Quality-Aware Images

Zhou Wang, *Member, IEEE*, Guixing Wu, Hamid Rahim Sheikh, *Member, IEEE*, Eero P. Simoncelli, *Senior Member, IEEE*, En-Hui Yang, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*—We propose the concept of *quality-aware image*, in which certain extracted features of the original (high-quality) image are embedded into the image data as invisible hidden messages. When a distorted version of such an image is received, users can decode the hidden messages and use them to provide an objective measure of the quality of the distorted image. To demonstrate the idea, we build a practical quality-aware image encoding, decoding and quality analysis system,[1] which employs: 1) a novel reduced-reference image quality assessment algorithm based on a statistical model of natural images and 2) a previously developed quantization watermarking-based data hiding technique in the wavelet transform domain.**

***Index Terms*—Generalized Gaussian density (GGD), image communication, image quality assessment, image watermarking, information hiding, natural image statistics, quality-aware image, reduced-reference image quality assessment.**

I. Introduction

DIGITAL images are subject to a variety of distortions during compression, transmission, processing, and reproduction. In order to maintain, control and possibly enhance the quality of the image and video data being delivered, it is important for data management systems (e.g., network video servers) to be able to identify and quantify quality degradations on the fly. Since most of the image data will eventually be consumed by humans, the most reliable means of assessing image quality is subjective evaluation. However, subjective testing is expensive and time-consuming. On the other hand, most objective image/video quality assessment methods proposed in the literature [1]–[3] are not applicable in this scenario because they are *full-reference* (FR) methods that require access to the original images as references. Therefore, it is highly desirable to develop quality assessment algorithms that do not require full access to the reference images.

Unfortunately, *no-reference* (NR) or "blind" image quality assessment is an extremely difficult task. Most proposed NR quality metrics are designed for one or a set of predefined specific distortion types [4]–[10] that may not be generalized for evaluating images degraded with other types of distortions. Moreover, knowledge of the distortions that arise between the original and corrupted images is in general not available to image quality assessment systems. Thus, it is desirable to have a more general image quality assessment system that is applicable to a wide variety of distortions. However, to the best of our knowledge, no such method has been proposed and extensively tested.

One interesting recent development in image/video quality assessment research is to design *reduced-reference* (RR) methods for quality assessment [2], [3]. These methods do not require full access to reference images, but only needs partial information, in the form of a set of extracted features. Conceptually, RR methods make the quality assessment task easier than NR methods by paying the additional cost of transmitting side information to the users. The standard deployment of an RR method requires the side information to be sent through an ancillary data channel [3]. However, this restricts the application scope of the method because an additional data channel may be inconvenient or expensive to provide. An alternative solution would be to send the side information in the same channel as the images being transmitted. For example, the side information can be included as a component of the image data structure (e.g., as part of the header of the image format). However, this strategy would be difficult to implement in existing large-scale, heterogeneous networks such as the Internet, because it requires all the users in the communication network to adopt a new image format, or amend all the existing image formats to allow the side information to be included. Besides, lossy data transmission and typical image format conversion may cause loss of the original image headers.

In this paper, we propose the concept of *quality-aware image*, in which extracted features of the reference image are embedded as hidden messages. When a distorted version of such an image is received, the users can decode the hidden messages and use them to help evaluate the quality of the distorted image using an RR quality assessment method. There are several advantages of this approach.

Manuscript received August 30, 2004; revised May 31, 2005. The work of Z. Wang and E. P. Simoncelli was supported in part by the Howard Hughes Medical Institute. The work of G. Wu and E.-H. Yang was supported in part by the National Sciences and Engineering Research Council of Canada, in part by the Premier's Research Excellence Award, Canada Foundation for Innovation, in part by the Ontario Distinguished Research Award, and in part by the Canada Research Chairs Program. The work of H. R. Sheikh and A. C. Bovik was supported in part by the National Science Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhigang Fan.

Z. Wang was with Center for Neural Science, New York University, New York, NY 10012 USA. He is now with Department of Electrical Engineering, The University of Texas at Arlington, Arlington, TX 76019 USA (e-mail: zhouwang@ieee.org).

E. P. Simoncelli is with Center for Neural Science and the Courant Institute for Mathematical Sciences, New York University, New York, NY 10012 USA (e-mail: eero.simoncelli@nyu.edu).

G. Wu and E.-H. Yang are with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: g2wu@bbcr.uwaterloo.ca; ehyang@bbcr.uwaterloo.ca).

H. R. Sheikh is with Texas Instruments, Inc., Dallas, TX 75243 USA (e-mail: hamid.sheikh@ieee.org).

A. C. Bovik is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: bovik@ece.utexas.edu).

Digital Object Identifier 10.1109/TIP.2005.864165

[1]A MATLAB implementation of the system is available online at http://www.cns.nyu.edu/ lcv/qaware.

- It uses an RR method that makes the image quality assessment task feasible (as compared to FR and NR methods).
- It does not affect the conventional usage of the image data because the data hiding process causes only invisible changes to the image.
- It does not require a separate data channel to transmit the side information.
- It allows the image data to be stored, converted and distributed using any existing or user-defined formats without losing the functionality of "quality-awareness," provided the hidden messages are not corrupted during lossy format conversion.
- It provides the users with a chance to partially "repair" the received distorted images by making use of the embedded features.

This study is largely inspired by [11]–[13], where a pseudo-random bit sequence or a watermark image is hidden inside the image being transmitted. The bit-error rate or the degradation of the watermark image measured at the receiver side is then used as an indication of the quality degradation of the host image signal. These methods are perhaps the first attempts to use information hiding technologies for the estimation of image quality degradation. Nevertheless, strictly speaking, these methods are not image quality assessment methods because no extracted features about either the reference or the distorted images are actually used in the quality evaluation process. Instead, the distortion processes that occur in the distortion channel are gauged, in the hope that such estimated channel distortion would correlate well with perceptual image degradation incurred during transmission through the channel. However, such a connection is obscured by the nature (e.g., complexity) of the image signals and the types of image distortions, which have variable effects on perceived image quality. In addition, these methods provide no clue about how the received distorted images can be repaired.

Information hiding or digital watermarking has been an active research area in the last decade. Traditionally, these techniques have been designed for security-related applications such as copyright protection and data authentication. Recently, researchers have attempted to broaden their application scope to nonsecurity oriented applications [14], [15]. Quality-aware images mainly belong to the second category (see Section V for discussions), and they bring about new challenges in the selection and design of information hiding techniques.

II. Quality-Aware Image

A. Framework

A system diagram of quality-aware image encoding, decoding and quality analysis system is shown in Fig. 1. A feature extraction process is first applied to the original image, which is assumed to have perfect quality. The quality-aware image is obtained by embedding these features as invisible messages into the original image. The quality-aware image may then pass through a "distortion process" before it reaches the receiver side. Here the "distortion process" is general in concept. It can be a distortion channel in an image communication system, with possibly lossy compression, noise contamination and/or postprocessing involved. It can also be any other processes that may alter the image.

