

A Haar Wavelet Approach to Compressed Image Quality Measurement

Yung-Kai Lai

Welltel Networks, Irvine, California 92618

and

C.-C. Jay Kuo

*Department of Electrical Engineering—Systems, University of Southern California,
Los Angeles, California 90089-2564*

Received August 22, 1997; accepted August 24, 1999

The traditional mean-squared-error and peak-signal-to-noise-ratio error measures are mainly focused on the pixel-by-pixel difference between the original and compressed images. Such metrics are improper for subjective quality assessment, since human perception is very sensitive to specific correlations between adjacent pixels. In this work, we explore the Haar wavelet to model the space–frequency localization property of human visual system (HVS) responses. It is shown that the physical contrast in different resolutions can be easily represented in terms of wavelet coefficients. By analyzing and modeling several visual mechanisms of the HVS with the Haar transform, we develop a new subjective fidelity measure which is more consistent with human observation experience. © 2000 Academic Press

Key Words: image fidelity assessment; compression artifact measure; human visual system (HVS); Haar transform; wavelet transform.

1. INTRODUCTION

The objective of lossy image compression is to store image data efficiently by reducing the redundancy of image content and discarding unimportant information while keeping the quality of the image acceptable. Thus, the tradeoff in lossy image compression is between the number of bits required to represent an image and the quality of the compressed image. This is usually known as the rate–distortion tradeoff. The number of bits used to record the compressed image can be measured easily and objectively. However, the “closeness” between the compressed and the original images is not a purely objective measure, since human perception plays an important role in determining the fidelity of the compressed image.

At present, the most widely used objective distortion measures are the mean squared error (MSE) and the related peak signal–noise ratio (PSNR). They can easily be computed to represent the deviation of the distorted image from the original image in the pixelwise sense. However, in practical viewing situations, human beings are usually not concentrated on pixel differences alone, except for particular applications such as medical imaging, where pixelwise precision can be very important. The subjective perceptual quality includes surface smoothness, edge sharpness and continuity, proper background noise level, and so on. Image compression techniques induce various types of visual artifacts that affect the human viewing experience in many distinct ways, even if the MSE or PSNR level is adjusted to be about equal. It is generally agreed that MSE or PSNR does not correlate well with the visual quality perceived by human beings, since MSE is computed by adding the squared differences of individual pixels without considering the spatial interaction among adjacent pixels. Some work tries to modify existing quantitative measures to accommodate the factor of human visual perception. One approach is to improve MSE by putting different weights to neighboring regions with different distances to the focal pixel [31]. Most approaches can be viewed as curve-fitting methods to comply with the rating scale method.

In order to obtain an objective measure for perceived image fidelity, models of the human visual system (HVS) should be taken into account. It is well known that the HVS has different sensitivities to signals of different frequencies. Since the detection mechanisms of the HVS have localized responses in both the space and frequency domains, neither the space-based MSE nor the global Fourier analysis provides a good tool for the modeling. Since the late 1970's, researchers have started to pay attention to the importance of the human visual system (HVS) and tried to include the HVS model in image fidelity or quality metrics [20, 21]. The development of the HVS model at that time was not mature enough and the proposed model could not interpret human visual perceptual phenomena very well. Recently, Karunasekera proposed an objective distortion measure to evaluate the blocking artifact of block-based compression techniques [25]. Watson [40] and van den Branden Lambrecht [27, 28] proposed more complete models and extended their use to compressed video.

The major difficulty in modeling HVS, however often neglected by subjective fidelity assessment research, is in the computation of the contrast in complex images. In this work, we explore the Haar wavelet, which has good space–frequency localization properties to evaluate the physical contrast. It is shown that the contrast can be easily represented in an expression of transform coefficients. Some visual phenomena can also be modeled by multiresolutional analysis since the contrast is defined in every resolution. Contrasts in different resolutions are then combined with models of visual mechanism to yield a new gray-scale image fidelity measure. The new objective error measure is defined as the aggregate contrast response mismatch between the original and compressed images. The proposed metric is more consistent with human subjective ranking and capable of describing various compressed image artifacts. Our effort can have an impact on the development of new compression methods that concentrate more on the overall perceptual fidelity rather than pixelwise error minimization.

This paper is organized as follows. The HVS model is first discussed in Section 2. We propose a new definition of contrast with respect to complex images by taking the Haar wavelet transform of the image, and using wavelet coefficients to estimate the local contrast at each resolution in the image in Section 3. By using the multiresolution and space–frequency

localization properties of wavelets, several observed inconsistencies in the psychophysical literature can be explained naturally. HVS mechanisms mentioned in Section 2 such as the suprathreshold perception response, the frequency masking effect, and the directional preference can be conveniently incorporated in the wavelet framework. Based on this framework, we propose an objective metric to compute the extent of the perceived contrast in every resolution and derive an error measure by examining the response differences of contrasts between the original and compressed images at each resolution in Section 4.1. Psychophysical experiments are conducted in Section 4.2 to demonstrate the validity and effectiveness of the Haar filter. We use these experiments to conclude the independence of two visual masking variables. In addition, human perception deficiency at oblique orientations is measured in these experiments. In Section 5, the effectiveness of the proposed image fidelity measure is tested with natural images. The effect of viewing distance effect on the perception of image compression artifacts is also discussed. Concluding remarks are given in Section 6.

2. HUMAN VISUAL SYSTEM (HVS) MODELS

Visual perception is the result from a series of optical and neural transformations. The light is projected onto the retina through the cornea and lens to form an optical image. This retina image is then sensed by photoreceptors on the retina and transformed into neural responses to reach the optic nerve. The optic nerve carries these signals to the visual cortex in the brain for further processing. Since both photoreceptors and cortical cells transform incoming signals into some particular representation, they form the core of the visual system. Many of the composing mechanisms are originating from these two core elements as well as the optical mechanism of the eye. The objective of psychophysical research is to model the overall transfer function of the visual system.

2.1. Contrast Threshold and Sensitivity

Generally speaking, human visual perception is a function of both the luminance difference between the background and the stimuli and the background adaptation level.

Let L_{\max} and L_{\min} be the maximum and minimum luminance of the waveform around the point of interest. Michelson's contrast, defined as

$$C = (L_{\max} - L_{\min}) / (L_{\max} + L_{\min}), \quad (1)$$

is found to be nearly constant when used to represent the just noticeable luminance difference [22]. For stimuli of uniform luminance seen against a uniform background, another contrast measure called Weber's fraction is defined as

$$C = \Delta L / L, \quad (2)$$

where ΔL is the luminance difference and L is the background luminance. We see that for simple patterns, Weber's fraction and Michelson's contrast differ by a factor of 2.

Physiological experiments showed that many of the cortical cells are focused on certain regions in their receptive fields and only sensitive to the contrast in certain frequency bands. The overall visual perception of object luminance or contrast is therefore the aggregate performance of each cell's frequency response [13]. Since the HVS cannot provide an infinite

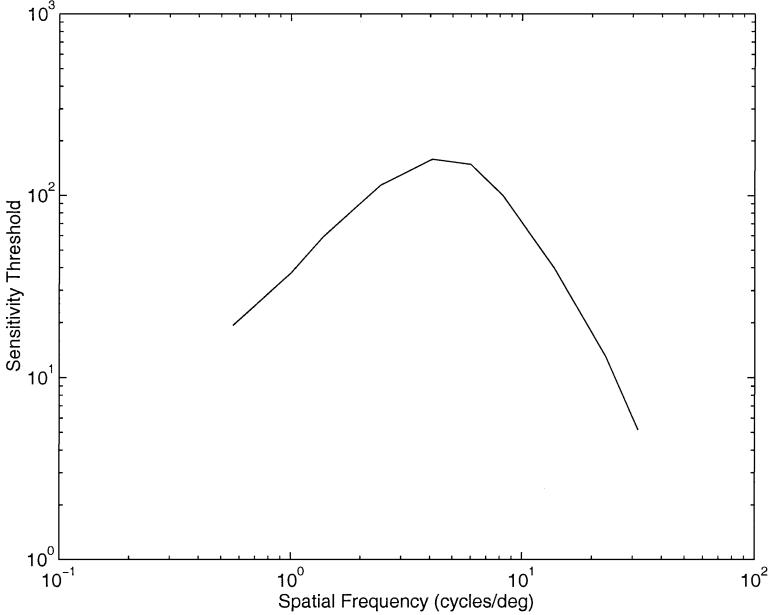


FIG. 1. A typical contrast sensitivity curve for human beings.

contrast resolution, a contrast threshold exists at every spatial frequency. This threshold represents the just noticeable contrast in each frequency band.

