Subjective Evaluation of Spatial Resolution and Quantization Noise Tradeoffs

Soo Hyun Bae, *Student Member, IEEE,* Thrasyvoulos N. Pappas, *Fellow, IEEE,*
and Biing-Hwang Juang, *Fellow, IEEE*

Abstract— **Most full-reference fidelity/quality metrics compare the original image to a distorted image at the same resolution assuming a fixed viewing condition. However, in many applications, such as video streaming, due to the diversity of channel capacities and display devices, the viewing distance and the spatiotemporal resolution of the displayed signal may be adapted in order to optimize the perceived signal quality. For example, at low bitrate coding applications an observer may prefer to reduce the resolution or increase the viewing distance to reduce the visibility of the compression artifacts. The tradeoff between resolution/viewing conditions and visibility of compression artifacts, requires new approaches for the evaluation of image quality that account for both image distortions and image size. In order to better understand such tradeoffs, we conducted subjective tests using two representative still image coders, JPEG and JPEG 2000. Our results indicate that an observer would indeed prefer a lower spatial resolution (at a fixed viewing distance) in order to reduce the visibility of the compression artifacts, but not all the way to the point where the artifacts are completely invisible. Moreover, the observer is willing to accept more artifacts as the image size decreases. The subjective test results we report can be used to select viewing conditions for coding applications. They also set the stage for the development of novel fidelity metrics. The focus of this paper is on still images, but it is expected that similar tradeoffs apply to video.**

Index Terms— **Scalability, image quality, image fidelity, noise visibility, just noticeable distortion, JND, human visual perception.**

I. INTRODUCTION

RECENT advances in video capture and display technologies and digital communications have led to the development of a wide variety of video services. The spatiotemporal resolution of the video signals that these services provide depends on the video capture device, the transmission bandwidth, and the display system. In many scenarios, a video sequence may be transmitted to a variety of users with different bandwidths and different display devices. This gives rise to the need of a scalable scheme for maximal results. A scalable scheme may be implemented at the source, where a scalable coder is employed, or embedded in the transmission protocol, which requires that the source bit-stream be transcoded

somewhere along the transmission path to support the service needs in quality while minimizing the throughput demand on the network. In any event, the transmitted video streams need to be adjusted according to the channel bandwidth and the display device of the user.

The adaptive scaling scheme also involves the user's viewing conditions. As an example, in low bitrate coding applications, the compressed image or video may be too distorted and the viewer may prefer to reduce the resolution (or size) or increase the viewing distance to reduce the visibility of the compression artifacts. Such kind of situation has at least two implications that warrant investigation. First, existing subjective evaluations or objective image fidelity/quality metrics based on a fixed set of viewing conditions may not be reasonably extrapolated for measuring the perceived image quality under such potentially mismatched viewing conditions. Second, since most fidelity metrics (and coding schemes that are guided by them) were designed under the assumption of fixed viewing conditions, they only measure the noise/distortion visibility for a given image, and do not account for changes in the signal visibility when the viewing conditions (e.g., resolution or viewing distance), and hence, the signal itself, changes. For example, when the viewing distance is infinite, the noise becomes invisible, but so does the signal!

In order to cope with the new application scenarios, there is a need for a fundamental change in our understanding of image fidelity assessment, both subjective and objective. The goal of this paper is to highlight this need and to explore some of the tradeoffs that must be addressed by those who design subjective experiments or develop fidelity metrics.

We present a subjective study that measures the effects of viewing conditions on perceived image quality. In particular, our goal is to explore tradeoffs between spatial resolution and image compression artifacts in order to obtain the optimal display conditions for an image that has been compressed by a given algorithm at a given bitrate. As we will discuss below, there are similar tradeoffs between viewing distance and compression artifacts. The focus of this paper is on still images, but we expect that similar tradeoffs apply to video. Of course, formal experiments would be necessary to quantify such tradeoffs, as well as tradeoffs that arise from varying the temporal resolution of the video sequence.

First we clarify two terms: image fidelity and image quality since these two have been interchangeably referred to in the literature. In this paper, our use of these two terms carries the following somewhat unconventional notions. Image fidelity is a measure that always involves a reference; it can be either

objective under a prescribed set of conditions or subjective for same or different conditions. In the case that an objective image fidelity metric employs a perceptual model, it is referred to as a perceptual image fidelity metric. On the other hand, image quality denotes the perceived characteristics of an image. Such an assessment may or may not involve a reference signal; its measurement is primarily based on the viewers' long-term psychophysical experience. In this paper, when there is an explicit comparison with a reference image, we will use the term image fidelity. When the comparison is indirect or implicit, we will use the term image quality.

The paper is organized as follows. In the remainder of this introduction, we discuss the motivation, methodology, and prior work. Section II reviews objective image fidelity metrics. The setup for the subjective tests is presented in Section III. Section IV presents and analyzes the experimental results. The concluding remarks are provided in Section V.

A. Motivation and Methodology

The primary goal of this paper is to test the *hypothesis* that there exist *unconventional* tradeoffs between the spatial resolution of a given image and its perceived quality, and to provide guidelines for determining the most efficient spatial resolution at a given level of noise. This is in contrast to other psychophysical experiments that try to quantify parameters of an algorithm, e.g., visual thresholds. Our main interest here is noise that arises from compression artifacts, but while rigorous studies are necessary, we expect to see similar types of tradeoffs in other distortions and applications.

Before introducing the methodology of perceptual analysis, let us first state three fundamental assumptions of the human psychophysical behavior as related to visual perception.

1) Relatively, a human observer perceives an image of higher spatial resolution as having higher quality, provided that the image is a natural one without obvious known artifacts.
2) Relatively, a human observer rates an image as of lower quality if it contains more compression artifacts.
3) There exist varying threshold levels of noise visibility below which an observer cannot perceive the noise. These are commonly referred to as just-noticeable distortion (JND) levels.

With these axiomatic assumptions, one key hypothesis to be verified and an objective to be accomplished in our current study can be stated as follows.

1) The visibility of compression artifacts (or noise visibility) decreases as the spatial resolution decreases. At the same time, the signal visibility (or presence in a more general sense) also decreases with decreasing spatial resolution, albeit at a different rate from the noise visibility. Consequently, there exist some *unconventional* tradeoffs between the spatial resolution of a given image and its perceived quality.
2) Quantification of the thresholds for noise visibility may provide a guideline for determining the most efficient spatial resolution at a given noise level.

In order to to verify the above hypothesis and to gain a better understanding of different tradeoffs and the related quantification of perceptual parameters, we conducted two subjective experiments.

The goal of the first experiment is to find the just-noticeable level of distortion (JND) for a given image at a given resolution and viewing distance, that is, to find the lowest bitrate at which no compression artifacts are visible. We refer to this as the *just-noticeable noise perception assessment (JNNP).* Conversely, for a fixed bitrate and viewing distance, this experiment can be used to find the highest resolution at which the compression artifacts are not visible. In principle, the JND levels can be predicted using existing perceptual fidelity metrics [1]. The goal of this experiment is then to verify the predictive capability of the metrics.

The goal of the second experiment is to explore the tradeoffs between noise and signal visibility. For a given encoding rate and viewing distance, the goal of this experiment is to determine the spatial resolution that gives the best overall subjective quality. We refer to this as the *relative perceived quality assessment (RPQ).* In this experiment, each subject was presented with a set of images at various spatial resolutions while the resolution of the display (in pixels per degree) was fixed. This setup allows easy administration of the assessment procedure, as the only thing that changes is the size of the displayed image. The subjects were asked to select one displayed image that is of the highest subjective quality. The subjects were instructed to base their decisions on the overall image quality, i.e., including both distortion artifacts and signal visibility. The images were obtained by downsampling (i.e., anti-alias filtering and decimation) from the same image, which was encoded and reconstructed using two still image coders (JPEG and JPEG 2000) at several encoding rates. In other words, the encoding was all done at the highest resolution. An alternative would be to downsample the original image first, and then encode it at different bit rates. Both alternatives are reasonable, but they address different applications. Our primary interest here is in applications in which the receiver has no control over the transmitter or encoder, and is simply trying to optimize the displayed image quality.

We also made an attempt to relate the results of our subjective tests to existing objective image fidelity metrics. As we saw above, the existing fidelity metrics assume that the original and distorted image are at the same resolution and viewing conditions, and hence, *cannot* provide quantitative estimates of image quality across resolutions or viewing distances. Moreover, most of the existing metrics have been calibrated for CRT displays, and a direct comparison with our experiments, which were conducted with flat panel displays, is difficult.

Based on the test results, we experimentally verified that there exist sensible tradeoffs as hypothesized. Furthermore, we found that the acceptable noise visibility level under varying signal visibility conditions (spatial resolution) is in general significantly higher than the conventional JND thresholds under fixed viewing conditions. This can be understood as a human observer tends to be willing to accept more compression arti-

facts than what may be predicted by existing fidelity measures as long as the operating resolution does not unduly increase the visibility of the artifacts. These findings suggest that a new measure of visual fidelity must take into account the spatial resolution discrepancy between the original image and the viewed image, and that many image coding applications can achieve higher coding performance by exploiting the human psychophysical behavior in a variety of viewing conditions. The ultimate goal, of course, is to develop novel fidelity metrics for a wide range of applications that involve variable bitrates and display devices. Indeed, the subjective evaluations we present in this paper set the stage for the development of such metrics.