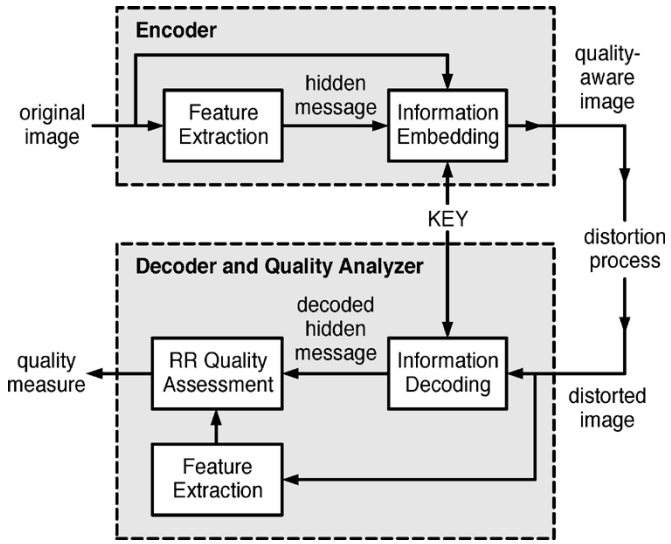


Fig. 1. Quality-aware image encoding, decoding, and quality analysis system.

At the receiver side, the hidden messages are first decoded from the distorted quality-aware image. In order for correct decoding of the messages, the key for information embedding and decoding is shared between the sender and the receiver. Depending on the application environment, there may be different ways to distribute the embedding key. One simple solution is to attach the key to the decoder software and/or publish the key, so that it can be easily obtained by all potential users of quality-aware images. Note that the key is independent of the image and can be the same for all quality-aware images, so it does not need to be transmitted with the image data. The decoded messages are translated back to the features about the reference image. Next, another feature extraction procedure corresponding to the one at the sender side is applied to the distorted image. The resulting features are then compared with those of the reference image to yield a quality score for the distorted quality-aware image.

In order to improve robustness, error detection/correction coding techniques may be applied before the information embedding process. Nevertheless, the hidden messages may still be decoded incorrectly when the distortions are extremely severe. At the receiver side, the system must be able to detect such situations (based on the error detection and correction code) and report a failure message, instead of a quality score.

B. Design Considerations

Designing an effective quality-aware image system is a challenging task. On the one hand, in order to provide effective quality prediction, the RR quality assessment system desires to know as much information as possible about the reference image. Therefore, the information hiding system would need to embed a fairly large amount of information. On the other hand, in order for the hidden messages to be invisible and for these messages to survive a wide variety and degree of distortions, the amount of information that can be embedded is limited. The

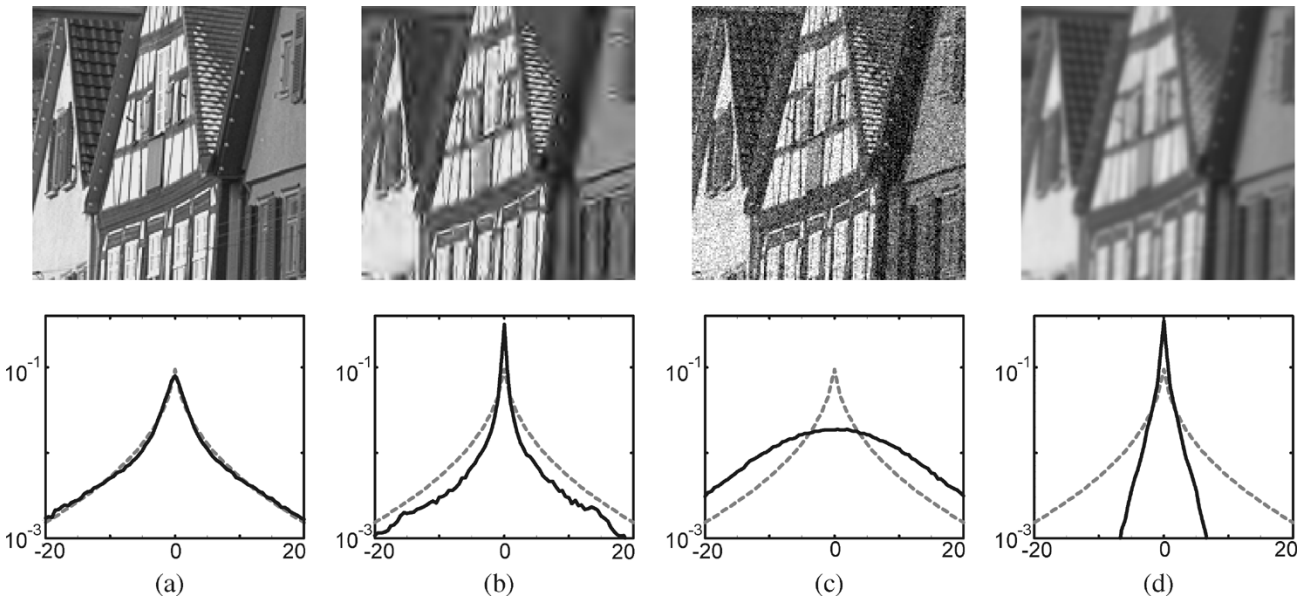


Fig. 2. Comparisons of wavelet coefficient histograms (solid curves) calculated from the same horizontal subband in the steerable pyramid decomposition [22]. (a) Original (reference) "buildings" image (cropped for visibility). (b) JPEG2000 compressed image. (c) White Gaussian noise contaminated image. (d) Gaussian blurred image. The histogram of the original image coefficients is well fitted by a generalized Gaussian density model (dashed curves).

RR quality assessment system must observe this limit and carefully select a set of features that can be encoded within the limit. These features must be highly relevant to image quality degradations. They must also provide an efficient summary about the reference image.

Another issue that may need to be considered is that many data hiding techniques tend to change certain statistical features of the original image (e.g. [16] and [17]). This could potentially conflict with quality assessment systems because these systems may rely on the way that these statistical features change as an indication of quality degradation.

To summarize, a successful quality-aware image system must provide a good trade-off between data hiding load, embedding distortion, robustness, and the accuracy of image quality prediction.

C. Simple Example

Perhaps the simplest way to implement a quality-aware image system is to embed a certain number of (perhaps randomly selected) reference image pixels as hidden messages. For synchronization purpose, the positions of theses pixels also need to be embedded. At the receiver side, the decoded reference image pixels are compared with the corresponding distorted image pixels, and certain distortion/quality metric, such as mean-squared error (MSE) and peak signal-to-noise ratio (PSNR), are estimated.