The contrast threshold value is a function of the spatial frequency, which can be determined experimentally. A typical contrast sensitivity curve defined as the reciprocal of the contrast curve is shown in Fig. 1 [13]. As shown in the figure, HVS has the highest luminance sensitivity around 3–10 cycles per degree, and the sensitivity attenuates at both high and low frequency ends. It was shown by Campbell *et al.* [6] that, over a wide range of spatial frequencies, the contrast threshold of a grating is determined only by the amplitude of its fundamental Fourier components. Based on this observation, Fourier frequency analysis is widely used in vision research. According to this model, an image artifact can be sensed only if its contrast is above the visual threshold after probability summation and contrast integration.

There has been work devoted to parameterize the contrast sensitivity curve. Daly [9] and Barten [1, 2] summarized various experimental results from the literature and determined the contrast sensitivity as the function of several variables. Two variables are of particular importance, i.e., the display size and the background luminance level. Experimental results [8] suggest that the angular display size of gratings affects the contrast sensitivity at low spatial frequencies. This deviation may be due to the consequence that a smaller number of stimuli cycles produces smaller perceived contrast at the threshold [7]. In addition, Peli [33] showed that the suprathreshold perception is unaffected with variable grating sizes up to 4° . Thus, we do not consider the display size as a variable in this work. On the other hand, although the contrast threshold is approximately constant for various background luminance levels for extremely low frequency patterns such as the staircase grating, this constancy does not hold for the threshold sensitivity curve at a wide range of spatial frequencies. By taking this factor into consideration and using a parabola in log–log coordinates as a reasonable approximation [41], we obtain the relationship between the contrast threshold C_{T0} and the

spatial frequency f as

$$\log(1/C_{T0}) = (p_1 \log(L) + q_1)(\log f)^2 + (p_2 \log(L) + q_2) \log f + (p_3 \log(L) + q_3), \quad (3)$$

where L represents the background luminance around the fixation point and p_i and q_i are model parameters. The computation of L from different resolutions will be discussed in Section 3.2.

2.2. Channel Interactions

Although cells narrowly tune to different frequency bands, they are not strictly band-limited and interactions among adjacent frequency channels are well observed. Two primary effects were often discussed in the literature, i.e., the summation effect and the masking effect. The summation effect is an interchannel effect saying that the neighboring frequency channels contribute to the total contrast. Consequently, a subthreshold contrast may still produce a small response if there exist other excitory stimuli in nearby frequency channels. However, since the summation effect is far less important than the masking effect [35], it is not considered in this work. The masking effect is another interchannel effect which states that the visibility of a stimulus at some frequency could be impaired by the presence of other stimuli in nearby frequency channels. One well-known example is the blocking artifact in images compressed by block transform methods. Since the blocking artifact consists of high-frequency edge components, it is less visible in textured regions. This effect can be viewed as a reduction of contrast sensitivity threshold at certain spatial frequencies in certain regions.

Several masking models were investigated by Klein *et al.* in [26]. Generally speaking, there is no single model which can be used to account for all masking phenomena. For the multiple spatial channel HVS model, at least two variables are required to model the masking effect. They are the frequency separation and the masker contrast. The sensitivity threshold is lowered when the frequency separation between the masker and the signal decreases and/or when the masker contrast is larger with respect to the signal contrast [12, 30]. If we assume these two variables are separable, the new contrast threshold C_T after masking becomes

$$C_T = C_{T0} \prod_{i=1}^N g_i(C_i, C_0) h_i(f_i, f_0), \quad (4)$$

where

$$g_i(C_i, C_0) = \begin{cases} 1 & \text{if } C_i \leq k_0 C_0, \\ k_1 (C_i / C_0)^{k_2} & \text{if } C_i > k_0 C_0 \end{cases} \quad (5)$$

and

$$h_i(f_i, f_0) = \begin{cases} k_3 (f_i / f_0)^{k_4} & \text{if } f_i < f_0, \\ k_5 (f_i / f_0)^{k_6} & \text{if } f_i > f_0, \end{cases} \quad (6)$$

and where C_{T0} is the original contrast threshold, C_0 is the contrast of the signal, C_i is the contrast of the masker in the i th channel, and f_0 and f_i are spatial frequencies of the signal

channel and the i th masking channel, respectively. Both the separability assumption (4) and parameters k_0 through k_5 in (5) and (6) will be examined and determined in Section 4.2.B.

2.3. Suprathreshold Contrast

Since the subthreshold stimuli, i.e., stimuli with a contrast lower than the threshold, cannot be sensed by human beings, only the suprathreshold contrast is of concern in image perception. As the suprathreshold contrast becomes larger, the equal-response curve morphs from the inverse-U shape near the threshold (Fig. 1) to a flat horizontal line at high contrast levels. In other words, at high suprathreshold contrast levels, the visual responses to all spatial frequencies (below the optical cut-off frequency) become approximately the same [19].

Since visual responses to the suprathreshold contrast involves human subjective rating, it is difficult to use psychophysical methods to measure it. However, even though it is difficult to find a precise formula for its modeling, it is generally agreed that the estimated response R is a function of the spatial frequency and follows a power law [29],

$$R = k(C - C_T)^p, \quad (7)$$

where C is the suprathreshold contrast, C_T is the contrast threshold at the specific frequency, and the exponent p varies between 0.4 and 0.52 [7]. The value of p is chosen to be 0.45 and the scaling or normalization factor k is set to 1 in this work.

2.4. Directional Preference

Besides spatial locations and frequencies, HVS also responds differently to various orientations of stimuli. Campbell *et al.* [4, 5] demonstrated psychophysically that HVS is most sensitive to stimuli in the vertical and horizontal directions and least sensitive to stimuli in the 45° and 135° directions. Different from spatial frequency selectivity, the bandwidth of orientational selectivity varies considerably from cell to cell. Cortical cells from foveal and near nonfoveal cortical regions have a wide range of orientation bandwidth from 10° to over 180° , with the median bandwidth being about 45° [13]. Phillips and Wilson [35] used masking experiments to show that orientation bandwidths vary from about $\pm 30^\circ$ at 0.5 cpd to $\pm 15^\circ$ at 11.3 cpd. These results encourage the use of filters in four different orientations: 0° , 45° , 90° , and 135° . In our work, the contrast sensitivity function in the oblique directions also takes a form similar to that of (3) but with different parameters (see Section 4.3).

3. NEW CONTRAST DEFINITION BASED ON HAAR WAVELET

3.1. Space-Frequency Localization

Limited by its spatial location on the retina, each photoreceptor can only focus on a certain region of the visual field to form a channel. Furthermore, each photoreceptor is only sensitive to signals of a certain spatial frequency range. Cortical cells pool the responses of all photoreceptors on the same retinal location, and there are many channels tuned to the same spatial frequency band, say, 2 cycles/degree, but at different locations in the receptive field. Thus, we can say that the frequency response of visual stimuli is not only band-limited in the frequency domain but also localized in the space domain. For example, the frequency

response of the fixation point is characterized by the typical contrast threshold function as shown in Fig. 1. However, from the spatial inhomogeneity [16] phenomenon, the high-frequency response will further attenuate as the eccentricity from the focal point increases. This means that responses at the same (high) spatial frequency in different locations in the receptive field are governed by different channels, thus providing an evidence that visual channels are narrowly tuned in specific spatial locations.

The commonly used Fourier frequency analysis is, however, a global process which gives all spatial components the same weighting. It is well known as Heisenberg's uncertainty principle that exact localization in both space and frequency domains cannot be achieved simultaneously [24]. The Gabor transform [15], which is a Gaussian-windowed Fourier transform, was proved to achieve the limit of the Heisenberg inequality. Gabor gratings have thus been widely used in modern psychophysical experiments. Parameters of the Gaussian window were chosen based on researchers' preferences, and various degrees of localization were achieved [34]. That is, by varying the Gaussian envelope parameters, the passbands of Gabor gratings were overlapped to a different extent. There are some limitations in the Gabor representation. First, it is difficult to analyze the stimuli whose frequency responses fall into the overlapped band. Second, since the Gabor filter is an IIR (infinite impulse response) filter, truncation is still needed for practical implementation. However, localization is not fully ensured after truncation.

3.2. Contrast Computation with Haar Wavelets

A major difficulty of HVS modeling, though one often neglected by researchers, is the computation of the contrast in complex images. Michelson's contrast defined in (1) is based on the staircase pattern, which has distinct L_{\max} and L_{\min} . In psychophysical experiments with sinusoidal gratings, the stimuli also have unique peak (maximum) and trough (minimum) luminance. In Gabor experiments, on the other hand, the contrast is defined at the largest ripple, which is located at the focal point. It is, however, extremely difficult to define the contrast for natural images since there are no unique or obvious maximum and minimum luminance values to be recorded even with the Gaussian envelope localization. For example, for a one-dimensional grating composed of two sinusoidal waveforms with different frequencies [32],

$$f(x, y) = I_0(1 + a_1 \cos(2\pi\omega_1x) + a_2 \cos(2\pi\omega_2x)), \quad (8)$$

where $\omega_1 < \omega_2$. The grating is shown in Fig. 2. We see that the contrast is approximately equal to $a_2/(1 - a_1)$ at point A, where the slow-varying waveform $a_1 \cos(2\pi\omega_1x)$ is at its minimum luminance, while the contrast is about $a_2/(1 + a_1)$ at point B, where $a_1 \cos(2\pi\omega_1x)$ is at its maximum. Therefore, the contrast of this grating is different everywhere along the x -direction. A good definition of contrast should be able to handle such cases.