B. Prior Work

The study of image quality as a function of viewing distance, resolution, and picture size was the topic of early work by Westerink and Roufs [2], who found that angular resolution (in cycles per degree) and picture angle each influence image quality independently. They found that subjective quality increases with resolution but saturates at approximately 25 cycles per degree. Their results also indicate a linear relationship between subjective quality and the logarithm of the picture angle. Their experiments, which used slide projections of complex scenes as stimuli, set the foundations of modern image quality analysis and had a major impact on the design of the high-definition TV standard. The effect of the display resolution, image size, and other display parameters on the perception of image quality was considered by Barten [3]–[5], who proposed a metric, the square root integral (SQRI), that given the modulation transfer function of a particular display device, can be used to estimate the optimal viewing distance for a given spatial resolution. However, neither Westerink and Roufs nor Barten took the effects of compression artifacts into consideration. Since then the topic has received little attention in the literature. In the mean time, new communications applications have evolved that require a new look at the parameters that influence image quality, and especially tradeoffs between them, e.g., angular (or spatial) resolution versus picture angle (or image size). Kuhmünch *et al.* [6] considered the tradeoff between temporal and spatial resolution in the context of scalable video coding. They proposed a video fidelity metric that is based on the concepts introduced by Webster *et al.* in [7] and used it to find the ratio between spatial and temporal scaling that maximizes perceived quality as measured by the metric. The proposed metric obtains separate estimates of static and dynamic image quality, and then combines them additively or multiplicatively. Although their work provides a valuable approach for video scalability, they do not address the noise versus signal visibility tradeoff considered in this paper. Feghali *et al.* in [8] considered tradeoffs between temporal resolution and quantization, and proposed a new, empirically-derived, metric that takes into account quantization errors, frame rate, and motion speed. The viewing conditions were fixed, however. Koumaras *et al.* [9], [10] considered subjective and objective estimates of upper and lower bounds in the perceived quality of video clips with different resolutions and spatio-temporal activity levels. Finally, Schilling and Cosman

[11] evaluated a number of progressive coders based on the assumption that the time it takes a human observer to recognize a given image relates to image fidelity.

II. OBJECTIVE IMAGE FIDELITY METRICS

The study of multimedia communication applications (speech, image, and video), which inevitably involve some types of signal distortion, requires the evaluation of signal fidelity in its reconstructed form. This can be done with objective criteria, but since the ultimate user is usually a human, any such evaluation or assessment should consistently reflect human perceptual preference. In this section, we review objective measures for still image fidelity that have been proposed with varying degrees of success in predicting the subjective preference.

The most commonly used objective fidelity measure is the mean squared error (MSE), most commonly expressed as peak signal-to-noise ratio (PSNR), which is known to be inadequate as a measure of perceptual distortion. A number of perceptual measures have also been proposed. These measures have relied on certain explicit low-level models of human perception that account for sensitivity to subband noise as a function of the spatial frequency, the local luminance, and the contrast or texture masking [1], [12]. Another recently proposed class of fidelity measures, known as Structural SIMilarity (SSIM) [13], uses implicit perceptual models to account for high-level characteristics of the human visual system (HVS). Such measures take into account point-by-point distortions that may not be relevant to perception of quality, such as spatial translation and intensity shifts, as well as contrast and scale changes. We review all these measures in this section.

A. Perceptual Metrics with Explicit Visual Model

Most of the existing perceptually based image fidelity measures incorporate explicit models of human perception [1], [12]. As mentioned earlier, these measures assume that the reference and the processed (reconstructed) images have the same resolution, and are viewed at a prescribed distance, i.e., under identical viewing conditions. Most measures are based on a multi-scale spatial frequency decomposition using methods such as discrete wavelet transform (DWT), filterbank (i.e., the subband method), or discrete cosine transform (DCT). Perceptual sensitivity usually can be better addressed in the transform domain than with the original pixel array.

Our focus here is on measures that have been developed specifically for image compression applications. We assume that a proper transformation that is commensurate with the adopted perceptual model has been performed, resulting in the set of coefficients $\{b_{i,k}\}$, where k denotes the index in the transform domain and i is the location index of the transformed image block. For each coefficient of the decomposed signal, such measures employ a noise visibility threshold $t_{i,k}$, referred to as the just noticeable distortion level or *JND*, to quantify the perceived distortion. This threshold accounts for human visual sensitivity to spatial frequency, local luminance, and contrast/texture masking. The distortion for the overall image

is then computed as:

$$D_p = \frac{1}{N} \left(\sum_{i,k \in \Re} \left(\max \left\{ \frac{\left| \hat{b}_{i,k} - b_{i,k} \right|}{t_{i,k}}, 1 \right\} \right)^{Q_s} \right)^{\frac{1}{Q_s}}, \quad (1)$$

where $\hat{b}_{i,k}$ is the coefficient of the processed (or distorted) image corresponding to $b_{i,k}$, and N is the total number of coefficients (equivalently, pixels). The value of Q_s is empirically determined from psychophysical experiments. (For a discussion, see [1].) Note that when the difference of the two coefficients is below the visibility threshold, the noise is essentially invisible and the actual value of the difference does not matter. When the difference is higher than the threshold, it is normalized by the JND level. Thus, D_p represents the average distortion expressed in JNDs.

In order to be consistent with traditional error metrics, we express the perceptual metric in terms of "visual decibels (dB)." We define the "masked peak signal-to-noise ratio (MP-SNR)" as

$$\mathrm{MPSNR} = 10 \log_{10} \frac{255^2}{D_p^2}. \quad (2)$$

Note that the maximum value of MPSNR is 48.13 dB, which corresponds to the perceptually transparent condition. For this paper, we considered several measures, one developed by Safranek and Johnston for subband coders [14], one by Watson for DCT coders [15][6], and another by Watson *et al.* for wavelet-based coders [16]. A detailed description of the measures can also be found in [1]. Even though these measures were developed for near-threshold applications, they have also been used in supra-threshold applications [17], [18]. More systematic studies of the supra-threshold case can be found in [19]–[22].

B. Structural Similarity Metrics

In contrast to the perceptual metrics we described above, the Structural SIMilarity (SSIM) metrics, proposed by Wang *et al.* [13], [23], are not based on measurements of noise sensitivities; instead, they attempt to take into account higher-level functionalities of the HVS, and in particular, they attempt to make explicit use of the "structural" information in the viewing field. An important property of the SSIM measures is that they are only supposed to respond to significant structural changes, while perceptually insignificant point-by-point distortions, such as contrast and intensity changes, are not substantially penalized. These measures assume that the structural information is available or can be extracted from the image. Thus, they are expected to be more effective in measuring supra-threshold compression distortions, which affect the structure of an image.

There are several SSIM implementations, both in the image/pixel space and in the wavelet domain. The basic SSIM index proposed in [13] is a real number in the range $[-1, 1]$ and is computed based on the second order statistics of the reference and the distorted image as follows:

$$S(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x \mu_y + K_1)(2\sigma_{xy} + K_2)}{(\mu_x^2 + \mu_y^2 + K_1)(\sigma_x^2 + \sigma_y^2 + K_2)}, \quad (3)$$

where $\mathbf{x}$ and $\mathbf{y}$ are two nonnegative image signals (or image patches), μ_x and μ_y are the mean intensities, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance of $\mathbf{x}$ and $\mathbf{y}$, and K_1 and K_2 are small real constants relative to μ_x or μ_y. A more general form of this metric can be found in [13]. The spatial domain SSIM has been shown to provide good quality prediction across a variety of artifacts, but is highly sensitive to spatial translation.

The complex wavelet domain implementation (CWSSIM) [24] allows imperceptible spatial translations, and also small rotations and scaling changes. The CWSSIM of a given subband is given by

$$S_c(\mathbf{c}_x, \mathbf{c}_y) = \frac{2 \left| \sum_i c_{x,i} c_{y,i}^* \right| + K}{\sum_i |c_{x,i}|^2 + \sum_i |c_{y,i}|^2 + K}, \quad (4)$$

where $\mathbf{c}_x$ and $\mathbf{c}_y$ are the wavelet coefficients corresponding to two images or image patches, $\mathbf{c}^*$ denotes the complex conjugate of $\mathbf{c}$, and K is a small positive constant.

Note that the mean of the wavelet coefficients (except the baseband) is zero due to the bandpass property of the wavelet transform. The overall metric value is computed as the mean of the CWSSIM subband indices. Brooks *et al.* [25], [26] proposed a variation of this metric, whereby the subband indexes are weighted based on human sensitivity to subband noise. The weighted CWSSIM incorporates an explicit model of subband sensitivity to noise, and thus, provides a link to the perceptual metrics described above.

Overall, even though SSIM metrics have introduced a new way of looking at image fidelity, they are also limited by the fact that the reference and test images are at the same resolution and viewing conditions.