Such a system, although simple, is quite weak in several aspects. First, it requires a high data hiding rate. For example, for a 512×512, 8 bits/pixel gray scale image, to embed 1% of the image pixels (together with 2×9 bits for encoding each pixel position) requires a total of 68 146 bits, a heavy load for most robust information hiding systems. Second, such a small number of pixels is unlikely to allow accurate estimation of the distortion metrics, unless the distortion between the reference and distorted images is independently and identically distributed noise. The obvious drawbacks of this simple example lead us to consider image features that are more efficient in summarizing image information and more effective in evaluating image quality.

III. Implementation

A. RR Quality Assessment

Here, we propose a new RR quality assessment method based on statistics computed for natural images in the wavelet transform domain. Wavelet transforms provide a convenient framework for localized representation of signals simultaneously in space and frequency. They have been widely used to model the processing in the early stages of biological visual systems and have also become the preferred form of representations for many image processing and computer vision algorithms. In recent years, natural image statistics have played an important role in the understanding of sensory neural behaviors of the human visual system [18]. In the image processing literature, statistical prior models of natural images have been employed as fundamental ingredients in a large number of image coding and estimation algorithms (e.g., [19]–[21]). They have also been used for image quality assessment purposes (e.g., [8]).

Fig. 2 shows the histograms of the coefficients computed from one of the wavelet subbands in a steerable pyramid decomposition [22] (a type of redundant wavelet transform that avoids aliasing in subbands). It has been pointed out that the marginal distributions of such oriented bandpass filter responses of natural images are highly kurtotic [with sharp peaks at zero and much longer tails than Gaussian density, as demonstrated in Fig. 2(a)] and have a number of important implications to sensory neural coding of natural visual scene [23]. In [24] and [25], it was demonstrated that many natural looking texture images

can be synthesized by matching the histograms of the filter responses of a set of well-selected bandpass filters. Psychophysical visual sensitivity to histogram changes of wavelet-textures had also been studied (e.g., [26], [27]). In Fig. 2, it can be seen that the marginal distribution of the wavelet coefficients changes in different ways for different types of image distortions. Such histogram changes in images contaminated with white Gaussian noise have been observed previously and used for image denoising [19], [20].

Let $p(x)$ and $q(x)$ denote the probability density functions of the wavelet coefficients (assumed to be independently and identically distributed) in the same subband of two images, respectively. Let $\mathbf{x} = \{x_1, \ldots, x_N\}$ be a set of N randomly and independently selected coefficients. The log-likelihoods of $\mathbf{x}$ being drawn from $p(x)$ and $q(x)$ are

$$l(p) = \frac{1}{N}\sum_{n=1}^{N} \log p(x_n) \quad \text{and} \quad l(q) = \frac{1}{N}\sum_{n=1}^{N} \log q(x_n) \tag{1}$$

respectively. Now, assume that $p(x)$ is the true probability density distribution of the coefficients. Based on the law of large numbers, when N is large, the difference of the log-likelihoods (or, equivalently, the log-likelihood-ratio) between $p(x)$ and $q(x)$ asymptotically approaches the Kullback–Leibler distance [28] (KLD) between $p(x)$ and $q(x)$

$$l(p) - l(q) \longrightarrow d(p\|q) = \int p(x) \log \frac{p(x)}{q(x)} dx. \tag{2}$$

In previous work, a number of authors have pointed out the relationship between KLD and log-likelihood function and used KLD to compare images, mainly for classification and retrieval purposes [29]–[32]. KLD has also been used to quantify the distributions of image pixel intensity values for the evaluation of compressed image quality [33], [34]. Here, we use KLD to quantify the difference between wavelet coefficient distributions of a perfect quality reference image and a distorted image [denoted later on as $p(x)$ and $q(x)$, respectively)]. To make an effective estimation, the coefficient histograms for both images must be available. The latter can be easily computed from the received distorted image. The difficulty is in obtaining the coefficient histogram of the reference image at the receiver side. Transmitting all the histogram bins as hidden messages would result in either a heavy data load (when the bin step size is fine) or weaker statistical characterization (when the bin step size is coarse).

One important discovery in the literature of natural image statistics is that the marginal distribution of the coefficients in individual wavelet subbands can be well-fitted with a two-parameter generalized Gaussian density (GGD) model [35]

$$p_m(x) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x|}{\alpha}\right)^{\beta}} \tag{3}$$

where $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$ (for $a > 0$) is the Gamma function. One fitting example is shown in Fig. 2(a) as the dashed line. This model provides a very efficient means to summarize the coefficient histogram of the reference image, so that only two model parameters $\{\alpha, \beta\}$ need to be transmitted to the receiver as hidden messages. This model has been explicitly used in previous work for image compression [21] and texture image retrieval [32]. In addition to the fitting parameters α and β, we also embed the prediction error as a third parameter, which is defined as the KLD between $p_m(x)$ and $p(x)$

$$d(p_m\|p) = \int p_m(x) \log \frac{p_m(x)}{p(x)} dx. \tag{4}$$

In practice, this quantity has to be evaluated numerically using histograms

$$d(p_m\|p) = \sum_{i=1}^{L} P_m(i) \log \frac{P_m(i)}{P(i)} \tag{5}$$

where $P(i)$ and $P_m(i)$ are the normalized heights of the ith histogram bins, and L is the number of bins in the histograms.

At the receiver side, we wish to compute an approximation to (2), the KLD between the probability distribution of the original image $p(x)$ and that of the distorted image $q(x)$. Since we do not have the probability distribution of the original image, we replace the expectation over $p(x)$ with an expectation over the model density $p_m(x)$

$$\hat{d}(p\|q) = \int p_m(x) \log \frac{p(x)}{q(x)} dx \tag{6}$$

$$= d(p_m\|q) - d(p_m\|p). \tag{7}$$

The second term is simply the KLD between the original probability distribution and the model (4), which is embedded in the image by the encoder. The first term is the KLD between $p_m(x)$ and $q(x)$

$$d(p_m\|q) = \int p_m(x) \log \frac{p_m(x)}{q(x)} dx. \tag{8}$$

This is computed at the receiver side from the histogram bins of the distorted wavelet coefficients [analogous to (5)]. Note that, unlike the encoding side, we avoid fitting $q(x)$ with a GGD model, which may not be appropriate for the distorted data. Finally, the overall distortion between the distorted and reference images is defined as

$$D = \log_2 \left(1 + \frac{1}{D_0} \sum_{k=1}^{K} \left| \hat{d}^k(p^k\|q^k) \right| \right) \tag{9}$$

where K is the number of subbands, p^k and q^k are the probability density functions of the kth subbands in the reference and distorted images, respectively, $\hat{d}^k$ is the estimation of the KLD between p^k and q^k, and D_0 is a constant used to control the scale of the distortion measure (but has no impact on the performance of the algorithm).