Since (1) is defined as the ratio of the luminance difference and the background adaptation level, both values should be obtained if one wishes to devise a good definition of the contrast in complex images. Hess *et al.* [23] defined the contrast at the i th spatial frequency band as

$$C_i = \frac{AC_i}{DC}, \quad (9)$$

where AC_i is the filtered AC coefficient at that specific frequency band and DC is the

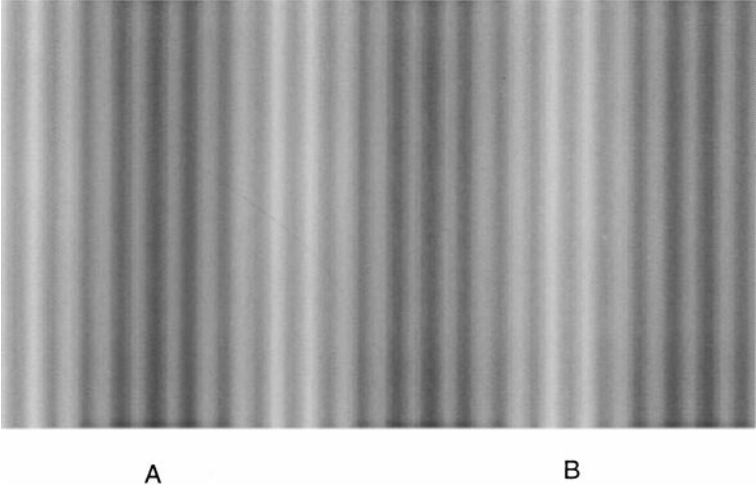


FIG. 2. A composite grating example for contrast computation.

DC (zero-frequency) value computed based on the whole image or $1/4$ or $1/16$ subimages. Clearly, this value is pre-determined and not adaptive to model the contrast at different resolutions with their respective space–frequency localizations. Peli [32] used localized cosine log filters to define the contrast at the i th spatial frequency band as

$$C_i = \frac{AC_i}{\sum_{j=0}^{i-1} AC_j}, \quad (10)$$

where the denominator takes the sum of responses from all frequencies lower than the target frequency band. It has a good adaptive property since these filters are well localized and perfectly reconstructive. Both of the above approaches suggest that the background adaptation level can be computed for the low-frequency response, since it represents more global variation. On the other hand, the high-frequency response acts more like a differential operator. This basic idea does make a lot of sense intuitively. The disadvantage of these approaches is that it is difficult to show its coherence with the contrast definition presented earlier mathematically. We will show below that the Haar wavelet transform approach provides a good framework to generalize the contrast definition from simple to complex cases both intuitively and mathematically.

The wavelet transform provides a good space–frequency localization property [10] and can be implemented using the multichannel filter banks. Compactly supported wavelets such as the Daubechies filters [10] can be implemented with FIR (finite impulse response) filters. The space–frequency localization is optimized among all possible FIR filters with the given length for the Daubechies filters. The Haar wavelet is the simplest basis function in the compactly supported wavelet family. It provides the capability to compute the contrast directly from the responses of low and high frequency subbands. For the Haar wavelet, filter coefficients for the low- and high-frequency filter banks are given by

$$h_0[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = -1, 0, \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$h_1[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = 0, \\ -\frac{1}{\sqrt{2}} & n = -1, \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

respectively. Assume that a discrete-time input signal $x[n]$ is the staircase contrast pattern

$$x[n] = \begin{cases} L_{\max} & n < 0 \\ L_{\min} & n \geq 0. \end{cases} \quad (13)$$

The responses $y_{0,1}[n]$ and $y_{1,1}[n]$ at the first resolution after filtering with $h_0[n]$ and $h_1[n]$, respectively, are

$$y_{0,1}[n] = \begin{cases} \sqrt{2}L_{\max} & n < -1 \\ \frac{1}{\sqrt{2}}(L_{\max} + L_{\min}) & n = -1 \\ \sqrt{2}L_{\min} & n > 0, \end{cases} \quad (14)$$

$$y_{1,1}[n] = \begin{cases} \frac{1}{\sqrt{2}}(L_{\max} - L_{\min}) & n = -1 \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

Thus, contrast C_1 in the interval $(-1, 0)$ and the 1st (finest) resolution can be computed via the ratio of $y_{1,1}[n]$ and $y_{0,1}[n]$, i.e.,

$$C_1 = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} = \frac{y_{1,1}[-1]}{y_{0,1}[-1]},$$

which is consistent with Michelson's contrast definition as given in (1). At the second (second finest) resolution, the low-frequency band response $y_{0,1}[n]$ is downsampled by 2 and fed into the same filter bank. The responses are

$$y_{0,2}[n] = \begin{cases} 2L_{\max} & n < -1 \\ L_{\max} + L_{\min} & n = -1 \\ 2L_{\min} & n > -1, \end{cases} \quad (16)$$

$$y_{1,2}[n] = \begin{cases} L_{\max} - L_{\min} & n = -1 \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Again, we can compute the contrast at the second resolution as

$$C_2 = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} = \frac{y_{1,2}[-1]}{y_{0,2}[-1]}.$$

Following this path, the contrast at the i th resolution can be computed as

$$C_i = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} = \frac{y_{1,i}[-1]}{y_{0,i}[-1]}, \quad (18)$$

i.e., the ratio of high- and low-band responses evaluated at $n = -1$. Figure 3 illustrates this constant-ratio relationship across resolutions.

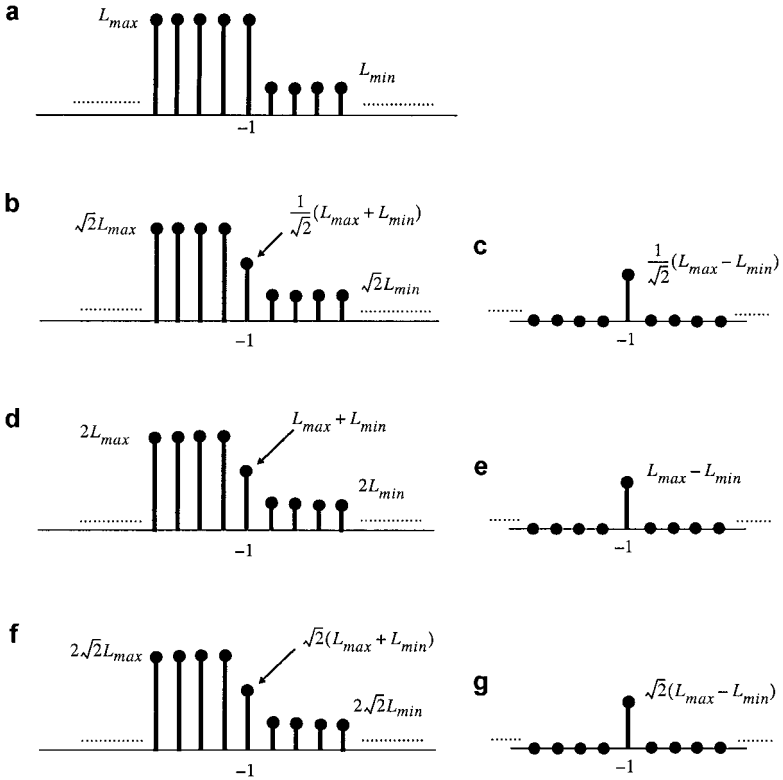


FIG. 3. Contrast computation using various filter responses where (a) is the original staircase signal, (b) and (c) are low and high band responses at the 0th resolution, and (d), (e), (f), and (g) are responses at the first and second resolutions.

It is worthwhile to point that the dyadic wavelet transform satisfies the uncertainty principle in that the supported spatial radius is doubled when the center frequency of the highpass band is halved. In addition, the supported radii of the highpass filter and the lowpass filter are exactly the same. This is a very desirable property since the background adaptation level, i.e., the mean luminance of the signal, should be obtained from the same supported radius as that of the bandpass filter extracting the frequency components to form the contrast. According to (14), (16), and subsequent computation, the background luminance level L in (3) at the i th resolution can be computed as

$$L = (\sqrt{2})^{-i} y_{0,i}[-1] \quad (19)$$

Even though the new contrast is derived based upon the staircase pattern, it can be directly applied to more complex cases such as the example in Fig. 2. Taking $\omega_1 = 0.004$, $\omega_2 = 0.0625$, and $a_1 = a_2 = 0.25$, the contrasts at points A and B in Fig. 2 should be 0.2 and 0.33, respectively. Based on the half-band decomposition, the fast varying term $a_2 \cos(2\pi\omega_2x)$ will be separated from the slowly varying term $a_1 \cos(2\pi\omega_1x)$ at the fourth resolution. We show in Fig. 4 the low- and high-frequency bank responses as well as the computed contrast. We see that the Haar wavelet can predict the contrast at different spatial locations accurately.