C. Signal and noise visibility

As will be discussed in the next section, the aim of our subjective experiments is to measure the effect of viewing conditions (viewing distance or spatial resolution) on the perceived image quality. As the amount of distortion increases, say due to reduced bandwidth or bitrate, it is conjectured that the perceived quality may benefit from a reduction in the spatial resolution, or similarly, from an increase in the viewing distance. Of course, changing the viewing distance is not equivalent to changing the spatial resolution, as one has to take into account the specifics of the sampling rate conversion and characteristics of the display device [27]. However, to a first order approximation, we can safely assume that reducing the resolution is equivalent to increasing the viewing distance by the same factor. The fidelity measures that have been proposed so far cannot provide a quantitative estimate of the image quality when the image resolution or the viewing distance has been altered. One obvious approach to evaluating the visual fidelity across spatial resolution is to evaluate the images at the highest resolution using one of the available measures. A problem with this approach is that it requires image interpolation, and it is not clear what interpolation techniques should be used since they may introduce unexpected artifacts. More importantly, this approach is essentially inconsistent with the way the reconstructed images are viewed or used by a viewer.
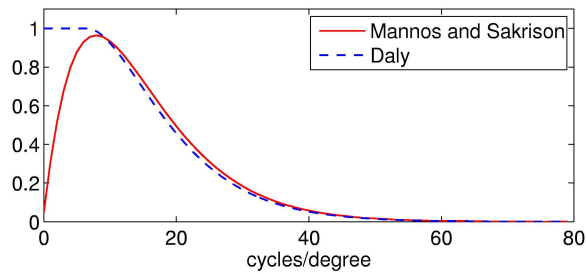
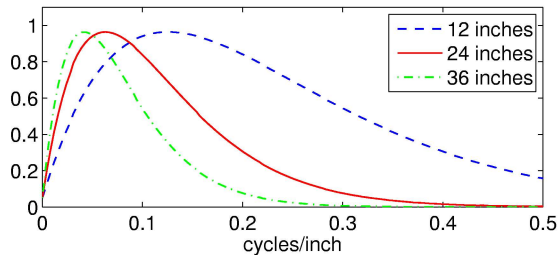Fig. 1. The Mannos-Sakrison and Daly eye contrast sensitivity functions.

Fig. 2. The Mannos-Sakrison frequency response with frequency expressed in cycles/inch at several viewing distances.

In order to explain in simple terms the issues involved, we consider the contrast sensitivity function (CSF) of the eye. However, we should point out that the CSF is too simple of a model of the eye for the purpose of evaluating the perceived image fidelity. An estimate of the CSF by Mannos and Sakrison [28] is shown in Fig. 1 in solid red line. According to the estimate, the eye is a bandpass filter, showing increasing sensitivity as the spatial frequency rises, peaking at about 8 cycles per degree, and then decaying exponentially towards higher frequencies. Also shown in dashed blue line is the frequency response used by Daly [29]; it models the eye as a lowpass filter instead of a bandpass one. As illustrated in Fig. 2, the frequency response changes substantially with the viewing distance. For example, as the viewing distance increases, the peak of the response shifts to lower frequencies, making high frequency details less visible (for more details, see [30]).

The key parameters for the CSF are the viewing distance (in inches) and the resolution of the display device (in pixels per inch). Alternatively, one can specify the viewing distance in image heights and the image height in pixels (assuming identical horizontal and vertical display resolution). In either case, one must derive the "display visual resolution" in pixels per degree [16]. Note that, since the CSF demonstrates a bandpass characteristic, it is possible that (low-frequency) image degradations become more visible as the viewing distance increases [12], [31]. In order to compensate for this undesirable effect, several researchers have proposed "flattening" of the eye response [29], [32]–[35] as shown in Fig. 1.

Based on the above discussion, it is clear that increasing the viewing distance reduces the visibility of compression artifacts. However, increasing the viewing distance also decreases the visibility of the signal itself. In the extreme case of an infinite viewing distance, no artifacts are visible, but neither is the signal. Thus, there is a need to find a viewing distance that achieves the best balance between the signal and the noise visibility so as to achieve the best overall image quality.

III. SUBJECTIVE EXPERIMENTS

To obtain a better understanding of how the spatial resolution affects the perceived signal quality in the context of compression artifacts, we conducted subjective image quality evaluation experiments. Early psychophysical experiments for analyzing the effect of spatial resolution on image quality assessment were conducted by Westerink and Roufs [2]. Such experiments have formed the basis of modern image quality analysis and had a major impact on the design of the high-definition TV standard. The goal of this paper is to consider specific tradeoffs that were not in that early study, and that are encountered in current communications applications, which encompass a wide variety of display devices and channel capacities. In particular, our goal is to explore tradeoffs between spatial resolution and image compression artifacts in order to obtain the optimal display conditions for an image that has been compressed by a given algorithm at a given bitrate.

We conducted two experiments. In both cases, the viewing distance was fixed. We used a series of compressed images at different bitrates that were carefully chosen to cover a wide range of perceptual quality. These images were then down-sampled using near-optimal sinc-function anti-aliasing filters. As we noted in the introduction, all the encoding was done at the highest resolution. The goal of the *just-noticeable noise perception assessment* experiment was, at each resolution, to find the highest compression level at which the artifacts are not visible, or conversely, for a fixed bitrate and viewing distance, to find the highest resolution at which the compression artifacts are not visible. The goal of the *relative perceived quality assessment* experiment was to determine the resolution that gives the best overall subjective quality for a given encoding rate.

A. Image generation and experimental setup

In our experiments we used a number of test images and compression techniques. Figure 3 shows thumbnails of the four test images, "Bank," "Lena," "Bike," and "Woman," all with spatial resolution 512×512 pixels and 256 quantization levels. The "Bike" and "Woman" images were cropped from the ISO 400 image set. Note that the "Bank" image has the highest amount of detail, followed by "Bike," "Lena," and "Woman." Note also that in "Lena" the background is blurred, while in both "Lena" and "Bike" there is some visible noise, most likely scanning artifacts.

The test images were first encoded using two different coding algorithms, JPEG [36], [37] and JPEG 2000 [38], [39]. JPEG is the most widely used method for perceptually lossless or lossy image compression. It is based on a DCT decomposition. The JPEG quantization table was based on the baseline contrast sensitivity thresholds determined by Watson [1], [15], computed at six image heights. JPEG 2000 is the most recent standard for image compression, and is based on a DWT decomposition. The number of decomposition levels

Fig. 3.   Images for subjective tests, "Bank", "Lena", "Bike", and "Woman" (in raster scan order).

was set at 5 and the codeblock size was set at 32. Since these compression methods have different coding efficiencies, the bitrates for each image and technique were carefully selected to obtain reasonable differences in perceived quality between images at consecutive bitrates. The selected bitrates for the two compression methods and four images are shown in Table I. In all cases, the highest bitrate was 1.0, at which there are no distortions visible to human observers.

For each coder and bitrate, the reconstructed images were then downsampled by factors of 4/3, 2, 8/3, 4, 16/3, and 8 to obtain seven different resolutions, using integer upsampling and downsampling combinations with near-optimal sinc-function anti-aliasing filters. The viewing distance was fixed at six image heights of the highest resolution (512×512) image, which is equal to 35.4 inches. The stimuli were displayed in a blue background and viewed in a darkened room on a Dell 1905FP flat panel liquid-crystal display (LCD) screen with contrast ratio 800:1 and a resolution of 86.78 pixels/inch. The maximum luminance of this display is $250 \, \mathrm{cd/m^2}$ and the average luminance is $125 \, \mathrm{cd/m^2}$. The resulting spatial resolutions, viewing angles to the image stimuli, and the maximum spatial frequencies are listed in Table II.

Six observers (hereafter, sbj1 to sbj6) took part in the experiments. All were binocular with normal color vision and normal (two) or corrected (four) visual acuity. Five were male and one female. Their age was between 20 and 30 years old. The subject group included a balanced mixture of critical and casual viewers; their familiarity with image processing did not seen to have any significant effect on the test results. All six subjects participated in both experiments. Prior to the test, each observer was presented with the original images and typical examples of compressed images, in order to become familiarized with the test environment. During the test, the

observers were given enough time to make their decisions. In the relative test, they were also allowed to view the original test images (at the highest resolution) at any time during the test. The ordering of images and encoders was randomized to avoid any biases, but as we will see below, the bitrates were not.

For a review of the general methodology of designing a psychophysical test, the reader is referred to [40]–[42].

B. Just-Noticeable Noise Perception Assessment

The goal of the just-noticeable noise perception assessment is to find the distortion level at which the compression artifacts become invisible. In theory, the JND can be predicted using existing perceptual fidelity measures. However, this is not always true in practice, thus making this experiment necessary. This experiment is in contrast to the relative perceived quality test, where we test the expectation that the subjects are willing to accept some compression artifacts in order to obtain a larger image, in which case they should select a higher resolution than the one that corresponds to the JND level.

This experiment consists of a series of two-alternative forced choice (2AFC) tests. Each observer was presented with two images, the original and the encoded image, in random order, and was asked to select the one without compression artifacts. Both images were downsampled to a given resolution using the same algorithm. An example of the stimulus is shown in Figure 4. The test was repeated ten times at each resolution and distortion level (bitrate). If all the answers are correct, this indicates that the noise is visible. If the noise is not visible, then the observer's selection should be random, i.e., 50% of the answers should be correct. In a 2AFC experiment like this, the threshold value for the correct answer is typically taken as the midpoint between the ideal percentages, which corresponds to 75% correct [43]. Therefore, if an observer gives the correct answer eight times or more, then we conclude that the noise is visible; otherwise it is invisible. The lowest level of distortion at which the noise is invisible is the JND. A similar experimental setup was used in [14], [44].

In principle, there should be ten trials for each bitrate, image, and spatial resolution, which adds up to over a thousand trials for each observer. However, since our goal is to determine the threshold of perception, the length of the test can be significantly reduced through the use of a dynamic threshold determination method. Rather than carrying out an exhaustive test, each observer was presented with a limited range of bitrates. The initial bitrate was chosen in the middle of the range of bitrates. If the compression artifacts at this bitrate were determined to be visible, then the bitrate was increased by one step. This procedure was repeated until the artifacts became invisible. Similarly, if the compression artifacts at the initial bitrate were determined to be invisible, then the bitrate was decreased one step at a time until they became visible. In both cases, once the critical coding bitrate was determined, an additional test at the next higher or lower bitrate was conducted, in order to validate the results.