B. Feature Extraction

Fig. 3 illustrates our implementation of the feature extraction system at the encoder side. We first apply a three-scale four-orientation steerable pyramid transform [22] to decompose the image into 12 oriented subbands (four for each scale) and the highpass and lowpass residuals, as demonstrated in Fig. 4. For

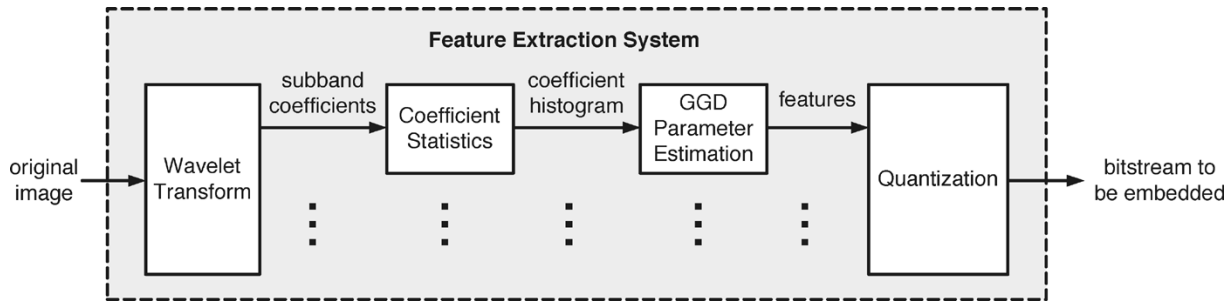


Fig. 3. Feature extraction system at the encoder side.

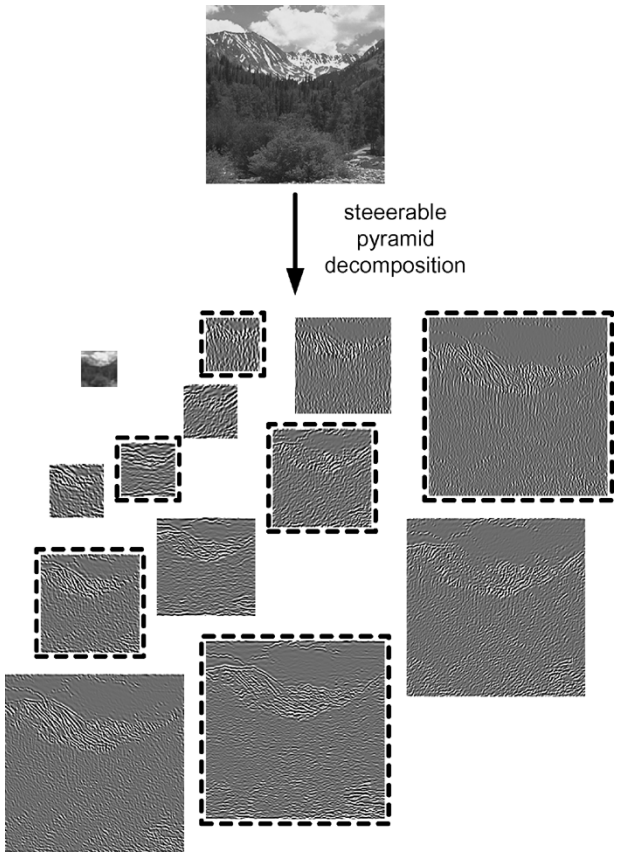


Fig. 4. Steerable pyramid decomposition [22] of image (highpass residual band not shown). A set of selected subbands (marked with dashed boxes) are used for GGD feature extraction.

each subband, the histogram of the coefficients is computed and then its feature parameters $\{\alpha, \beta, d(p_m||p)\}$ are estimated using a gradient descent algorithm to minimize the KLD between $p(x)$ and $p_m(x)$. Six of the 12 oriented subbands (as shown in Fig. 4) are selected for feature extraction. The major criterion for selecting these subbands is to reduce the data rate of RR features while at the same time, maintain the quality prediction performance. Specifically, in the Fourier domain, the adjacent steerable pyramid subbands (in both scale and orientation) have significant overlaps, but there is essentially no overlap between nonadjacent subbands. Therefore, the six subbands marked in Fig. 4 are selected to reduce the use of redundant information. Furthermore, in our tests, selecting the other six oriented subbands or all the 12 oriented subbands gives similar overall performance of image quality prediction.

The extracted scalar features are quantized to finite precision. Both β and $d(p_m||p)$ are quantized into 8-bit precision, and α is represented using 11-bit floating point, with 8 bits for mantissa and 3 bits for exponent. These quantization precision parameters were hand picked to represent the features in a limited number of bits while maintaining a reasonable approximation of the features. The final result is a total of $(8+8+8+3)\times 6 = 162$ bits that are embedded into the image.

C. Information Embedding

To embed the extracted features into the image, we choose to use an existing dithered uniform scalar quantization watermarking method in the wavelet transform domain. This method is a simple case of the class of quantization-index-modulation information embedding techniques [36], which allow for "blind" decoding (decoding does not require the access to the reference image) and achieve a good tradeoff between data-hiding rate and robustness. The information embedding system is illustrated in Fig. 5.

We first use a five-scale separable wavelet transform (specifically, a quadrature mirror filter transform [37]) to decompose the reference image into 16 subbands, including the horizontal, vertical and diagonal subbands at each scale, and a low frequency residual band. In order to embed one bit of information $m \in \{0,1\}$ into a wavelet coefficient c, the coefficient is altered according to the following rule:

$$c_q = Q\left(c + d(m)\right) - d(m) \equiv Q^m(c) \tag{10}$$

where c_q is the altered coefficient, $Q(\cdot)$ is a base quantization operator with quantization step size Δ, and $d(m)$ is a dithering operator defined as

$$d(m) = \begin{cases} -\frac{\Delta}{4}, & \text{if } m = 0 \\ \frac{\Delta}{4}, & \text{if } m = 1. \end{cases} \tag{11}$$

At the receiver side, a distorted coefficient c_d is obtained and used to estimate the embedded bit based on the minimum distance criterion

$$\hat{m}(c_d) = \arg\min_{m\in\{0,1\}} \left\| c_d - Q^m(c_d) \right\|. \tag{12}$$

We embed the hidden messages into the horizontal, vertical and diagonal subbands at the fifth scale (counted from fine to coarse) of the wavelet decomposition. We choose to use these low-frequency components because they usually have