There are several reasons to define multiple contrasts in different resolutions. First of all, since human contrast sensitivity is highly dependent on the spatial frequency, multiple contrasts can be used to address different variations at different resolutions across the image

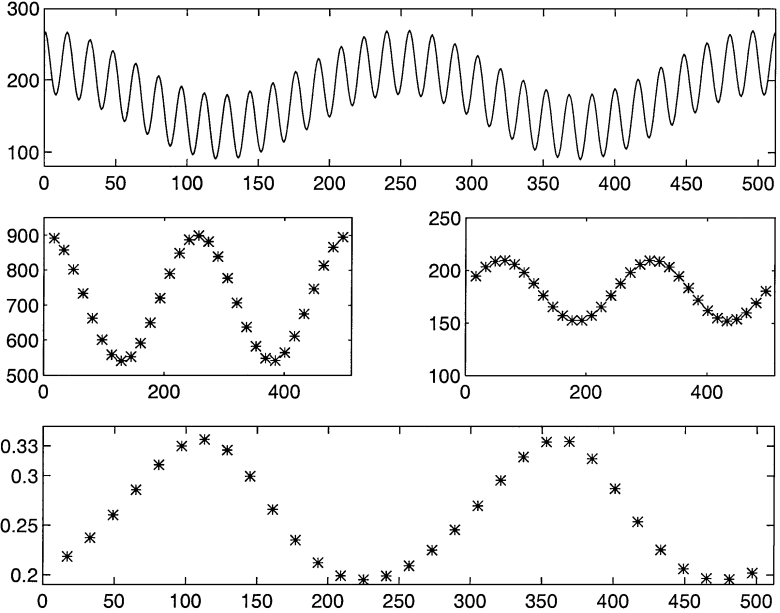


FIG. 4. The Haar decomposition of Fig. 2: the waveform (top) low-frequency response at the fourth resolution (middle- left), high-frequency response at the fourth resolution (middle right), the ratio (computed contrast) of these two responses (bottom).

[32]. Second, the uncertainty principle requires the response in different frequency bands to have different supported radii as stated above. Furthermore, it was shown that each frequency channel in HVS has the bandwidth of about one octave [32]. The dyadic wavelet transform satisfies this requirement naturally. Finally, perfect reconstruction is possible with responses obtained from different scales, and no visual information will be lost during the process. In contrast, to perfectly reconstruct the visual information using the Gabor analysis, all filters must have the same length and, as a result, the space-frequency localization property is less flexible.

4. NEW WAVELET-BASED FIDELITY MEASURE

4.1. Fidelity Measure System and Metric

Based on the discussion in Sections 2 and 3, we propose a new fidelity measure system as shown in Fig. 5 and detail the process below.

1. Wavelet Decomposition

Both the original and distorted images are passed through the system for dyadic Haar wavelet decomposition in four orientations, i.e., 0° , 45° , 90° , and 135° . The oblique

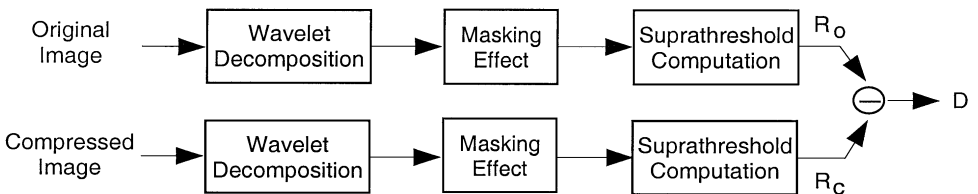


FIG. 5. A block diagram of the proposed fidelity measure system.

decomposition is performed on diagonally adjacent pixels, thus the central spatial frequency is $1/\sqrt{2}$ times of that at horizontal and vertical directions at the same decomposition level. Contrasts C are computed at every pixel and every resolution of interest with (18). Contrast thresholds C_{T0} at each resolution are computed via (3) with respect to their center frequencies f .

2. Masking Effect

The contrast threshold C_{T0} is adjusted according to (4) at each resolution to incorporate the masking effect.

3. Suprathreshold Computation

Equation (7) is used to give the suprathreshold response from computed contrasts C and adjusted contrast thresholds C_T .

4. Summation of Error Measure

Let subscripts c and o represent the compressed and the original images, respectively, and let (i, j) indicate the coordinate of the pixel in the image. Then, the perceptual error measure D for the entire image is pooled and the Minkowski metrics taken as

$$D = \left(\frac{1}{N} \sum_{k=1}^N \left(\sum_{j=1}^V \sum_{i=1}^H |R_{c,k}(i, j) - R_{o,k}(i, j)| \right)^\beta \right)^{1/\beta}, \quad (20)$$

where V and H are the vertical and horizontal sizes of the image, respectively, N is the number of filtering channels across all frequency bands in the four direction, and β is an empirical parameter related to the psychometric function and probability summation with values from 2.0 to 4.0 [3]. We choose $\beta = 4$ in the experiment.

Note that the error measure D is dimensionless since the contrast itself is dimensionless.

4.2. Experimental Calibration and Validation

The following psychophysical experiments were conducted on a 17" Silicon Graphics color graphic display GDM-17E11. The luminance range of the display was adjusted from 0 to 80 cd/m^2 (candela/square meter) using a Photoresearch spectroradiometer. There were 256 discrete gray scales present in the experiments. The relationship of the luminance L versus the gray scale G is measured and approximated by

$$L = \begin{cases} (0.0785G - 1.3270)^{1.4925} & \text{if } G \geq 28 \\ (0.0159G + 0.5437)^{10} & \text{if } G < 28. \end{cases}$$

This relation can be used to transfer the display gray levels to the actual luminance. The transfer characteristics is plotted in Fig. 6. This curve was used to compute the actual contrast in the following experiments. This display has a smaller gamma value than ordinary displays [36], but the influence on the following experiments is not critical.

A. Validation of Haar wavelet. The fact that cortical cells have a Gaussian-shaped reception profile [11] is often used to support the argument that the Gabor filter is preferable in vision experiments. Since the Haar filter does not possess the same Gaussian-shaped passband as the Gabor filter, one may suspect the validity of using the Haar filter in vision analysis. To validate the use of the Haar wavelet, we measured the contrast threshold by using both Gabor and Haar filtered patches. The spatial frequencies of test patches range

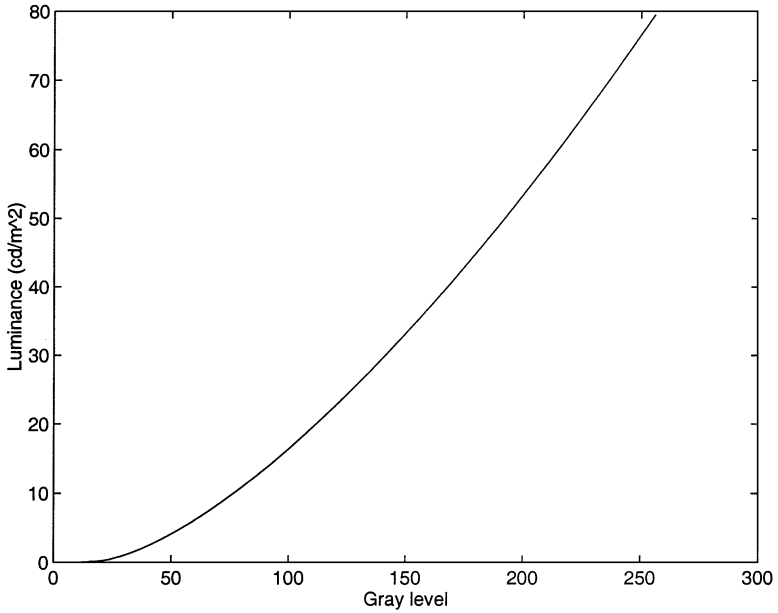


FIG. 6. The plot of the luminance versus the gray level for the color graphic display used in experiments.

from 0.069 to 19.2 cycles per degree. This range covers virtually the whole frequency band we would sense from digital images. The result is shown in Fig. 7, where the sensitivity threshold, defined as the reciprocal of the contrast threshold, is plotted as a function of spatial frequency. The closeness of these two curves confirms that the Haar filter has a comparable performance in comparison with the Gabor filter.