TABLE I
CODING BITRATES FOR THE TEST IMAGES

Image	Coder	Bitrates (bits/pixel)
Bank	JPEG	1.0, 0.8, 0.7, 0.6, 0.5, 0.45, 0.4, 0.3, 0.25, 0.2
	JPEG 2000	1.0, 0.75, 0.6, 0.5, 0.3, 0.2, 0.18, 0.15, 0.12, 0.1, 0.05
Bike	JPEG	1.0, 0.8, 0.6, 0.5, 0.4, 0.3, 0.27, 0.23
	JPEG 2000	1.0, 0.75, 0.5, 0.4, 0.3, 0.2, 0.1, 0.08, 0.05, 0.04, 0.03
Lena	JPEG	1.0, 0.5, 0.4, 0.35, 0.3, 0.27, 0.25, 0.23, 0.2
	JPEG 2000	1.0, 0.5, 0.4, 0.35, 0.3, 0.2, 0.1, 0.08, 0.05, 0.03
Woman	JPEG	1.0, 0.6, 0.5, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.14
	JPEG 2000	1.0, 0.5, 0.3, 0.2, 0.1, 0.08, 0.06, 0.04

TABLE II
SPATIAL RESOLUTIONS AND VIEWING CONDITIONS

Spatial resolution	512	384	256	192	128	96	64
Viewing angle (degree)	9.53	7.15	4.77	3.58	2.39	1.79	1.19
Maximum spatial frequency (cycles/degree)	53.74	40.31	26.87	20.15	13.44	10.08	6.72

Fig. 4.   Test images for the just noticeable noise perception assessment test ("Bank" coded by JPEG at 0.2 bits/pixel).

C. Relative Perceived Quality Assessment

The goal of the relative perceived quality assessment test is to study the effect of spatial resolution on the perceived quality of the displayed images. When the spatial resolution decreases, the visibility of the compression artifacts is reduced and so is the visibility of the signal. One key question to be answered is how the visibility reductions for the signal and the noise relate to each other perceptually. This subjective experiment was designed to answer the question by asking an observer to take both effects into consideration as she/he selects the resolution that maximizes the perceived quality. In the subjective experiment, each subject was presented with a set of images at different spatial resolutions and asked to select the one of highest subjective quality. All the images in the set were downsampled from the same decoded image. Figure 5 illustrates the setup, a seven-alternative forced choice (7AFC) stimulus array. The images are ordered clockwise in increasing resolution in order to facilitate pair-wise comparisons between adjacent resolutions. While the images and compression techniques were randomized, the bitrates were presented in descending order. This is because the observers may be confused if the quality jumps around too much, and also, by moving from higher to lower quality, the observers have a better sense of what they are looking for.

Fig. 5.   Test images for the relative perceived quality assessment test ("Lena" coded by JPEG 2000 at 0.1 bits/pixel).

IV. EXPERIMENTAL RESULTS

We now examine the results of the subjective experiments. Due to a variety of factors, it is reasonable to expect that the results of such subjective evaluations will include a number of outliers. Thus, as is typical in subjective evaluations, in each test we excluded the most extreme (high or low) vote. That is, in the just noticeable noise perception assessment, for each image and resolution, we excluded the observer with the most extreme bitrate, while in the relative perceived quality assessment, for each image and bitrate, we excluded the observer with the most extreme resolution. When there were two extreme votes, we excluded the one with the lower numerical value. We then used both the median and the mean of the remaining votes as the final outcome of the assessment.

A. Just Noticeable Noise Perception Assessment

Figure 6 plots the noise visibility thresholds that were obtained from the Just Noticeable Noise Perception Assessment. Note that the axes of these plots have been interchanged in
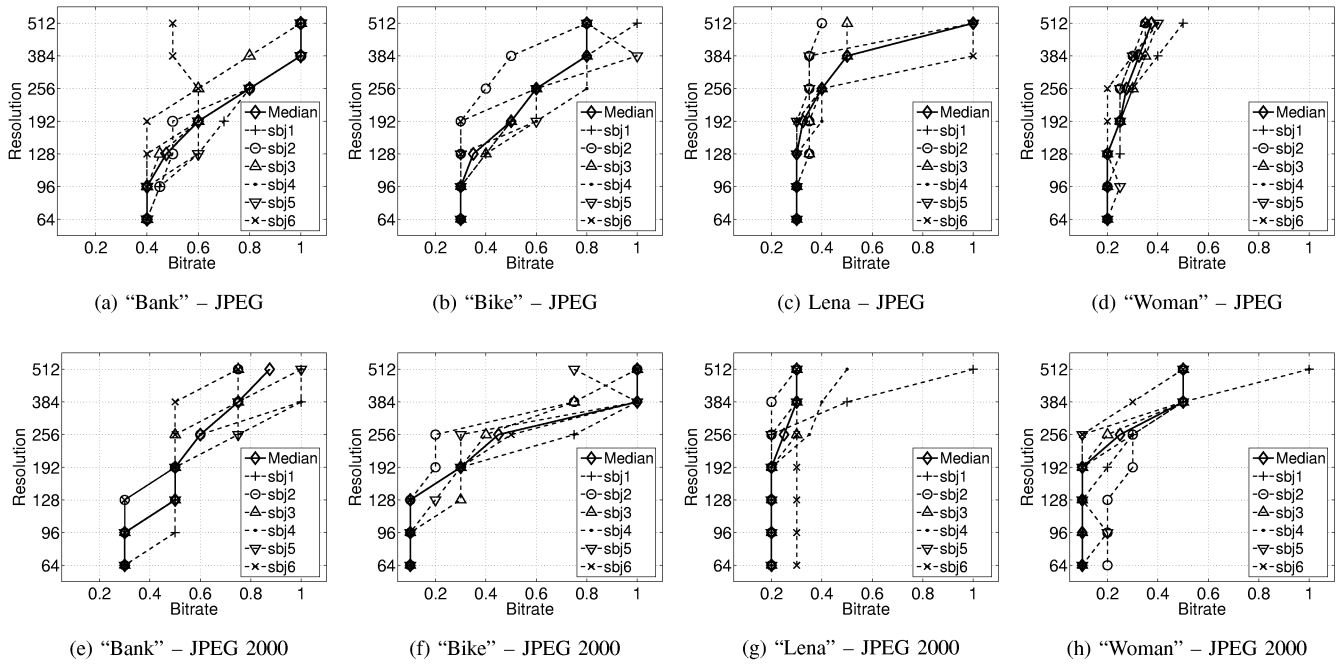
Fig. 6.   Noise visibility thresholds for all the subjects. Their votes are indicated by the dotted lines and their median value is by the solid line.

order to provide consistency with the figures that follow. In general, we expect the noise visibility threshold to decrease as the spatial resolution increases, and thus, the selected bitrate to increase with increasing resolution. Observe that this is not true for all of the results. Several violations of this expectation can be seen in Fig. 6 (a), (b), (d), (f), and (h). For example, note the graph for sbj5 in the "Bike" image of Fig. 6 (b) and (f). This may be explained by the fact that the inherent noise in the original uncompressed image could influence the observer's choice. Image compression algorithms typically drop high frequencies, and this results in image smoothing. When the original image is noisy (or even textured, e.g., in a face), the observers may then consider the smoother compressed image as superior to the noisy original and, thus, label it as original.

The overall JND level (median) is shown in solid line in the figure. At a given resolution, this represents the bitrates at which the compression artifacts become invisible. Note that at low resolutions, the JND levels for images generated by JPEG 2000 correspond to lower bitrates than those for JPEG images, while at higher resolutions, the results are mixed. This should be expected because at lower rates JPEG 2000 provides superior rate-distortion performance, while at higher rates JPEG outperforms JPEG 2000 for highly detailed images (like "Bank", "Bike") but is less efficient for images with low detail, whose quality is not affected by a certain amount of blurring (like Lena and "Woman").

B. Relative Perceived Quality Assessment

Figure 7 depicts the most preferred spatial resolution from the results of the relative perceived quality assessment. As in Fig. 6, the x-axis and the y-axis correspond to the coding bitrate and the spatial resolution, respectively. Note that in Fig. 7 (c), (e), and (g) one can see several measurement

inconsistencies. As we saw above, the final result (solid lines) was obtained as the median values of the subjects' votes after excluding one extreme measurement for each test. Note that as the bitrate decreases, the subjective quality decreases, and that the observers prefer to reduce the spatial resolution in order to reduce the visibility of the compression artifacts. The question is whether they always prefer the resolution at which no compression artifacts are visible, or they are willing to accept some artifacts instead of a further decrease in image resolution (and size). To answer this question, we now compare the results of the two subjective experiments.