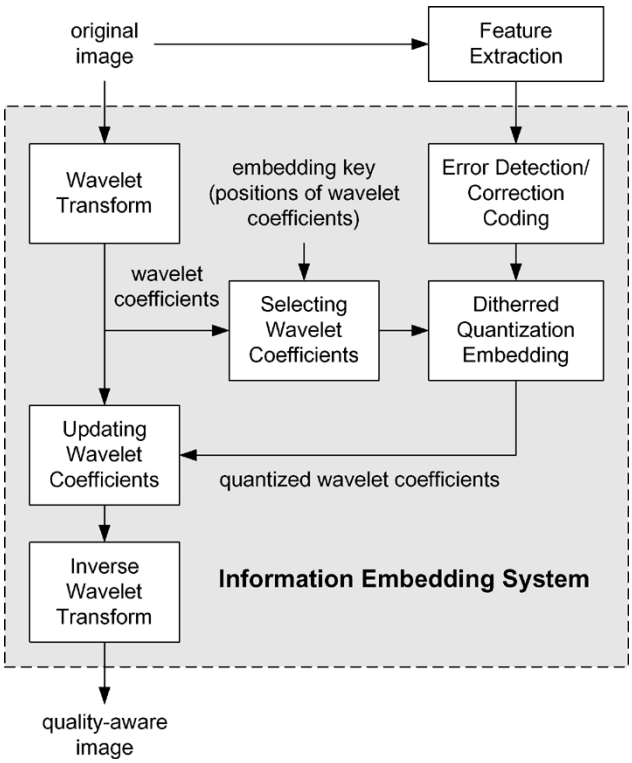


Fig. 5. Information embedding system.

high signal energy and are less likely to be significantly altered during typical image processing operations. Moreover, such a selection avoids conflict with the proposed RR quality assessment method, which is based on detecting the statistical changes of the wavelet coefficients at the finer scales. To further improve robustness, two error detection/correction techniques are employed. First, a 16-bit cyclic redundancy check (CRC) code [38] is computed and attached to the 162 information bits. Second, the resulting 178 bits are further encoded using a binary (15,5,7) BCH code [38], which can correct up to 3 bits of errors out of every 15 bits. As a result, a total of 540 bits are generated. The same number of wavelet coefficients are randomly selected from the fifth scale of the wavelet transform, and every bit is encoded into one coefficient using (10). The positions of the coefficients are shared between the sender and receiver as the embedding key.

At the receiver side, we first apply the same wavelet transform to the received image. The embedded 540 bits are then extracted from the corresponding wavelet coefficients using (12), and decoded with the BCH system. The decoded 178 bits are split into the corresponding 162 information bits and 16 CRC bits. We then calculate a new set of CRC bits using the decoded information bits and compare them with the decoded CRC bits. If any of the CRC bit is incorrect, the system reports a failure message. Otherwise, the extracted 162 information bits are converted back into scalar features about the reference image and relayed to the quality assessment system. Finally, a quality score of the distorted image is reported.

In several cases, a failure message may be reported. It could be that the received image is not a quality-aware image (no side information has been embedded) or the embedded information is desynchronized (e.g., by image editing). It could also be that the image quality degradation is very severe, such that the embedded information cannot be completely recovered. It is often useful to distinguish between the two cases, because in the latter case, a failure message can serve as an indication of low image quality. One way to make such a distinction is to look at the percentage R of correct CRC bits because statistically only in the latter case, R may be significantly higher than 50%. Following the general idea of [11]–[13], one can take an even further step to use R as an important factor for the prediction of image quality at very low quality range, although the accuracy may be complicated by the nature (e.g. complexity) of the images being evaluated.

IV. Test

A. Performance of Quality Assessment

In order to evaluate and compare the performance of image quality assessment algorithms, we built a large image database (the LIVE image database, available online [39]) and conducted an extensive subjective experiment to assess the quality of the images in the database. The database contains 29 high-resolution (typically 768 $\times$ 512) original images altered with five types of distortions at different distortion levels. The distorted images were divided into seven datasets. Datasets 1 (87 images) and 2 (82 images) are JPEG2000 compressed images; Datasets 3 (87 images) and 4 (88 images) are JPEG compressed images; and Datasets 5–7 (each containing 145 images) are distorted with white Gaussian noise, Gaussian blur, and transmission errors in the JPEG2000 bitstream using a fast-fading Rayleigh channel model, respectively. Subjects were asked to provide their perception of quality on a continuous linear scale and each image was rated by 20–25 subjects. The raw scores for each subject were converted into Z-scores and rescaled within each dataset to fill the range from 1 to 100. Mean opinion score and the standard deviation between subjective scores were then computed for each image.

Three measures are computed to quantify the performance of the proposed quality assessment method. First, following the procedure given in the Video Quality Experts Group (VQEG) Phase I FR-TV test [42], we use a logistic function to provide a nonlinear mapping between the objective and subjective scores

$$f(s) = \frac{a_1 - a_2}{1 + \exp\left(-\frac{(s-a_3)}{a_4}\right)} + a_2 \tag{13}$$

where s is the objective score and a_1, a_2, a_3, and a_4 are the model parameters, which are found numerically using a nonlinear regression process with MatLab optimization toolbox. After the nonlinear mapping, the correlation coefficient between the predicted and true subjective scores is calculated to evaluate *prediction accuracy*. Second, the Spearman rank-order correlation coefficient is employed to evaluate *prediction monotonicity*. Finally, to evaluate *prediction consistency*, the outlier ratio is used, which is defined as the percentage of predictions outside the range of ± 2 standard deviations between subjective scores.

To the best of our knowledge, no other RR method has been proposed that: 1) aims for general-purpose image quality assessment (as opposed to distortion- or application-specific) and 2) uses such small amount of information about the reference

TABLE I
PERFORMANCE EVALUATION OF IMAGE QUALITY MEASURES USING THE LIVE DATABASE [39]. JP2: JPEG2000 DATASET; JPG: JPEG DATASET; NOISE: WHITE GAUSSIAN NOISE DATASET; BLUR: GAUSSIAN BLUR DATASET; ERROR: TRANSMISSION ERROR DATASET

