 24, pp. B61–B69, Dec. 2007.

[28] W. Osberger, N. Bergmann, and A. Maeder, "An automatic image quality assessment technique incorporating high level perceptual factors," in *Proc. IEEE Int. Conf. Image Process.*, 1998, pp. 414–418.

[29] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1397–1410, Oct. 2001.

[30] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 3, pp. 129–132, Mar. 2002.

[31] Z. Wang, L. Lu, and A. C. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 243–254, Feb. 2003.

[32] E. P. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.

[33] J. Najemnik and W. S. Geisler, "Optimal eye movement strategies in visual search," *Nature*, no. 434, pp. 387–391, 2005.

[34] 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.

[35] A. A. Stocker and E. P. Simoncelli, "Noise characteristics and prior expectations in human visual speed perception," *Nature Neurosci.*, vol. 9, pp. 578–585, 2006.

[36] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[37] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[38] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[39] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 855–861, 2000.

[40] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COMM-31, no. 4, pp. 532–540, Apr. 1983.

[41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ: Wiley, 1991.

[42] 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.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "The SSIM index for image quality assessment," [Online]. Available: http://www.cns.nyu.edu/~lcv/ssim/

[44] 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.

[45] 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. Quality Metrics Consum. Electron.*, Scottsdale, AZ, Jan. 2007.

[46] 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/

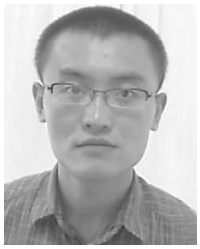
[47] M. Gaubatz and S. S. Hemami, "MeTriX MuX visual quality assessment package," [Online]. Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

[48] N. Ponomarenko, "PSNR-HVS-M download page," [Online]. Available: http://www.ponomarenko.info/psnrhvsm.htm
[49] E. C. Larson and D. M. Chandler, "Full-reference image quality assessment and the role of strategy: The most apparent distortion," [Online]. Available: http://vision.okstate.edu/mad
[50] D. Walther and C. Koch, "Saliency toolbox," [Online]. Available: http://www.saliencytoolbox.net/
[51] D. Walther and C. Koch, "Modeling attention to salient proto-objects," vol. 19, pp. 1395–1407, 2006.
[52] 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
[53] A. Ninassi, P. Le Callet, and F. Autrusseau, "Pseudo no reference image quality metric using perceptual data hiding," in *Proc. SPIE: Human Vis. Electron. Imag.*, San Jose, CA, Jan. 2006, vol. 6057.
[54] A. Ninassi, P. Le Callet, and F. Autrusseau, "Subjective quality assessment—IVC database," [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb
[55] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, "MICT image quality evaluation database," [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html
[56] 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. Quality Metrics Consum. Electron.*, Scottsdale, AZ, Jan. 2009.
[57] N. Ponomarenko and K. Egiazarian, "Tampere image database 2008 TID2008," [Online]. Available: http://www.ponomarenko.info/tid2008.htm
[58] E. C. Larson and D. M. Chandler, "Categorical image quality (CSIQ) database," [Online]. Available: http://vision.okstate.edu/csiq
[59] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Apr. 2000, available at [Online]. Available: http://www.vqeg.org/
[60] 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.

Zhou Wang (S'97–A'01–M'02) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin in 2001.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He was an Assistant Professor at The University of Texas at Arlington; a Howard Hughes Research Associate at New York University; and a Research Engineer at AutoQuant Imaging, Inc. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has more than 90 publications and one U.S. patent in these fields with more than 6,000 citations (Google Scholar). He is an author of *Modern Image Quality Assessment* (Morgan & Claypool, 2006).

Dr. Wang has served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009-present), IEEE SIGNAL PROCESSING LETTERS (2006–2010), and *Pattern Recognition* (2006-present), and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2007–2009) and EURASIP *Journal on Image and Video Processing* (2009–2010). He was a recipient of 2009 IEEE Signal Processing Society Best Paper Award, ICIP 2008 IBM Student Paper Award (as senior author), and 2009 Ontario Early Researcher Award.



Qiang Li (S'06–M'09) received the B.S. and M.S. degrees from Beijing Institute of Technology, and the Ph.D. degree from The University of Texas at Arlington in 2000, 2003 and 2009, respectively.

He is currently a Video Algorithm Engineer at Media Excel Inc., Austin, TX, where he works on perceptual video quality assessment and pre- and post-processing algorithms. His current research interests include objective video quality assessment and compression.

Dr. Li was a recipient of IBM Student Paper Award at the 2008 IEEE International Conference on Image Processing.