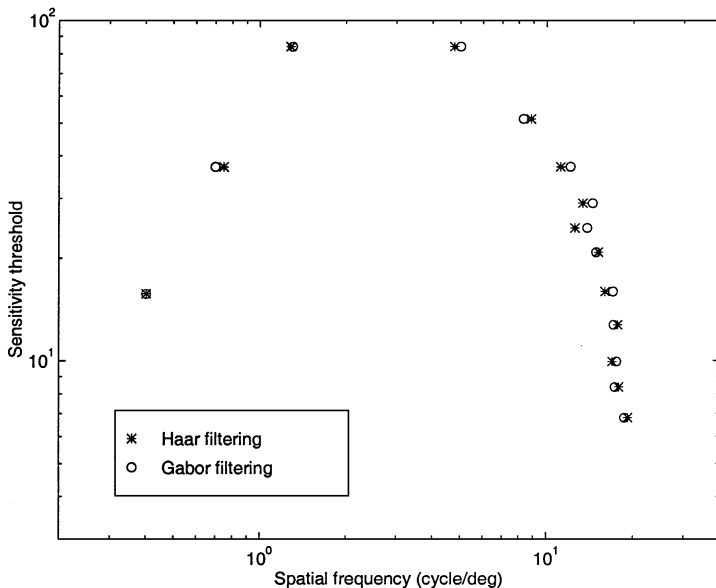


FIG. 7. Comparison of contrast sensitivity thresholds using the Gabor and Haar filters.

B. Suprathreshold masking. In Section 2.2, we formulated the masking function (4) by assuming that the two variables of masking, namely, the frequency separation and the contrast ratio between the target and masker, are independent and separable. To verify this assumption, psychophysical experiments were conducted to find the parameters of this model. The contrast ratio C_{mask}/C , where C_{mask} and C represent the contrast of masking and target signals, respectively, ranged from 0.5 to 2.5. The frequency ratio f_{mask}/f , in the meanwhile, ranged from -3 to 3 octaves. To isolate individual effects, we first fixed the frequency ratio, and varied the contrast of each signal to investigate the effect of the contrast ratio. The result is shown in Fig. 8a, and an exponential fitting function was determined from the data. We then varied both the contrast and the frequency ratios of target and masking

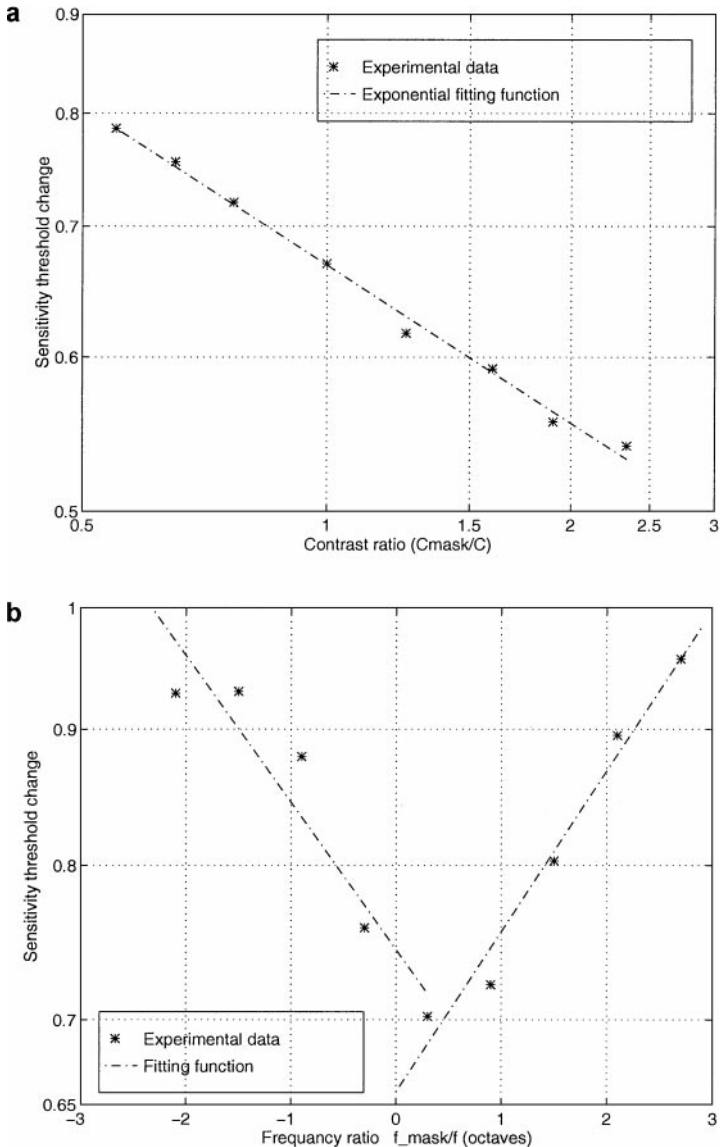


FIG. 8. Illustration of the masking effect: sensitivity threshold changes under different (a) contrasts and (b) frequency ratios.

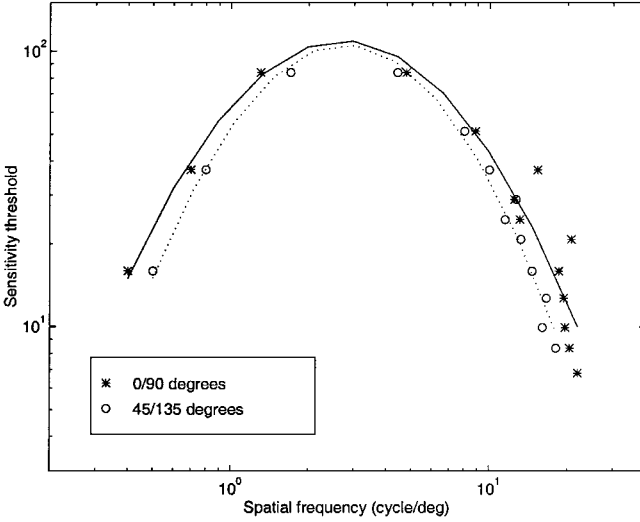


FIG. 9. Horizontal/vertical and diagonal sensitivity thresholds.

signals. During the computation process, we scaled experimental data with respect to their contrast ratios according to Fig. 8a. The scaled data show a very small amount of deviation, thus indicating that (4) is a very good approximation to the masking model. The means of experimental data and fitting functions are shown in Fig. 8b, which are in good agreement with those in [12]. By fitting the data, we obtain the following parameters in (5) and (6):

$$\begin{aligned} k_0 &= 0.22, & k_1 &= 1.5, & k_2 &= 0.27, & k_3 &= 1.34, \\ k_4 &= 0.18, & k_5 &= 1.52, & k_6 &= -0.20. \end{aligned}$$

C. Directional preference. The contrast threshold functions for four orientations (0° , 45° , 90° , and 135°) are measured. It is confirmed experimentally that there is no significant difference between the contrast thresholds of 0° (horizontal) and 90° (vertical) stimuli, nor between thresholds of 45° and 135° stimuli. The difference between thresholds of $0^\circ/90^\circ$ and $45^\circ/135^\circ$ stimuli is shown in Fig. 9, where we see that the sensitivity threshold is lower for diagonal stimuli. Parameters in (3) are obtained from the fitting functions. They are

$$\begin{aligned} p_1 &= -0.0062, & p_2 &= 0.16, & p_3 &= 0.24, \\ q_1 &= -0.53, & q_2 &= 0.52, & q_3 &= 3.28, \end{aligned}$$

for horizontal/vertical thresholds. For oblique thresholds, the same parameters are used for p_1 , p_2 , and p_3 while

$$q_1 = -0.65, \quad q_2 = 0.76, \quad q_3 = 3.06.$$

The contrast sensitivity curves for different orientations and L 's are shown in Fig. 10. The curves are consistent with Daly's [9] and Barten's [1, 2] results except at very low spatial frequencies, where the sensitivity is lower than the literature. This spatial frequency range, however, is seldom used in practical viewing situations.

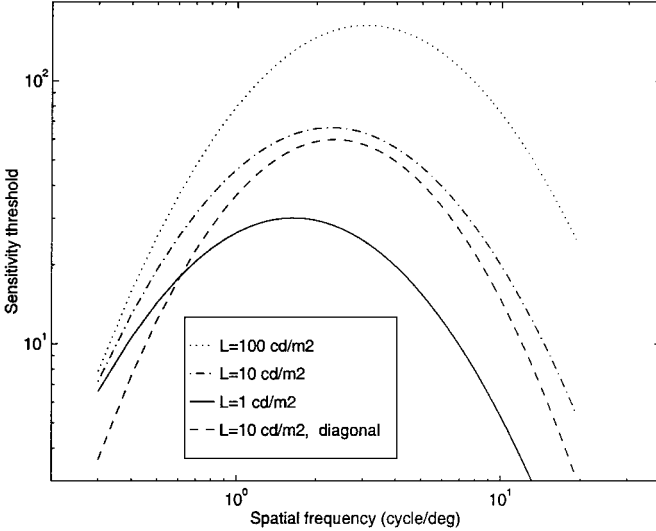


FIG. 10. Contrast sensitivity curves for horizontal/vertical and diagonal gratings.