Dataset		JP2 (1)	JP2 (2)	JPG (1)	JPG (2)	Noise	Blur	Error
number of images		87	82	87	88	145	145	145
method	type	Correlation Coefficient (prediction accuracy)						
Proposed	RR	0.9353	0.9490	0.8452	0.9695	0.8889	0.8872	0.9175
PSNR	FR	0.9337	0.8948	0.9015	0.9136	0.9866	0.7742	0.8811
Sarnoff [40]	FR	0.9706	0.9650	0.9589	0.9837	0.9631	0.9480	0.9144
MSSIM [41]	FR	0.9676	0.9669	0.9647	0.9856	0.9706	0.9361	0.9439
Wang *et al.* [7]	NR	N/A	N/A	0.9592	0.9808	N/A	N/A	N/A
Sheikh *et al.* [8]	NR	0.9258	0.9064	N/A	N/A	N/A	N/A	N/A
method	type	Rank-Order Correlation Coefficient (prediction monotonicity)						
Proposed	RR	0.9298	0.9470	0.8332	0.8908	0.8639	0.9145	0.9162
PSNR	FR	0.9231	0.8816	0.8907	0.8077	0.9855	0.7729	0.8785
Sarnoff [40]	FR	0.9668	0.9565	0.9528	0.8904	0.9411	0.9381	0.9048
MSSIM [41]	FR	0.9566	0.9677	0.9572	0.9441	0.9719	0.9425	0.9498
Wang *et al.* [7]	NR	N/A	N/A	0.9507	0.8880	N/A	N/A	N/A
Sheikh *et al.* [8]	NR	0.9192	0.8918	N/A	N/A	N/A	N/A	N/A
method	type	Outlier Ratio (prediction consistency)						
Proposed	RR	0.0690	0.0366	0.1839	0.0341	0.1793	0.1172	0.0621
PSNR	FR	0.0805	0.0976	0.0920	0.1818	0.0000	0.2069	0.1517
Sarnoff [40]	FR	0.0000	0.0366	0.0115	0.0000	0.0345	0.0276	0.0552
MSSIM [41]	FR	0.0000	0.0000	0.0000	0.0114	0.0000	0.0414	0.0345
Wang *et al.* [7]	NR	N/A	N/A	0.0230	0.0227	N/A	N/A	N/A
Sheikh *et al.* [8]	NR	0.0575	0.0610	N/A	N/A	N/A	N/A	N/A

image as compared to the proposed method. Therefore, we compare the proposed method with a set of general-purpose FR models as well as application-specific NR models. These models include PSNR (FR), Lubin's Sarnoff model (FR) [40], [43], [44], the mean structural similarity index (MSSIM, FR) [41], the JPEG quality index by Wang *et al.* (NR) [7], and the JPEG2000 quality assessment method by Sheikh *et al.* (NR) [8]. Although such comparison is unfair to one method or another in different aspects, it provides a useful indication about the relative performance of the proposed method. The performance evaluation results of all methods are summarized in Table I. It can be seen that the proposed method performs quite well for a wide range of distortion types. Specifically, for five of the seven datasets, it gives better prediction accuracy (higher correlation coefficients), better prediction monotonicity (higher Spearman rank-order correlation coefficients) and better prediction consistency (lower outlier ratios) than PSNR, which is the most widely used FR image quality metric in the image processing literature. In comparison with the NR models, the proposed method is inferior to Wang *et al.*'s method for the JPEG datasets (JPEG compressed images have distinct blocking effect, which is readily detected by an application-specific NR method), and performs better than Sheikh *et al.*'s method for the JPEG2000 datasets. Note that these application-specific NR methods are not applicable to other types of image distortions. A more complete test may include other distortion types (including mixed distortions) as well as validations across different distortion types, but the current testing results lead us to believe that the proposed method is a reasonable and useful choice for quality-aware image systems. It needs to be emphasized that none of the other methods being compared, or any other method we are aware of, can be used in this scenario.

B. Robustness of Information Embedding

The information embedding system is tested with four distortion types: JPEG2000 compression, JPEG compression, white Gaussian noise contamination, and Gaussian blur. For convenience, we define the distortion levels as compression bit rate (bits/pixel) for JPEG2000 compression, quality factor (which controls the quantization step of discrete cosine transform coefficients) for JPEG compression, noise standard deviation for white noise contamination, and standard deviation of blurring filter for Gaussian blur, respectively. The same 29 original images in the LIVE database [39] are used for the test. We first generate ten quality-aware images (each uses a different randomly generated embedding key) for each of the test images. For any given distortion type and level, we distort the 290 quality-aware images accordingly and check if the hidden messages can be correctly decoded (by comparing the CRC bits, see Section III-B).

Since the RR quality assessment system can provide useful quality prediction only when the hidden messages are fully recovered, we use correct decoding rate (defined as the percentage of the images whose embedded messages are completely recovered) as the criterion for evaluating the robustness of the system. The test results are shown in Fig. 6, which covers the transition range (from %0 to 100% correct decoding rate) of distortion levels for each distortion type.

V. CONCLUSION

The major contributions of the paper include the following.

1) introduction of the concept of quality-aware image, and discussion of its design considerations;
2) implementation of a practical quality-aware image encoding, decoding and quality analysis system;
3) development of a simple and effective RR image quality assessment algorithm based on a wavelet-domain statistical model of natural images;
4) expansion of the application scope of information hiding technologies.

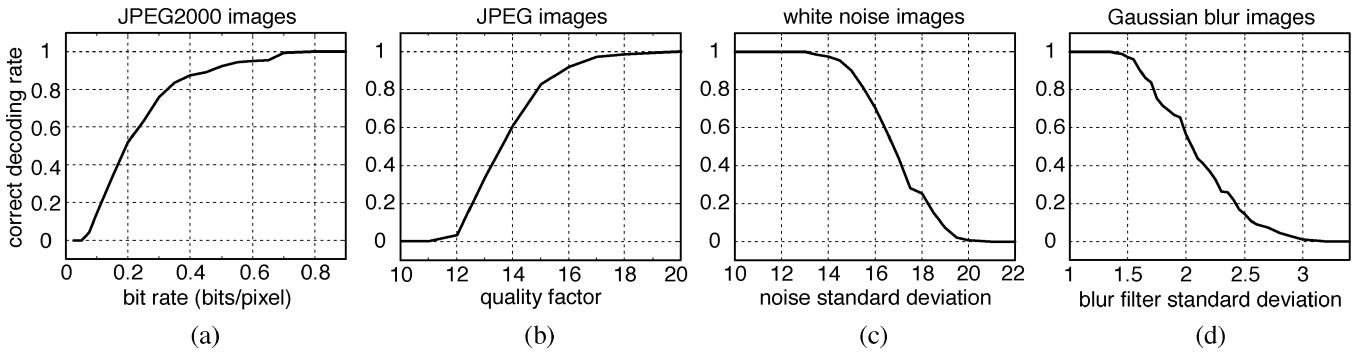


Fig. 6. Robustness test of the information embedding system. The quality/distortion level is defined as (a) bit rate (bits/pixel) for JPEG2000 compressed images; (b) quality factor for JPEG compressed images; (c) noise standard deviation for white noise contaminated images; and (d) standard deviation (pixels) of blurring filter for Gaussian blurred images.

Like other FR and RR approaches, the proposed quality assessment method assumes the existence of a perfect-quality reference image. In the case that this assumption does not hold, only an NR method can provide useful quality evaluation of the images.