C. Comprehensive Analysis and Statistical Validation of Experimental results

Figures 8 and 9 plot the results of the two tests (from Figs. 6 and 7, respectively) on one graph for each image and compression technique. Black solid lines with triangle markers correspond to the median of the measured JND levels. The error bars show the standard deviation and are centered on the mean values, which are marked with blue triangles. The red dashed lines connect the error bars to outline the area within a standard deviation of the measured JND values. The medians of the observer votes for the relative perceived quality test are shown in black solid lines with circle markers. The corresponding error bars are centered on the mean values, which are marked with blue circles. Note that since the tests included a discrete set of bit rates and spatial resolutions, the medians may be more meaningful than the means, but the means and error bars are also included as an indication of the most likely range of values of the respective quantities. Note that for the most part, the most preferred noise-level/resolution points of the relative perceived quality test are well outside the JND range. This is true for both JPEG and JPEG 2000 encoding. Table III shows the ratio of the observer votes in
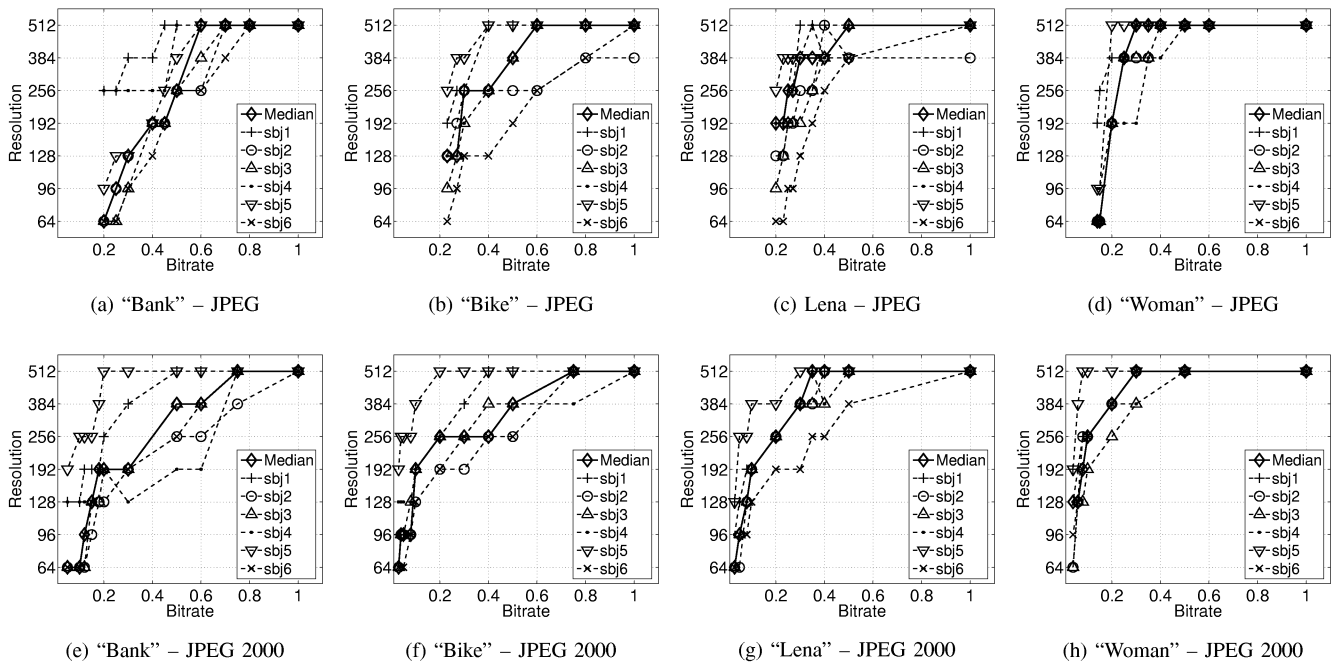
Fig. 7. Most preferred resolution for all the subjects. Their votes are indicated by the dotted lines and their median value is by the solid line.

TABLE III
RATIO OF OBSERVERS' VOTES IN THE RELATIVE PERCEIVED QUALITY TEST THAT FALL WITHIN A STANDARD DEVIATION OF THE JND (%)

	JPEG	JPEG 2000
Bank	1.67	9.09
Bike	14.58	10.06
Lena	18.52	28.33
Woman	26.67	12.50
Average	15.32	15.00

the relative perceived quality test that fall within a standard deviation of the JND level for each image and coding method. Based on these observations, we conclude that the difference between the noise visibility levels that correspond to the most preferred resolution and the JND levels is significant.

A couple of additional observations are in order. At the maximum spatial resolution of 512×512, the relative perceived quality levels saturate before the maximum coding bitrate of 1.0 bpp is reached. For "Lena" coded by JPEG 2000 in Fig. 9 (c), the JND level is lower than the relative perceived quality level at the highest resolution (512×512). Our interpretation of this somewhat unexpected result is that the original "Lena" image has visible noise artifacts that get wiped out by JPEG 2000 compression. Thus, in the JNP test, the observers prefer the lower rate images. On the other hand, in the relative perceived quality test, the observers simply pick the largest image in spite of the noise artifacts. A similar observation holds for the "Woman" image. In this case, there is no noise; it is just that smoother faces look better than textured faces.

For comparison, we also include the predictions of the Safranek-Johnston fidelity metric [14], shown as image intensities using the *heated object color map* [45]. As we discussed, existing metrics assume that the original and distorted image are at the same resolution, and cannot provide

quantitative estimates of image quality across resolutions. Thus, the metric values were computed on the downsampled original and compressed images, and are shown as independent horizontal stripes for each resolution. Note that the predictions of the Safranek-Johnston fidelity metric are mostly consistent with the results of our JNNP test, i.e., the JNNP test line is close to the points where the distortion map reaches its maximum (becomes white). However, there are also significant deviations, e.g., in Figures 7(c) and 8(c). We also compared the results of the other objective measures we discussed in Section II, such as the SSIM and its variations, Watson's DCT metric, as well as wavelet based metrics. We found that their JND predictions do not correspond well with our subjective tests.

For a given spatial resolution, we can say that the difference between the perceptually noise-transparent level and the level that corresponds to the most preferred resolution represents the *perceptual tolerance* of the observer. This perceptual tolerance can be expressed in visual dBs as the difference of the corresponding MPSNRs, or can be converted to JNDs by inverting (2). Thus, for the JPEG encoded "Bank" image, at 256×256, the perceptual tolerance is 0.64 visual dB or 1.076 JNDs. Figures 10(a) and 11(a) show the perceptual tolerances for each image and coding technique as a function of spatial resolution, averaged over the observers. The error bars show the standard deviations. Observe that the perceptual tolerance increases as the spatial resolution decreases. This is true for all images and both compression schemes. Figures 10(b) and 11(b) show the averages over all the images; the standard deviations account for variations both in observer preferences and image content. Note again, that there is a significant difference from the JND level, but the standard deviations are perhaps too high to reliably predict the most preferred resolution for each image.
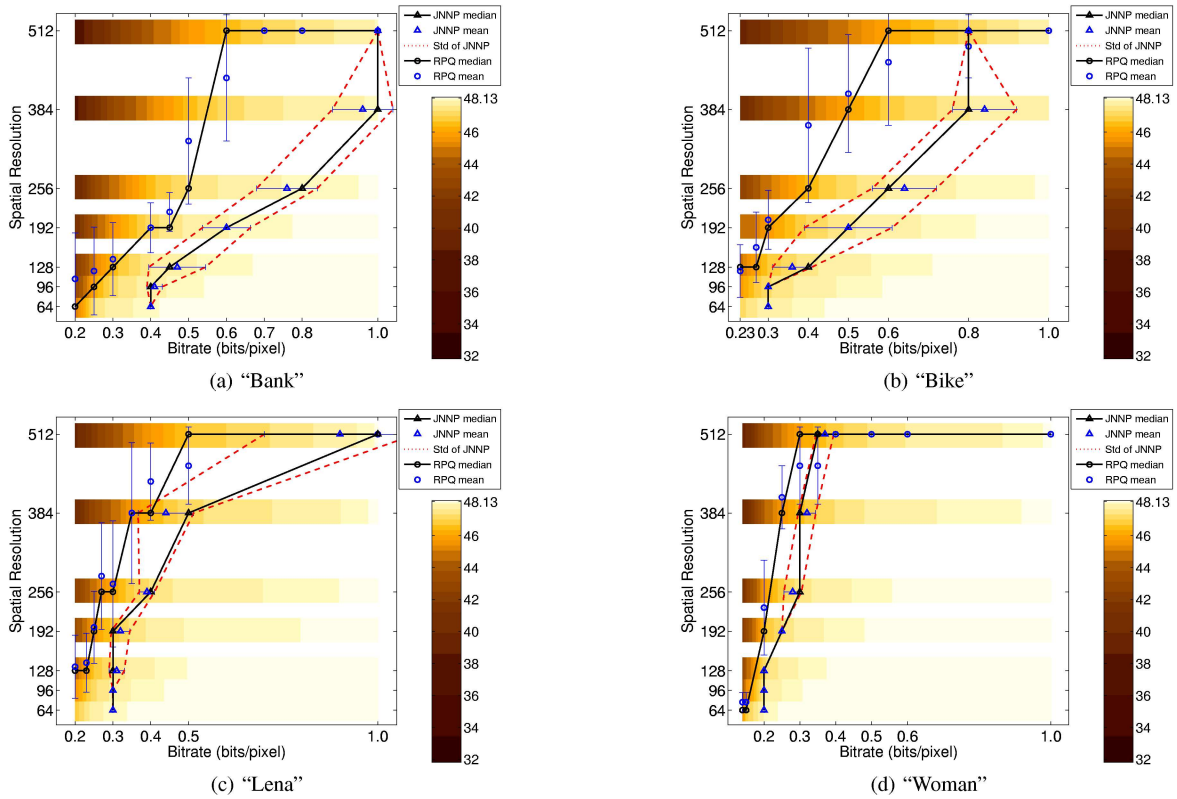
Fig. 8.   Most preferred resolution configurations on distortion map. Fidelity metric is Safranek-Johnston Metric. Image coder is JPEG.
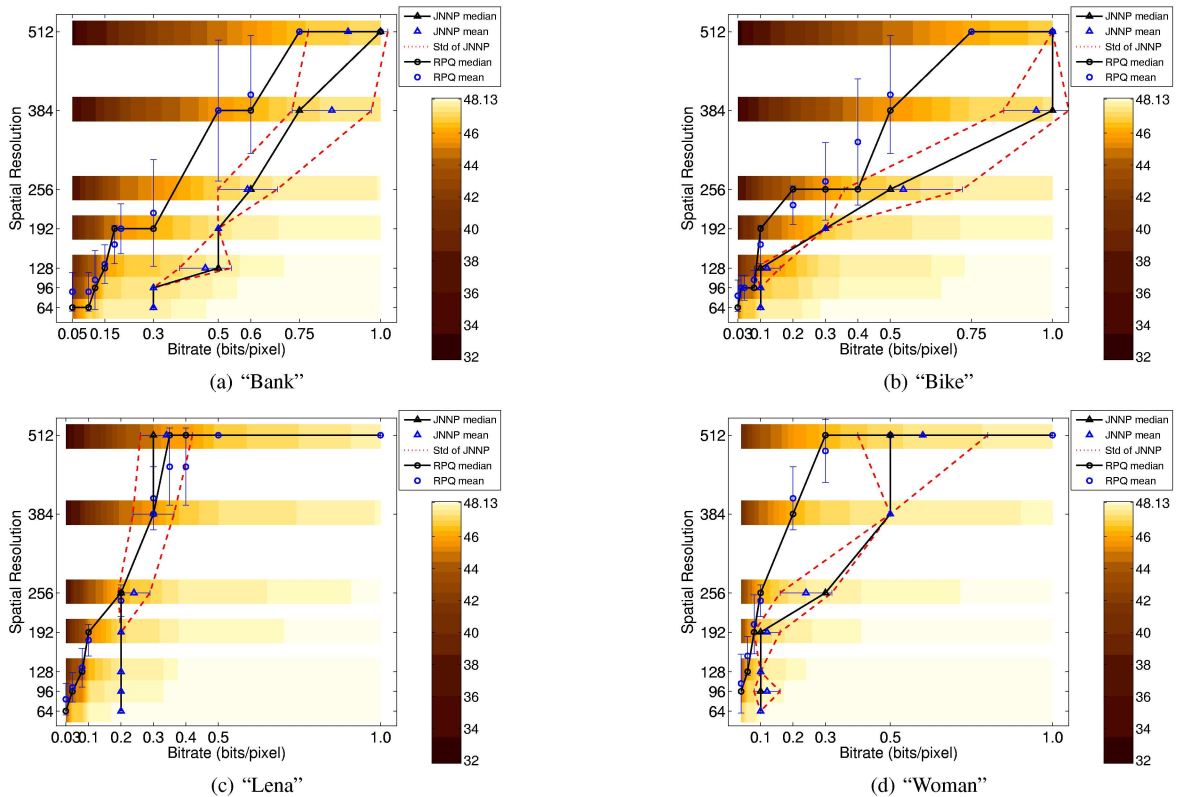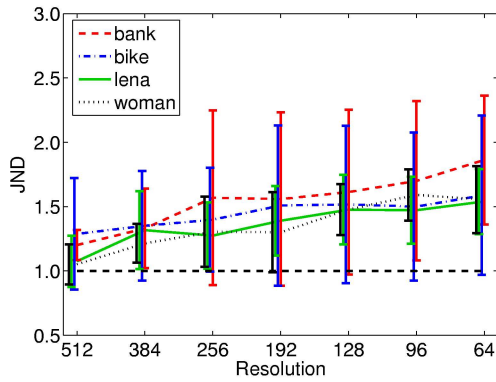