5. APPLICATION TO COMPRESSION ARTIFACT MEASURE

5.1. Perceptual Difference Map

Compressed Lena images of size 256×256 were used for image fidelity assessment with the proposed new fidelity assessment system. Two types of compression schemes were applied: block DCT-based compression (i.e., JPEG) and wavelet-based compression. We evaluate the performance of our perceptual distortion measure by examining the perceptual error map, defined as the sum of β -weighted response differences at each pixel,

$$D_{\text{permap}}(i, j) = \frac{1}{N} \sum_{k=1}^N (|R_{c,k}(i, j) - R_{o,k}(i, j)|)^\beta, \quad (21)$$

where the variables are as defined in (20), against the pixelwise error map which is used in MSE and PSNR computation,

$$D_{\text{pxlmap}}(i, j) = (G_c(i, j) - G_o(i, j))^2, \quad (22)$$

where $G_c(i, j)$ and $G_o(i, j)$ represent the grayscale values at pixel location (i, j) of the compressed and the original images, respectively. Since these two error maps are computed by different methods and are of different magnitude, we normalize them by equalizing the energy of the two maps for fair comparison. The viewing distance in this section is set to five times the width of the image.

The JPEG compression standard is a block-based method [38]. It does not consider the correlation among adjacent blocks, and the blocking artifact usually appears at low bit rates, presented as blocky edges along block boundaries. This artifact is visually annoying, but cannot be fully represented by the pixel-difference-based PSNR measure. We used an image compressed with the default quantization table with a bit rate of 0.19 bpp and PSNR = 23.36 dB. The original image and the compressed image are shown in Fig. 11. The resulting difference maps between Fig. 11a and 11b are shown in Fig. 12. We see

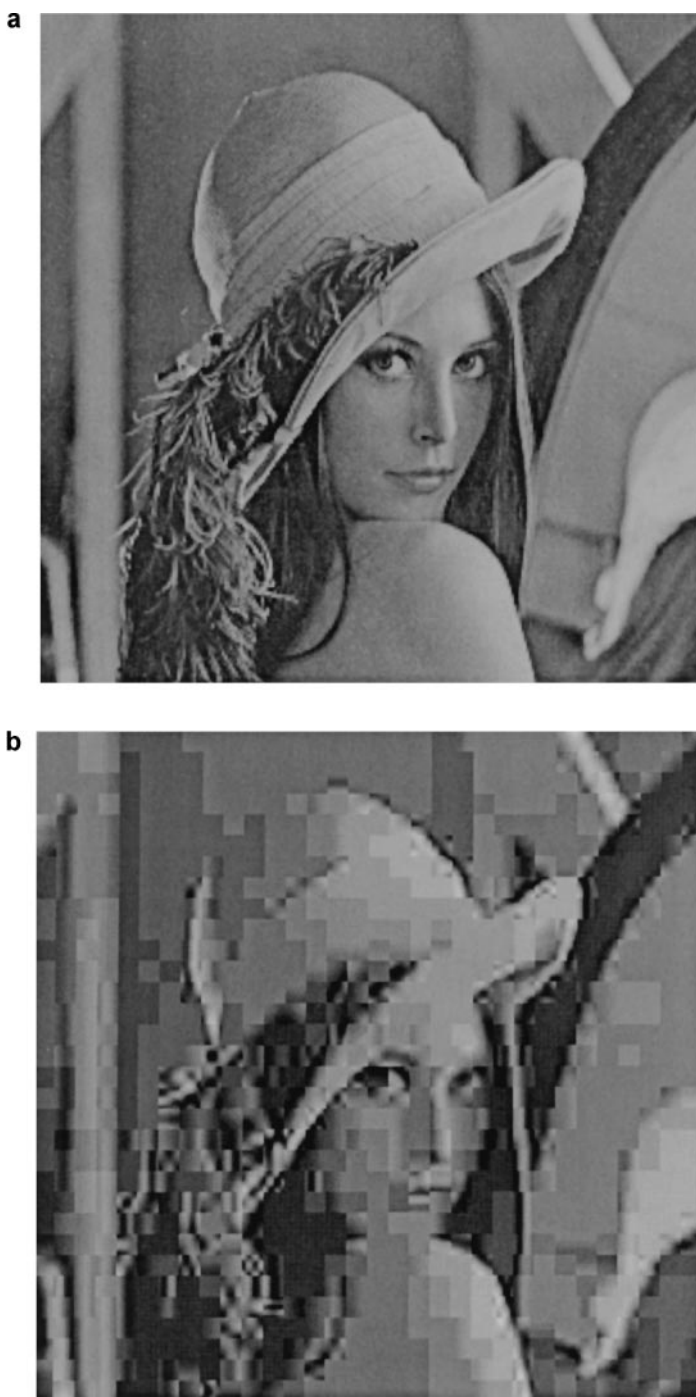


FIG. 11. (a) Original Lena image and (b) JPEG-compressed Lena at 0.19 bpp with PSNR = 23.36 dB.

that most of the energy of the pixelwise difference map concentrates in texture regions, since the pixel difference is large in these regions at low bit rates. The blocking artifact is mostly detected in homogeneous regions with slow slopes such as the shoulder, but is not detected in extremely flat regions such as the background, where the background noise is

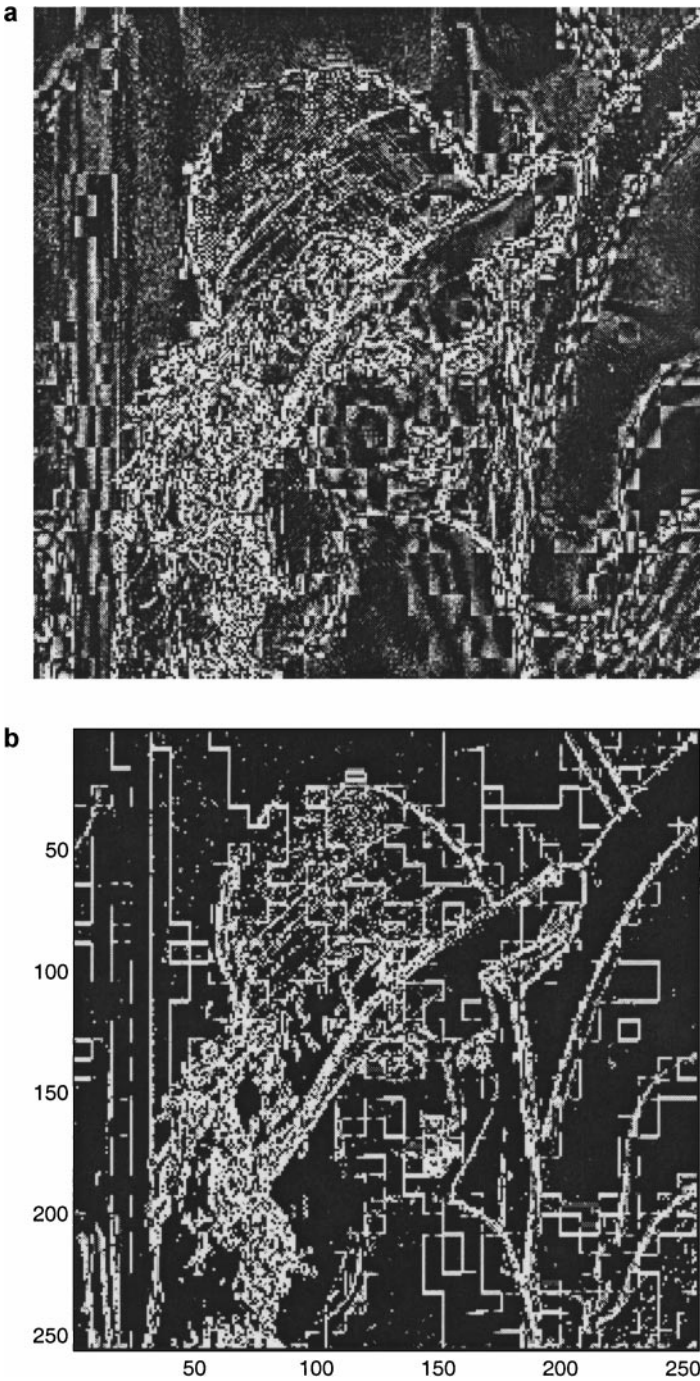


FIG. 12. Difference maps of the DCT-compressed image: (a) pixelwise difference and (b) perceptual difference.

more dominant. On the perceptual difference map, in contrast, the texture region difference is decimated due to the masking effect, which is more consistent with human viewing experiences. The blocking artifact is more dominant in flat or smooth regions with low slopes, since its sharp characteristics generate large contrasts at every resolution.



FIG. 13. Wavelet-compressed Lena image at 0.4 bpp with PSNR = 28.18 dB.

The main artifact for wavelet-based coding algorithms is the ringing artifact, which appears as ripples around the edges due to the truncation and quantization of wavelet components. The ringing artifact also appears in DCT compressed images but is not as severe as the blocking artifact. We used an image coded at a bit rate of 0.4 bpp with PSNR = 28.18 dB. The compressed image is shown in Fig. 13 and the corresponding two difference maps are shown in Fig. 14. The energy of the pixelwise difference map again concentrates on texture regions, and the ringing artifact is rather insignificant by comparison. The ringing artifact can be more easily detected in the perceptual difference map by comparing textured patterns located in the vertical and curved strips of the background in Figs. 14a and 14b.