In the future, the research initiated in this paper can be extended in several directions: The algorithm presented in Section III is only a specific implementation of the general framework of quality-aware image system (Fig. 1). The current method can be improved in many ways. For example, different RR image quality assessment algorithms could be employed. The improved algorithms may include more statistical image features (e.g., joint statistics of wavelet coefficients), which may lead to better quality prediction accuracy. For another example, different information hiding techniques could be used to enhance the robustness to a broader range of distortion types. The current method is sensitive to geometric transformations, gain attack and perhaps some other types of malicious attacks. The general concept of quality-aware images does not exclude itself from being employed in security-related applications. For example, in a pay-per-view scenario, an image could be paid according to its quality degradation. However, given the limited capability of the existing robust image watermarking techniques (including the one we are currently using), we propose to use it mainly for nonsecurity oriented applications, in which nobody will benefit from "removing" or "destroying" the embedded information, and therefore, the images are less likely to encounter malicious attacks (though the precise definition of malicious attacks could vary for different application environment). This is different from security-related applications such as copyright protection, where robustness to malicious attacks [45] is an essential issue.

The general approach may also be used beyond the scope of image quality assessment. For example, suppose an image is subject to a number of distortion stages. One can embed the quality scores measured at the intermediate stages into the image as additional hidden messages. The end receiver can then trace back to find the critical processing stages that have caused significant quality degradations. Inspired by the work of using data hiding techniques for error concealment (e.g., [46] and [47]), we can have another interesting application of the embedded features, which we refer to as "self-repairing images." The idea is to "repair" a distorted image by forcing some of its statistical properties to match those of the original image. Similar idea has been successfully used for texture synthesis (e.g., [24], [25], and [48]). Finally, the principle idea may be applied to other types of signals to create quality-aware (and possibly self-repairing) video, audio, and multimedia, etc.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Portilla, Dr. H. Farid, and the anonymous reviewers for valuable comments.

REFERENCES

[1] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, A. Bovik, Ed. New York: Academic, 2000.

[2] VQEG: The Video Quality Experts Group. [Online]. Available: http://www.vqeg.org/

[3] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marqueseds, Eds. Boca Raton, FL: CRC, 2003, pp. 1041–1078.

[4] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process. Lett.*, vol. 4, no. 4, pp. 317–320, Nov. 1997.

[5] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 2000, pp. 981–984.

[6] Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifact in digital video," *Proc. IEEE*, vol. 90, no. 1, pp. 154–169, Jan. 2002.

[7] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Processing*, Rochester, NY, Sep. 2002, pp. 477–480.

[8] 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.

[9] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Process. Image Commun.*, vol. 19, pp. 163–172, Feb. 2004.

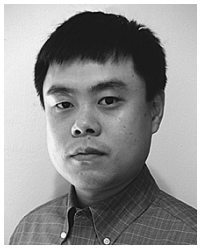
[10] X. Li, "Blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, Sep. 2002, pp. 449–452.

[11] O. Sugimoto, R. Kawada, M. Wada, and S. Matsumoto, "Objective measurement scheme for perceived picture quality degradation caused by MPEG encoding without any reference pictures," *Proc. SPIE*, vol. 4310, pp. 932–939, 2001.

[12] M. C. Q. Farias, S. K. Mitra, M. Carli, and A. Neri, "A comparison between an objective quality measure and the mean annoyance values of watermarked videos," in *Proc. IEEE Int. Conf. Image Processing*, Rochester, NY, Sep. 2002, pp. 469–472.

[13] P. Campisi, M. Carli, G. Giunta, and A. Neri, "Blind quality assessment system for multimedia communications using tracing watermarking," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 996–1002, Apr. 2003.

[14] M. Barni *et al.*, "What is the future for watermarking? (Part I)," *IEEE Signal Process. Mag.*, vol. 20, no. 9, pp. 55–59, Sep. 2003.
[15] ——, "What is the future for watermarking? (Part II)," *IEEE Signal Process. Mag.*, vol. 20, no. 11, pp. 53–57, Nov. 2003.
[16] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in *Proc. 3rd Int. Workshop Inf. Hiding*, Dresden, Germany, 1999, pp. 61–75.
[17] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," in *Proc. 5th Int. Workshop Inf. Hiding*, Noordwijkerhout, The Netherlands, 2002, pp. 340–354.
[18] E. P. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Ann. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.
[19] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. 3rd Int. Conf. Image Processing*, vol. I, Sep. 1996, pp. 379–382.
[20] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, Apr. 1999.
[21] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.
[22] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 3, pp. 587–607, Mar. 1992.
[23] D. J. Field, "What is the goal of sensory coding?," *Neural Comput.*, vol. 6, no. 4, pp. 559–601, 1994.
[24] D. Heeger and J. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. ACM SIGGRAPH*, Aug. 1995, pp. 229–238.
[25] S. C. Zhu, Y. N. Wu, and D. Mumford, "FRAME: Filters, random fields and maximum entropy-toward a unified theory for texture modeling," *Int. J. Comp. Vis.*, vol. 27, no. 2, pp. 1–20, 1998.
[26] C. Chubb, J. Econopouly, and M. S. Landy, "Histogram contrast analysis and the visual segregation of iid textures," *J. Opt. Soc. Amer. A*, vol. 11, no. 9, pp. 2350–2374, 1994.
[27] F. A. A. Kingdom, A. Hayes, and D. J. Field, "Sensitivity to contrast histogram differences in synthetic wavelet-textures," *Vis. Res.*, vol. 41, no. 5, pp. 585–598, 2001.
[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[29] J.-Y. Chen, C. A. Bouman, and J. P. Allebach, "Multiscale branch and bound image database search," in *Proc. SPIE/IS&T Conf. Storage Retrieval Image Video Databases V*, vol. 3022, Feb. 1997, pp. 133–144.
[30] J. De Bonet and P. Viola, "Texture recognition using a non parametric multi-scale statistical model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1998, pp. 641–647.
[31] N. Vasconcelos and A. Lippman, "A probabilistic architecture for content-based image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2000, pp. 216–221.
[32] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.
[33] İ. Avcibaş, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imag.*, vol. 11, pp. 206–223, Apr. 2002.
[34] J. A. Garcia, J. Fdez-Valdivia, R. Rodriguez-Sánchez, and X. R. Fdez-Vidal, "Performance of the Kullback-Leibler information gain for predicting image fidelity," in *Proc. IEEE Int. Conf. Pattern Recognition*, vol. III, 2002, pp. 843–848.
[35] S. G. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 2091–2110, Dec. 1989.
[36] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1423–1443, May 2001.
[37] E. P. Simoncelli and E. H. Adelson, "Subband transforms," in *Subband Image Coding*, J. W. Woods, Ed. Norwell, MA: Kluwer, 1990, ch. 4, pp. 143–192.
[38] J. B. Anderson, *Source and Channel Coding: An Algorithmic Approach*. Norwell, MA: Kluwer, 1991.
[39] H. R. Sheikh, 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/
[40] Sarnoff Corporation. JNDmetrix Technology. [Online]. Available: http://www.sarnoff.com/products_services/video_vision/jndmetrix/
[41] 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.
[42] VQEG. (2000, Apr.) Final report from the video quality experts group on the validation of objective models of video quality assessment. [Online]. Available: http://www.vqeg.org/
[43] 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, pp. 163–178.
[44] ——, "A visual discrimination mode for image system design and evaluation," in *Visual Models for Target Detection and Recognition*, E. Peli, Ed, Singapore: World Scientific, 1995, pp. 207–220.
[45] F. A. P. Petitcolas, "Watermarking schemes evaluation," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 58–64, Sep. 2000.
[46] Y. Liu and Y. Li, "Error concealment of digital images using data hiding," presented at the 9th DSP Workshop, Hunt, TX, Oct. 2000.
[47] P. Yin, B. Liu, and H. H. Yu, "Error concealment using data hiding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 3, Salt Lake City, UT, May 2001, pp. 1453–1456.
[48] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, pp. 49–71, Dec. 2000.