Fig. 9.   Most preferred resolution configurations on distortion map. Fidelity metric is Safranek-Johnston Metric. Image coder is JPEG 2000.
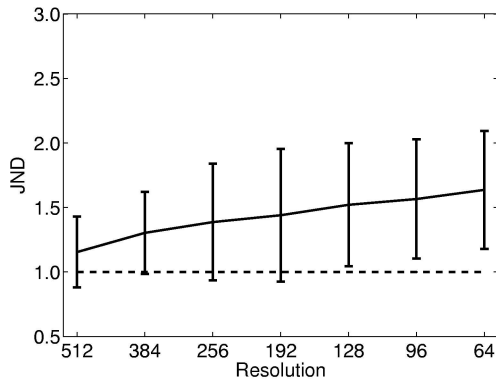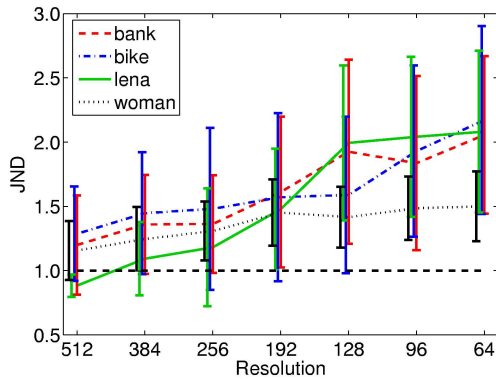
(a) Results for individual images

Fig. 10. Perceptual tolerance expressed in JNDs for JPEG compression.

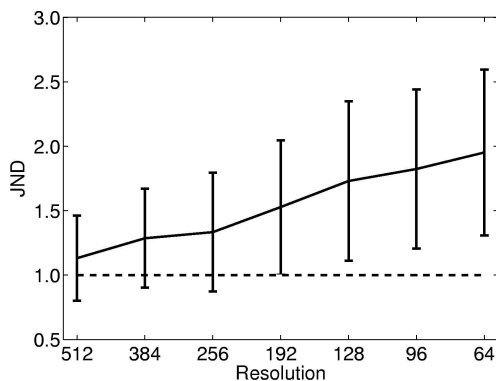(a) Results for individual images

Fig. 11. Perceptual tolerance expressed in JNDs for JPEG 2000 compression.

Another illustration of the results of the two subjective tests is provided in Tables IV and V, which list the Safranek-Johnston metric predictions for each resolution and bitrate, with JND levels shown as shaded cells and the most preferred resolutions shown in boldface numbers. For all the tables, the highest quality is at the lowest spatial resolution and highest bitrate (bottom-right), and the lowest quality is at the highest spatial resolution and lowest bitrate (top-left). Note that because of the way the two experiments were designed, each row contains one shaded box (the lowest bitrate for that resolution at which artifacts are not visible), and each column contains one boldface number (the most preferred resolution).

Overall, based on the results of the two experiments, we can conclude that in most cases the most preferred resolution is higher than the perceptually transparent resolution (JND). In other words, human observers are willing to accept some visible distortion in order to obtain higher resolution. Moreover, the amount of acceptable distortion, what we called perceptual tolerance, increases as the spatial resolution decreases.

V. CONCLUSIONS

We considered tradeoffs between spatial resolution and the visibility of compression artifacts. Such tradeoffs are not reflected in existing fidelity measures, which ignore the signal visibility and only measure the visibility of compression distortions, which decrease with image size. The analysis of such tradeoffs is of importance in applications that involve a wide variety of bitrates and display devices, including scalable image compression applications.

Based on three fundamental assumptions of human psychophysical behavior, we designed two subjective tests to experimentally verify the hypothesis that there are unconventional tradeoffs between spatial resolution and the visibility of compression artifacts. The goal of the just noticeable noise perception assessment test was to obtain the resolution at which no compression artifacts are visible. The goal of the relative perceived quality assessment test was to find, for each image and bitrate, the most preferred resolution on the basis of both image size and visibility of compression artifacts. We used two standard image coders (JPEG and JPEG 2000) and a set of representative images. Our results indicate that the tradeoffs that we hypothesized exist and that the most preferred resolution is higher than the resolution at which there are no visible distortion artifacts. Conversely, the distortion levels at the most preferred resolution are significantly higher than the JND thresholds. In other words, human observers are willing to accept some visible distortion in order to obtain higher resolution. Moreover, the amount of acceptable distortion, which we call perceptual tolerance, increases as the spatial resolution decreases. We also attempted to quantify this perceptual tolerance in terms of the masked peak signal-to-noise ratio. However, our results indicate that the variance of the measurements is too high to reliably predict the perceptual tolerance across different images and observers. The focus of this paper was on still images, but similar tradeoffs also apply to video. Of course, in the video case there are additional tradeoffs to be explored as the temporal resolution can also be varied.