5.2. *Effect of Viewing Distance*

Since the HVS contrast sensitivity threshold is characterized by the spatial frequency, defined as cycles per degree, one should expect the fidelity measure to vary with the ratio of D , the distance between the observer and the image, and the width W of the image. We used a DCT-compressed Lena image (0.34 bpp, PSNR = 26.43 dB) and a wavelet-compressed Lena image compressed by the embedded zerotree wavelet (EZW) algorithm (0.32 bpp, PSNR = 28.47) as test images. Figure 15 shows the relation between the ratio and the fidelity measure. As the distance between the observer and the image increases, the spatial frequencies of the details (high-frequency components) become even higher, so the visual system attenuation fails to capture the compression artifact. Therefore, the error will diminish as the viewing distance increases. On the other hand, once the viewing distance is decreased to a certain extent, the whole details of the image at the pixel level are perceivable. The fidelity measure will thus reach a maximum when the viewing distance is smaller than this distance. As the viewing distance becomes still smaller, however, due to the bandpass

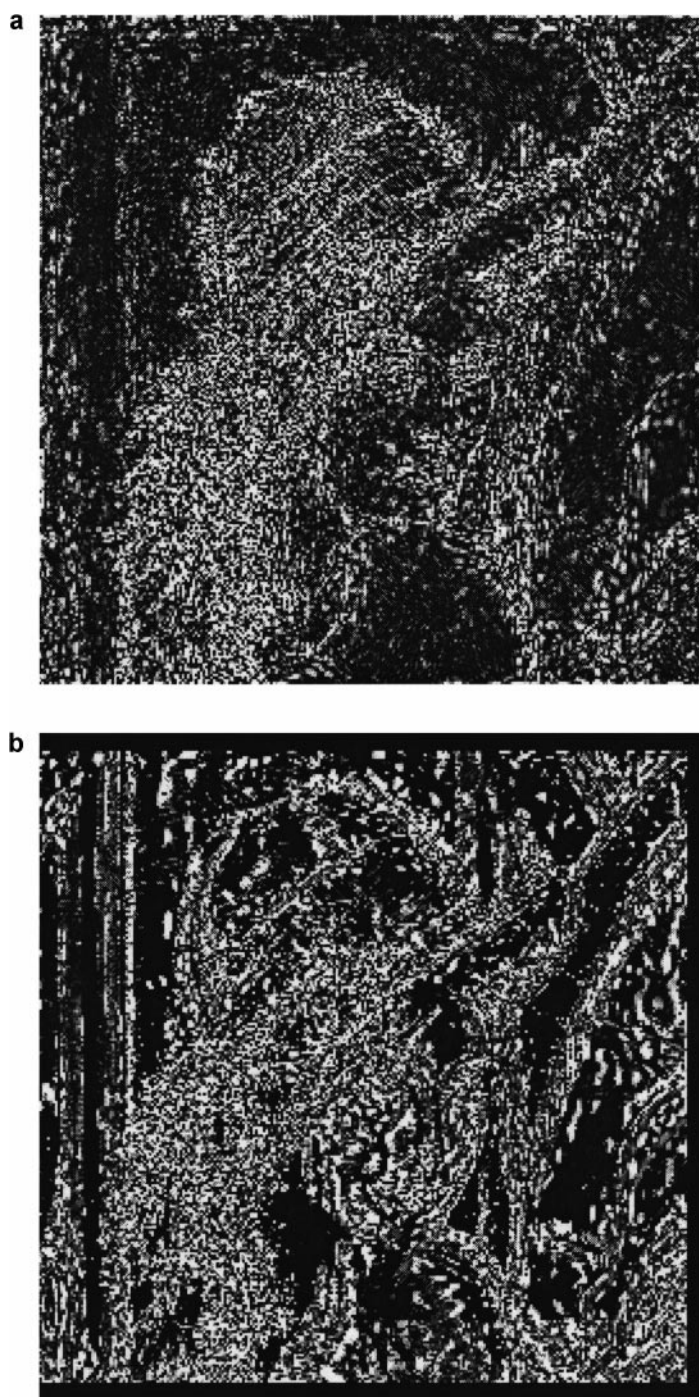


FIG. 14. Difference maps of the wavelet-compressed image: (a) pixelwise difference and (b) perceptual difference.

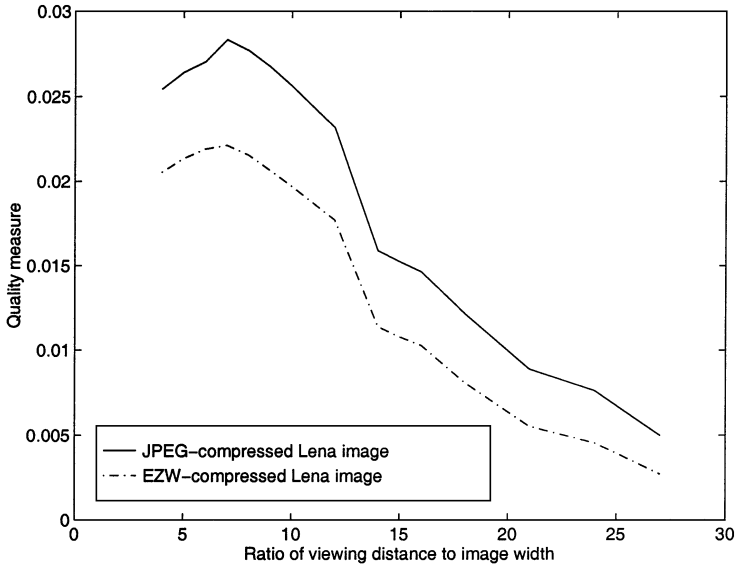


FIG. 15. Fidelity measure as a function of the ratio of the viewing distance and the image width.

characteristics of the HVS sensitivity as shown in Fig. 1, the contrast sensitivity at the low-frequency (coarse resolution) end starts to attenuate. The global variation represented by coarse resolution coefficients begins to extend further into the peripheral region and thus cannot be resolved by foveal vision. The perceived error is therefore lowered by some extent. The “best” viewing distance is about 3–8 times the image width, which is consistent with the rule of thumb in practical image viewing situations. We can also see from Fig. 15 that the EZW-compressed image has a lower fidelity error measure than that of the JPEG-compressed image, although the compression ratio is slightly smaller. This is consistent with the subjective ranking of human observers.

6. CONCLUSION AND FUTURE WORK

In this paper, we investigated a wavelet approach to modeling the human visual system (HVS) and proposed a new fidelity measurement system accordingly. The Haar wavelet was shown to provide local contrast values at each resolution, and the new contrast definition was then incorporated into computational models of visual mechanisms in the design of a new fidelity metric for image fidelity measurement. Experiments showed that Haar filters provide good ability to simulate HVS, and the resulting new metric was useful in measuring compressed image artifacts.

There are a few interesting problems worth further investigation in the future. First, our research work was mainly based on deterministic signal models. It is, however, better to model compressed images as stochastic signals. The generalization of the current work to the context of stochastic signals and the comparison between deterministic and stochastic models should be interesting and useful. Second, more thorough comparison of different still image quality measurement algorithms should be conducted. Although many of the published algorithms left some parameters unspecified, we were able to give good estimates of them for implementation. In our preliminary experiments, all these metrics, including ours, did very well in estimating the perceptual error, so that it is difficult to conclude

any decisive advantage of one algorithm over another. Third, masking in different orientations deserves special treatment. To fully investigate this mechanism with psychophysics experiments would be a demanding task and is beyond the scope of our research. In the literature, psychophysical experiments showed the orientational bandwidth could be as narrow as 10° and as wide as 180° , depending on which group of cortical cells they surveyed, what region (foveal or no foveal), and the test conditions. This work used 45° as an estimate for the orientational selectivity index, which is moderate in this image quality assessment research. Finally, the extension of this work to color image fidelity measures should have practical value. When all three dimensions one considered in color space, HVS computation becomes rather complex. A desired simplification is to find a transform which is able to project the color image onto three perceptually independent axes so that we can process each of the three projected images separately and combine the results. Similarly to luminance perception, chrominance perception is the aggregate response of many individual space–frequency localized channels. It is believed that the masking effect also exists in chrominance dimensions just as in the luminance case. However, this phenomenon has not yet been much discussed in the literature, and few experiments have been devoted to the determination of parameters of this effect. The interaction between luminance and chrominance is very asymmetric. For example, it has been shown that luminance masks have little effect on color contrast detection, while chromatic masks greatly reduce the detectability of luminance contrast. Interactions among chrominance channels and the cross-masking effect between luminance and chrominance dimensions should be investigated.