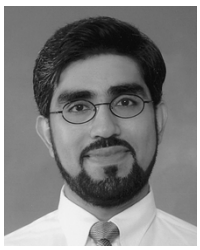
Zhou Wang (S'97–A'01–M'02) received the B.S. degree from Huazhong University of Science and Technology, Wuhan, China, in 1993, the M.S. degree from South China University of Technology, Guangzhou, China, in 1995, and the Ph.D. degree from The University of Texas at Austin in 2001.

He is currently an Assistant Professor at the Department of Electrical Engineering, The University of Texas (UT) at Arlington. Before joining UT Arlington in 2005, he was a Howard Hughes Research Associate at New York University, and a Research Engineer at AutoQuant Imaging, Inc. In the summers of 2000 and 2001, he was with IBM T. J. Watson Research Center. His research interests include image and video processing, quality assessment, and coding; multimedia communications; computer vision and pattern recognition; biomedical image and signal processing; and computational aspects of visual perception. He has more than 40 publications in these fields, including the lecture book Modern Image Quality Assessment (Morgan & Claypool, 2006).

Dr. Wang is an Associate Editor of IEEE SIGNAL PROCESSING LETTERS.

Guixing Wu received the B.S. degree from Huazhong University of Science and Technology, Wuhan, China, the M.S. degree from Shanghai Jiaotong University, Shanghai, China, and the Ph.D. degree from Unversity of Waterloo, Waterloo, ON, Canada, in 1993, 1995, and 2006, respectively.

Since 2001, he has been a Research Assistant at the Leitch-University of Waterloo Multimedia Communications Laboratory, Department of Electrical and Computer Engineering, University of Waterloo. His research interests include digital watermarking and multimedia compression.



Hamid Rahim Sheikh (S'93–M'04) received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 1998 and the M.S. and Ph.D. degrees from The University of Texas, Austin, in 2001 and 2004, respectively.

He is currently affiliated with Texas Instrument, Inc., Dallas, TX. His research interests include full-reference and no-reference quality assessment, application of natural scene statistics models and human visual system models for solving image and video processing problems, and image and video codecs and their embedded implementation.

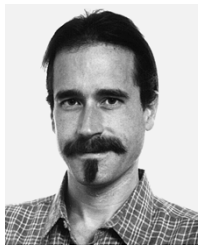
Eero P. Simoncelli (S'92–M'93–SM'04) received the B.S. degree in physics in 1984 from Harvard University, Cambridge, MA, studied applied mathematics at Cambridge University, Cambridge, U.K., for a year and a half, and then received the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1988 and 1993, respectively.

In 1993, he joined the Computer and Information Science Department, University of Pennsylvania, Philadelphia, as an Assistant Professor. He moved to New York University in September of 1996, where he is currently an Associate Professor in Neural Science and Mathematics. In addition, he became an Associate Investigator of the Howard Hughes Medical Institute, New York, in 2000, as part of their new program in Computational Biology. His research interests span a wide range of topics in the representation and analysis of visual images, in both machine and biological vision systems.

En-Hui Yang (M'97–SM'00) received the B.S. degree in applied mathematics from Huaqiao University, Quanzhou, China, in 1986, and the Ph.D. degree in mathematics from Nankai University, Tianjin, China, in 1991y.

He joined the faculty of Nankai University in June 1991 and was promoted to Associate Professor in 1992. From January 1993 to May 1997, he held positions of Research Associate and Visiting Scientist at the University of Minnesota, Minneapolis–St. Paul; the University of Bielefeld, Bielefeld, Germany; and the University of Southern California, Los Angeles. Since June 1997, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he is now a Professor and Canada Research Chair in information theory and multimedia compression. He is the founding Director of the Leitch University of Waterloo MULTIMEDIA Communications Laboratory, and a Co-Founder of SlipStream Data, Inc. His current research interests are multimedia compression, multimedia watermarking, multimedia transmission, digital communications, information theory, Kolmogorov complexity theory, source and channel coding, quantum information theory, and applied probability theory and statistics.

Dr. Yang is a recipient of several research awards, including the 1992 Tianjin Science and Technology Promotion Award for Young Investigators; the 1992 third Science and Technology Promotion Award of Chinese National Education Committee; the 2000 Ontario Premier's Research Excellence Award, Canada; the 2000 Marsland Award for Research Excellence, University of Waterloo; and the 2002 Ontario Distinguished Researcher Award.



Alan Conrad Bovik (S'80–M'81–SM'89–F'96) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois, Urbana-Champaign, in 1980, 1982, and 1984, respectively.

He is currently the Curry/Cullen Trust Endowed Chair in the Department of Electrical and Computer Engineering, The University of Texas, Austin, where he is the Director of the Laboratory for Image and Video Engineering (LIVE) in the Center for Perceptual Systems. During the Spring of 1992, he held a visiting position in the Division of Applied Sciences, Harvard University, Cambridge, MA. He is the editor/author of the *Handbook of Image and Video Processing* (New York: Academic, 2000). His research interests include digital video, image processing, and computational aspects of visual perception, and he has published over 350 technical articles in these areas and holds two U.S. patents.

Dr. Bovik was named Distinguished Lecturer of the IEEE Signal Processing Society in 2000, received the IEEE Signal Processing Society Meritorious Service Award in 1998, the IEEE Third Millennium Medal in 2000, the University of Texas Engineering Foundation Halliburton Award in 1991, and is a two-time Honorable Mention winner of the International Pattern Recognition Society Award for Outstanding Contribution (1988 and 1993). He was named a Dean's Fellow in the College of Engineering in 2001. He is a Fellow of the IEEE and has been involved in numerous professional society activities, including Board of Governors, IEEE Signal Processing Society, 1996–1998; Editor-in-Chief, IEEE TRANSACTIONS ON IMAGE PROCESSING, 1996–2002; Editorial Board, The PROCEEDINGS OF THE IEEE, 1998–present; and Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin in November 1994. He is a Registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.