TABLE IV

MPSNR VALUES (dB) OBTAINED BY THE SAFRANEK-JOHNSTON METRIC OVER DIFFERENT CODING RATES AND SPATIAL RESOLUTIONS FOR JPEG CODED IMAGES (JUST-NOTICEABLE LEVELS IN SHADED CELLS, MOST PREFERRED RESOLUTIONS IN BOLD NUMBERS)

Resolution/bpp	0.2	0.25	0.3	0.4	0.45	0.5	0.6	0.7	0.8	1.0
512	37.19	40.13	42.00	44.28	44.99	45.67	**46.44**	**46.99**	**47.36**	**47.96**
384	38.07	41.11	43.05	45.26	45.87	46.50	47.11	47.50	47.73	47.98
256	39.93	43.02	44.74	46.55	46.94	**47.35**	47.69	47.89	47.99	48.09
192	41.05	44.07	45.70	**47.18**	**47.44**	47.70	47.93	48.02	48.08	48.13
128	42.05	45.18	**46.50**	47.61	47.76	47.91	48.03	48.08	48.11	48.13
96	43.10	**46.12**	47.20	47.86	47.95	48.04	48.10	48.12	48.13	48.13
64	**44.92**	47.37	47.79	48.06	48.07	48.11	48.12	48.13	48.13	48.13

(a) "Bank" JPEG

Resolution/bpp	0.23	0.27	0.3	0.4	0.5	0.6	0.8	1.0
512	39.22	40.87	41.70	43.75	45.04	**46.02**	**47.11**	**47.91**
384	40.27	42.00	42.84	44.83	**46.01**	46.81	47.61	47.90
256	42.45	44.10	44.90	**46.45**	47.20	47.64	47.99	48.08
192	44.59	45.08	**45.83**	47.04	47.56	47.85	48.06	48.11
128	**44.81**	**46.10**	46.70	47.54	47.84	48.01	48.11	48.13
96	45.83	46.78	47.24	47.83	47.98	48.08	48.12	48.13
64	47.12	47.60	47.78	48.04	48.10	48.12	48.13	48.13

(b) "Bike" JPEG

Resolution/bpp	0.2	0.23	0.25	0.27	0.3	0.35	0.4	0.5	1.0
512	39.53	41.22	42.12	42.80	43.74	44.96	45.67	**46.69**	**48.09**
384	40.36	42.10	43.02	43.71	44.55	**45.70**	**46.36**	47.22	48.11
256	42.35	43.97	44.82	**45.45**	**46.07**	46.92	47.33	47.81	48.13
192	43.50	44.97	**45.75**	46.24	46.76	47.40	47.69	47.97	48.13
128	**44.61**	**45.90**	46.67	46.95	47.33	47.76	47.92	48.07	48.13
96	45.69	46.66	47.20	47.46	47.69	47.95	48.04	48.11	48.13
64	46.69	47.45	47.82	47.92	48.02	48.08	48.11	48.13	48.13

(c) "Lena" JPEG

Resolution/bpp	0.14	0.15	0.2	0.25	0.3	0.35	0.4	0.5	0.6	1.0
512	38.92	40.16	43.70	45.08	**45.94**	46.46	**46.86**	**47.35**	**47.63**	**48.09**
384	39.39	40.72	44.35	**45.72**	46.53	46.99	47.31	47.69	47.89	48.11
256	40.90	42.36	45.83	46.90	47.45	47.70	47.86	48.03	48.09	48.13
192	41.72	43.22	**46.54**	47.33	47.75	47.90	48.00	48.08	48.11	48.13
128	42.66	44.27	47.24	47.71	47.96	48.04	48.09	48.12	48.13	48.13
96	43.72	45.32	47.63	47.94	48.06	48.09	48.11	48.13	48.13	48.13
64	**45.47**	**46.84**	47.98	48.09	48.12	48.13	48.13	48.13	48.13	48.13

(d) "Woman" JPEG

This paper has highlighted the need for a fundamental change in our understanding of image quality assessment, both subjective and objective. The results of our subjective tests are expected to be applicable in the development of image fidelity measures that predict image quality over multiple resolutions and viewing conditions, and take into account both the visibility of the compression artifacts and the image size, i.e., the visibility of the signal itself. Such measures will be invaluable for scalable image and video compression applications.

ACKNOWLEDGMENT

The authors would like to thank all subjects for their participation in the experiments. They also want to thank Prof. David Neuhoff for several discussions.

REFERENCES

[1] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed.   Academic Press, 2005, pp. 939–959.

[2] J. H. D. M. Westerink and J. A. J. Roufs, "Subjective image quality as a function of viewing distance, resolution, and picture size," *SMPTE Journal*, vol. 98, pp. 113–119, Feb. 1989.

[3] P. G. J. Barten, "The SQRI method: A new method for the evaluation of visible resolution on a display," in *Proc. Society for Information Display*, vol. 28, 1987, pp. 253–262.

[4] ——, "The effects of picture size and definition on perceived image quality," in *IEEE Trans. Electron Devices*, vol. 36, 9, Sept. 1989, pp. 1865–1869.

[5] ——, "Subjective image quality of high-definition television images," in *Proc. Society for Information Display*, vol. 31, 1990, pp. 239–243.

[6] C. Kuhmünch, G. Kühne, C. Schremmer, and T. Haenselmann, "A video-scaling algorithm based on human perception for spatio-temporal stimuli," in *Multimedia Computing and Networking*, W. chi Feng and M. G. Kienzle, Eds., Proc. SPIE, Vol. 4312, San Jose, CA, Jan. 2001, pp. 13–24.

[7] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An objective video quality assessment system based on human perception,"

TABLE V

MPSNR VALUES (dB) OBTAINED BY THE SAFRANEK-JOHNSTON METRIC OVER DIFFERENT CODING RATES AND SPATIAL RESOLUTIONS FOR JPEG 2000 CODED IMAGES (JUST-NOTICEABLE LEVELS IN SHADED CELLS, MOST PREFERRED RESOLUTIONS IN BOLD NUMBERS)

Resolution/bpp	0.05	0.1	0.12	0.15	0.18	0.2	0.3	0.5	0.6	0.75	1.0
512	32.72	36.44	37.13	38.46	39.15	40.41	42.28	44.84	45.54	**46.60**	**47.20**
384	33.94	37.79	38.51	39.97	40.60	42.03	43.45	**45.72**	**46.32**	**47.20**	47.56
256	36.73	40.71	41.12	42.91	43.44	44.61	45.41	47.01	47.43	47.84	47.94
192	38.40	42.37	42.46	44.55	**44.95**	**45.74**	**46.29**	47.48	47.81	48.03	48.05
128	39.98	43.97	43.96	**45.91**	46.20	46.62	46.88	47.79	48.01	48.09	48.09
96	41.42	45.37	**45.40**	47.07	47.12	47.29	47.35	47.99	48.12	48.12	48.12
64	**44.55**	**47.00**	46.93	47.78	47.79	47.87	47.87	48.11	48.13	48.13	48.13

(a) "Bank" JPEG 2000

Resolution/bpp	0.03	0.04	0.05	0.08	0.1	0.2	0.3	0.4	0.5	0.75	1.0
512	31.85	33.19	34.05	36.14	37.68	40.61	42.13	43.77	44.59	**46.17**	**46.85**
384	32.71	34.38	35.13	37.46	39.14	42.11	43.54	45.13	**45.61**	47.01	47.33
256	35.47	37.68	38.24	40.39	42.25	**44.55**	**45.85**	**46.83**	46.99	47.80	47.88
192	37.20	39.73	40.06		**44.02**	45.60	**46.78**	47.40	47.47	48.01	48.02
128	38.80	41.76	42.20	43.59	45.57	46.55	47.50	47.76	47.78	48.09	48.09
96	40.87	**43.98**	**44.56**	**45.07**	46.86	47.17	47.91	47.97	47.97	48.12	48.12
64	**44.17**	46.49	46.75	46.86	47.83	47.87	48.10	48.11	48.11	48.13	48.13

(b) "Bike" JPEG 2000

Resolution/bpp	0.03	0.05	0.08	0.1	0.2	0.3	0.35	0.4	0.5	1.0
512	32.62	35.53	37.96	39.76	43.52	45.02	**45.76**	**45.76**	**46.79**	**47.80**
384	33.17	36.62	39.07	41.08	44.61	**45.77**	46.50	46.75	47.29	47.96
256	35.91	39.65	41.60	43.81	**46.50**	47.02	47.54	47.68	47.86	48.11
192	37.67	41.46	42.92	**45.26**	47.26	47.46	47.89	47.97	48.02	48.13
128	39.47	43.11	**44.09**	46.38	47.72	47.75	48.05	48.08	48.09	48.13
96	41.17	**44.94**	45.24	47.27	47.98	47.98	48.12	48.12	48.12	48.13
64	**44.54**	46.85	46.89	47.94	48.12	48.12	48.13	48.13	48.13	48.13

(c) "Lena" JPEG 2000

Resolution/bpp	0.04	0.06	0.08	0.1	0.2	0.3	0.5	1.0
512	39.63	41.71	42.85	43.83	45.66	**46.47**	**47.24**	**47.90**
384	40.56	42.53	43.72	44.66	**46.39**	47.09	47.63	48.01
256	42.31	44.36	45.35	**46.30**	47.47	47.82	48.03	48.11
192	43.20	45.25	**46.08**	47.00	47.83	48.01	48.11	48.12
128	44.04	**46.12**	46.71	47.56	48.04	48.10	48.13	48.13
96	**45.07**	47.06	47.26	47.92	48.12	48.13	48.13	48.13
64	46.84	47.86	47.90	48.10	48.13	48.13	48.13	48.13

(d) "Woman" JPEG 2000

in *Human Vision, Visual Processing, and Digital Display IV*, Proc. SPIE, Vol. 1913, San Jose, CA, Feb. 1993, pp. 15–26.

[8] R. Feghali, D. Wang, F. Speranza, and A. Vincent, "Quality metric for video sequences with temporal scalability," in *Proc. IEEE Int. Conf. Image Processing (ICIP-05)*, vol. III, Genova, Italy, Sept. 2005, pp. 137–140.

[9] H. Koumaras, A. Kourtis, and D. Martakos, "Evaluation of video quality based on objectively estimated metric," *Journal of Communications and Networks*, vol. 7, no. 3, pp. 235–242, Sept. 2005.