ACKNOWLEDGMENTS

This research has been funded in part by Intel and in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center with additional support from the Annenberg Center for Communication at the University of Southern California and the California Trade and Commerce Agency. The authors thank Dr. J. Li of Sharp Lab for his valuable discussion and support for this work.

REFERENCES

1. P. G. J. Barten, Evaluation of subjective image quality with the square-root integral method, *J. Opt. Soc. Am. A* **7**(10), 1990, 2024–2031.
2. P. G. J. Barten, Physical model for the contrast sensitivity of the human eye, in *Human Vision, Visual Processing, and Digital Display III*, SPIE Proceedings, Vol. 1666, pp. 57–72, 1992.
3. J. B. Bergen, H. R. Wilson, and J. D. Cowan, Further evidence for four mechanisms mediating vision at threshold: Sensitivities to complex gratings and aperiodic stimuli, *J. Opt. Soc. Am.* **69**, 1979, 1580–1587.
4. F. W. Campbell and J. J. Kulikowski, Orientational selectivity of the human visual system, *J. Physiol.* **187**, 1966, 437–445.
5. F. W. Campbell, J. J. Kulikowski, and J. Levinson, The effect of orientation on the visual resolution of gratings, *J. Physiol.* **187**, 1966, 427–436.
6. F. W. Campbell and J. G. Robson, Application of Fourier analysis to the visibility of gratings, *J. Physiol.* **197**, 1968, 551–566.
7. M. W. Cannon, Jr., Perceived contrast in the fovea and periphery, *J. Opt. Soc. Am. A* **2**, 1985, 1760–1768.

8. C. R. Carlson, Sine-wave threshold contrast-sensitivity function: Dependence on display size, *RCA Rev.* **43**, 1982, 675–683.
9. S. Daly, The visible differences predictor: An algorithm for the assessment of image fidelity, in *Human Vision, Visual Processing, and Digital Display III*, SPIE Proceedings, Vol. 1666, pp. 1–15, 1992.
10. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
11. J. G. Daugman, Two-dimensional spectral analysis of the cortical receptive field profiles, *Vision Res.* **20**, 1980, 847–856.
12. K. K. De Valois and E. Switkes, Simultaneous masking interactions between chromatic and luminance gratings, *J. Opt. Soc. Am.* **73**, 1983, 11–18.
13. R. L. De Valois and K. K. De Valois, *Spatial Vision*, Oxford Univ. Press, Oxford, 1988.
14. J. M. Foley, Human luminance pattern-vision mechanisms: Masking experiments require a new model. *J. Opt. Soc. Am. A* **11**, 1994, 1710–1719.
15. D. Gabor, Theory of communication, *J. Inst. Elect. Eng. (London)* **93**, 1946, 429–457.
16. M. A. García-Pérez, Space-variant visual processing: Spatially limited visual channels, *Spatial Vision* **3**(2), 1988, 129–142.
17. M. A. García-Pérez, The perceived image: Efficient modelling of visual inhomogeneity, *Spatial Vision* **6**(2), 1992, 89–99.
18. M. A. Georgeson, Contrast overconstancy, *J. Opt. Soc. Am. A* **8**, 1991, 579–586.
19. M. A. Georgeson and G. D. Sullivan, Contrast constancy: Deblurring in human vision by spatial frequency channels, *J. Physiol.* **252**, 1975, 627–656.
20. D. J. Granrath, The role of human visual models in image processing, *Proc. IEEE* **69**, 1981, 552–561.
21. C. F. Hall and E. L. Hall, A nonlinear model for the spatial characteristics of the human visual system, *IEEE Trans. Systems Man. Cybernet.* **7**, 1977, 161–170.
22. S. Hecht, The visual discrimination of intensity and the Weber–Fechner law, *Gen. Physiol.* **7**, 1924, 241.
23. R. F. Hess, A. Bradley, and L. Piotrowski, Contrast-coding in amblyopia. I. Differences in the neural basis of human amblyopia, *Proc. R. Soc. London Ser. B* **217**, 1983, 309–330.
24. G. Kaiser, *A Friendly Guide to Wavelets*, Birkhäuser, Boston, 1994.
25. S. A. Karunasekera and N. G. Kingsbury, A distortion measure for blocking artifacts in images based on human visual sensitivity, *IEEE Trans. Image Proces.* **4**, 1995, 713–724.
26. S. A. Klein, T. Carney, L. Barghout-Stein, and C. W. Tyler, Seven Models of Masking, in *Human Vision and Electronic Imaging II*, SPIE Proceedings, Vol. 3016, 1997.
27. C. J. van den Branden Lambrecht, A working spatio-temporal model of the human visual system for image restoration and quality assessment applications, in *Proceedings 1996 International Conference on Acoustics, Speech, and Signal Processing, May 1996*, pp. 2293–2296.
28. D. Costantini, C. J. van den Branden Lambrecht, G. L. Sicuranza, and M. Kunt, Motion rendition quality metric for MPEG coded video, in *Proceedings 1996 IEEE International Conference on Image Processing, Sept. 1996*, pp. 889–892.
29. G. E. Legge, A power law for contrast discrimination, *Vision Res.* **21**, 1981, 457–467.
30. G. E. Legge and J. M. Foley, Contrast masking in human vision, *J. Opt. Soc. Am.* **70**, 1980, 1458–1471.
31. H. Marmolin, Subjective MSE measures, *IEEE Trans. Systems Man. Cybernet.* **16**, 1986, 486–489.
32. E. Peli, Contrast in complex images, *J. Opt. Soc. Am. A* **7**, 1990, 2032–2040.
33. E. Peli, Suprathreshold contrast perception across differences in mean luminance: effects of stimulus size, dichoptic presentation, and length of adaption, *J. Opt. Soc. Am. A.* **12**, 1995, 817–823.
34. E. Peli, L. E. Arend, G. M. Young, and R. B. Goldstein, Contrast sensitivity to patch stimuli: Effects of spatial bandwidth and temporal presentation, *Spatial Vision* **7**(1), 1–14.
35. G. C. Phillips and H. R. Wilson, Orientation bandwidths of spatial mechanisms measured by masking, *J. Opt. Soc. Am. A* **1**, 1984, 226–232.
36. C. Poynton, Rehabilitation of gamma, in *Human Vision and Electronic Imaging III*, SPIE Proceedings, 1998.
37. R. St. John, B. Timney, K. E. Armstrong, and A. B. Szpak, Changes in perceived contrast of suprathreshold gratings as a function of orientation and spatial frequency, *Spatial Vision* **2**(3), 1987, 223–232.
38. G. Wallace, The JPEG still picture compression standard, *Comm. ACM* **34**, 1991, 31–44.

39. H.-J. Wang and C.-C. J. Kuo, A multi-threshold wavelet coder (MTWC) for high fidelity image compression, in *1997 International Conference on Image Processing (ICIP 97)*, IEEE Signal Processing Society, 1997.
 40. A. B. Watson, Perceptual-component architecture for digital video, *J. Opt. Soc. Am. A* **7**, 1990, 1943–1954.
 41. A. B. Watson and J. A. Solomon, A model of visual contrast gain control and pattern masking, *J. Opt. Soc. Am. A*. **14**, 1997, to appear.
-



C.-C. JAY KUO received the B.S. from the National Taiwan University, Taipei, Taiwan in 1980 and the M.S. and Ph.D. from the Massachusetts Institute of Technology, Cambridge, Massachusetts in 1985 and 1987, respectively, all in electrical engineering. Dr. Kuo was Computational and Applied Mathematics (CAM) Research Assistant Professor in the Department of Mathematics at the University of California, Los Angeles from October 1987 to December 1988. Since January 1989, he has been with the Department of Electrical Engineering–Systems and the Signal and Image Processing Institute at the University of Southern California, where he currently has a joint appointment as Professor of Electrical Engineering and Mathematics. His research interests are in the areas of digital signal and image processing, audio and video coding, wavelet theory and applications, multimedia technologies, and Internet and wireless communications. He has authored more than 380 technical publications in international conferences and journals. Dr. Kuo is a member of SIAM and ACM and a Fellow of IEEE and SPIE. He is Editor-in-Chief for the *Journal of Visual Communication and Image Representation* and served as Associate Editor for *IEEE Transaction on Image Processing* during 1995–1998 and *IEEE Transaction on Circuits and Systems for Video Technology* during 1995–1997. Dr. Kuo received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively.



YUNG-KAI LAI received the B.S. from National Taiwan University, Taipei, Taiwan in 1991, and the M.S. and Ph.D. from the University of Southern California, Los Angeles, California, in 1995 and 1998, respectively, all in electrical engineering. He is currently working with Welltel Network, Inc., in Irvine, California, where he develops models and algorithms for perceptual compression, encryption, and watermarking for audio, image, and video. His research interests include audio/image/video perceptual quality assessment, perceptually optimized compression and watermarking, and their application in Internet telephony networks. He is a member of IEEE, SPIE, and OSA.