[10] H. Koumaras, A. Kourtis, D. Martakos, and J. Lauterjung, "Quantified PQoS assessment based on fast estimation of the spatial and temporal activity level," *Multimed Tools Appl.*, vol. 34, pp. 355–374, 2007.

[11] D. Schilling and P. C. Cosman, "Image quality evaluation based on recognition times for fast image browsing applications," *IEEE Trans. Multimedia*, vol. 4, no. 3, pp. 320–331, Sept. 2002.

[12] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, pp. 177–200, 1998.

[13] 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. [Online]. Available: http://www.uta.edu/faculty/zhouwang/publications/ssim.html

[14] 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. ICASSP-89*, vol. 3, Glasgow, Scotland, May 1989, pp. 1945–1948.

[15] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Human Vision, Visual Proc., and Digital Display IV*, J. P. Allebach and B. E. Rogowitz, Eds., Proc. SPIE, Vol. 1913, San Jose, CA, Feb. 1993, pp. 202–216.

[16] 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.

[17] T. N. Pappas, T. A. Michel, and R. O. Hinds, "Supra-threshold perceptual image coding," in *Proc. Int. Conf. Image Processing (ICIP-96)*, vol. I, Lausanne, Switzerland, Sept. 1996, pp. 237–240.

[18] J. Chen and T. N. Pappas, "Perceptual coders and perceptual metrics," in *Human Vision and Electronic Imaging VI*, B. E. Rogowitz and T. N. Pappas, Eds., Proc. SPIE Vol. 4299, San Jose, CA, Jan. 2001, pp. 150–162.

[19] S. S. Hemami and M. G. Ramos, "Wavelet coefficient quantization to produce equivalent visual distortion in complex stimuli," in *Human Vision and Electronic Imaging V*, B. E. Rogowitz and T. N. Pappas, Eds., Proc. SPIE Vol. 3959, San Jose, CA, Jan. 2000, pp. 200–210.

[20] M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis," *J. Opt. Soc. Am. A*, vol. 18, no. 10, pp. 2385–2397, Oct. 2001.

[21] D. M. Chandler and S. S. Hemami, "Additivity models for suprathreshold distortion in quantized wavelet-coded images," in *Human Vision and Electronic Imaging VII*, B. E. Rogowitz and T. N. Pappas, Eds., Proc. SPIE Vol. 4662, San Jose, CA, Jan. 2002, pp. 105–118.

[22] ——, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *J. Opt. Soc. Am. A*, vol. 20, no. 7, pp. 1164–1180, July 2003.

[23] Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Structural approaches to image quality assessment," in *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed. Academic Press, 2005, pp. 961–974.

[24] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, Philadelphia, PA, 2005, pp. 573–576.

[25] A. C. Brooks and T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions," in *Human Vision and Electronic Imaging XI*, ser. Proc. SPIE, B. E. Rogowitz, T. N. Pappas, and S. J. Daly, Eds., vol. 6057, San Jose, CA, Jan. 2006, pp. 6057OU–1–6057OU–12.

[26] A. C. Brooks, X. Zhao, and T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1261–1273, Aug. 2008.

[27] G. Berbecel, *Digital Image Display: Algorithms and Implementation*. John Wiley & Sons, 2003.

[28] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 4, pp. 525–536, July 1974.

[29] S. Daly, "Subroutine for the generation of a two dimensional human visual contrast sensitivity function," Eastman Kodak, Rochester, NY, Technical Report 233203Y, 1987.

[30] T. N. Pappas, J. P. Allebach, and D. L. Neuhoff, "Model-based digital halftoning," *IEEE Signal Processing Mag.*, vol. 20, no. 4, pp. 14–27, July 2003.

[31] A. J. Ahumada, Jr. and H. A. Peterson, "An visual detection model for DCT coefficient quantization," in *AIAA Computing in Aerospace 9: A Collection of Technical Papers*, San Diego, CA, Oct. 1993, pp. 314–317.

[32] J. Sullivan, L. Ray, and R. Miller, "Design of minimum visual modulation halftone patterns," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 1, pp. 33–38, Jan./Feb. 1991.

[33] T. N. Pappas and D. L. Neuhoff, "Least-squares model-based halftoning," in *Human Vision, Visual Proc., and Digital Display III*, B. E. Rogowitz, Ed., Proc. SPIE, Vol. 1666, San Jose, CA, Feb. 1992, pp. 165–176.

[34] ——, "Least-squares model-based halftoning," *IEEE Trans. Image Processing*, vol. 8, no. 8, pp. 1102–1116, Aug. 1999.

[35] M. Analoui and J. P. Allebach, "Model based halftoning using direct binary search," in *Human Vision, Visual Proc., and Digital Display III*, B. E. Rogowitz, Ed., Proc. SPIE, Vol. 1666, San Jose, CA, Feb. 1992, pp. 96–108.

[36] *Information Technology-JPEG-Digital Compression and Coding of Continuous-Tone Still Image-Part I: Requirements and Guidelines*, ISO/IEC 10918-1 and ITU-T Recommendation T.81, 1994.

[37] J. L. M. William B. Pennebaker, Ed., *JPEG: Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.

[38] *Information Technology-JPEG 2000-Image Coding System-Part I:Core Coding System*, ISO/IEC 15 444-1, 2000.

[39] D. S. Taubman and M. W. Marcellin, "JPEG2000: Standard for interactive imaging," *Proc. IEEE*, vol. 90, no. 8, pp. 1336–1357, Aug. 2002.

[40] K. R. Murphy and C. O. Davidshofer, *Psychological Testing: Principles and Applications*, 6th ed. Upper Saddle River, NJ: Prentice Hall, 2005.

[41] M. J. Allen and W. M. Yen, *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole, 1979.

[42] L. R. Aiken, *Psychological Testing and Assessment*, 9th ed. Needham Heights, MA: Allyn & Bacon, 1997.

[43] S. H. Schwartz, *Visual Perception*, 2nd ed. Stamford, CT: Appleton&Lange, 1999.

[44] D. L. Neuhoff and T. N. Pappas, "Perceptual coding of images for halftone display," *IEEE Trans. Image Processing*, vol. 3, no. 7, pp. 341–354, July 1994.

[45] H. Levkowitz and G. T. Herman, "Color scales for image data," *IEEE Computer Graphics and Applications*, vol. 12, no. 1, pp. 72–80, Jan. 1992.

Soo Hyun Bae (S'04) received the B.S. degree in electronics engineering from Yonsei University, Korea, in 1999, and the M.S. degree in electrical and computer engineering from Georgia Institute of Technology, Atlanta, GA, in 2005. He is currently pursuing the Ph.D. degree in the School of Electrical and Computer Engineering at the Georgia Institute of Technology.

He spent the summer of 2005 as a visiting research assistant in the Image and Video Processing Laboratory of Northwestern University and the summer of 2007 as a research associate at NTT DoCoMo Labs USA. His research interests include data compression, information retrieval, perceptual model for image/video processing.

Thrasyvoulos N. Pappas (M'87, SM'95, F'06) received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1979, 1982, and 1987, respectively.

From 1987 until 1999, he was a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In 1999, he joined the Department of Electrical and Computer Engineering (now EECS), Northwestern University, as an associate professor. His research interests are in image and video quality and compression, perceptual models for image processing, image and video analysis, model-based halftoning, and multimedia signal processing.

Dr. Pappas has served as an elected member of the Board of Governors of the Signal Processing Society of IEEE (2004-2007), chair of the IEEE Image and Multidimensional Signal Processing Technical Committee, associate editor of the IEEE Transactions on Image Processing, and technical program co-chair of ICIP-01 and the Symposium on Information Processing in Sensor Networks (IPSN-04). He will also be technical program co-chair of ICIP-09. Dr. Pappas is a Fellow of SPIE. Since 1997 he has been co-chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging. He has also served as co-chair of the 2005 SPIE/IS&T Electronic Imaging Symposium.

Biing-Hwang Juang (S'79-M'81-SM'87-F'92) received his Ph.D. from University of California, Santa Barbara. He had worked at Speech Communications Research Laboratory (SCRL) and Signal Technology, Inc. (STI) on a number of Government-sponsored research projects. Notable accomplishments during the period include development of vector quantization for voice applications, voice coders at extremely low bit rates, 800 bps and around 300 bps, and robust vocoders for use in satellite communications. He was also a co-Principal Investigator for the project on co-channel separation of speech signals sponsored by the US Government. He subsequently joined Bell Laboratories in 1982, working in the area of speech enhancement, coding and recognition. Prof. Juang later became Director of Acoustics and Speech Research at Bell Labs, and at the turn of the century, Director of Multimedia Technologies Research at Avaya Labs (a spin-off of Bell Labs). His group continued the long heritage of Bell Labs in speech communication research. Prof. Juang has published extensively, including the book Fundamentals of Speech Recognition, co-authored with L.R. Rabiner, and holds about twenty patents. He has served as Editor-in-Chief for the IEEE Transactions on Speech and Audio Processing, and a number of positions in the IEEE Signal Processing Society, including Chair of its Fellow Evaluation Committee. Prof. Juang has received a number of technical awards, notable among which are several Best Paper awards in the area of speech communications and processing, the Technical Achievement Award from the Signal Processing Society of the IEEE, and the IEEE Third Millennium Medal. He is a Fellow of the IEEE, a Fellow of Bell Laboratories, a member of the US National Academy of Engineering and an Academician of Academia Sinica. Prof. Juang joined Georgia Tech in 2002 holding the Motorola Foundation Chair Professorship and is an Eminent Scholar of Georgia Research Alliance.