A Perceptually Tuned Subband Image Coder Based on the Measure of Just-Noticeable-Distortion Profile

Chun-Hsien Chou and Yun-Chin Li

Abstract—**To represent an image of high perceptual quality with the lowest possible bit rate, an effective image compression algorithm should not only remove the redundancy due to statistical correlation but also the perceptually insignificant components from image signals. In this paper, a perceptually tuned subband image coding scheme is presented, where a just-noticeable distortion (JND) or minimally noticeable distortion (MND) profile is employed to quantify the perceptual redundancy. The JND profile provides each signal being coded with a visibility threshold of distortion, below which reconstruction errors are rendered imperceptible. Based on a perceptual model that incorporates the threshold sensitivities due to background luminance and texture masking effect, the JND profile is estimated from analyzing local properties of image signals. According to the sensitivity of human visual perception to spatial frequencies, the full-band JND/MND profile is decomposed into component JND/MND profiles of different frequency subbands. With these component profiles, perceptually insignificant signals in each subband can be screened out, and significant signals can be properly encoded to meet the visibility threshold. A new quantitative fidelity measure, termed as peak signal-to-perceptible-noise ratio (PSPNR), is proposed to assess the quality of the compressed image by taking the perceptible part of the distortion into account. Simulation results show that near-transparent image coding can be achieved at less than 0.4 b/pixel. As compared to the ISO-JPEG standard, the proposed algorithm can remove more perceptual redundancy from the original image, and the visual quality of the reconstructed image is much more acceptable at low bit rates.**

I. Introduction

IT IS GENERALLY believed that the performance of current image coding techniques is not close enough to the fundamental limit on the bit rate at which high reception quality is maintained [1]–[4]. For supporting the future applications that demand high image quality at very low bit rates, the more efficient coding algorithms are expected. Among numerous approaches to reach this optimality, the perceptual coding that matches the compression algorithm to the human perception mechanism has been considered as the most promising solution, and has recently become an important area of research [3], [4].

It is well known that the statistics of image signals are quite nonstationary and the fidelity of the reconstructed images demanded by the human eye differs from pixel to pixel. Consequently, the essential task of perceptual coding is to effectively adapt the coding algorithm to the sensitivity of the human eye. A variety of methods have been proposed to incorporate certain psychovisual properties of the human visual system (HVS) into image coding algorithms [5]–[16], [25]–[29]. Some efforts have been made to exploit the sensitivity of the HVS to spatial frequency for adapting the quantizer stepsize in frequency domain [5]–[9]. The other efforts tried to make efficient use of spatial masking effects to hide distortion in spatial domain [12]–[15]. Nevertheless, due to the lack of an effective quantitative measure for evaluating image quality and the nonlinearity of the HVS, no image coding scheme has yet sufficiently integrated these psychovisual effects to offer a simple and efficient method for removing perceptual redundancies from still images.

In a recent review of image coding techniques, Jayant [3], [4] addressed a key concept of perceptual coding, namely, *just-noticeable-distortion* (JND). The ideal JND provides each signal being coded with a threshold level of error visibility, below which reconstruction errors are rendered imperceptible. The JND profile of a still image is a function of local signal properties, such as background intensity, activity of luminance changes and dominant spatial frequency. The mapping of these properties to the JND profile requires an effective perceptual model from extensive subjective experimentation. Once the JND profile of an image is obtained, the energy of perceptible distortion can be measured and the perceptual significance of each signal can be evaluated. Hence, an effective perceptual model that can produce authentic JND profiles from input images will be the prerequisite for perceptually lossless coding. The method based on the JND concept has been very successful in transparent coding of wideband audio [18]. Supposing transparent coding cannot be attained due to a tight bit-rate budget, a *minimally-noticeable-distortion* (MND) rather than JND profile would be required, such that the distortion presented in the reconstructed image is minimally perceptible and appears to be uniformly distributed over the image [3], [4]. The perceptual quality of the reconstructed image is accordingly expected to degrade gracefully if bit rate is reduced.

Subband coding is a promising approach to achieve high-quality digital image compression [19], [20]. The basic idea of subband coding is to decompose the spatial image signals into narrow frequency subbands, where each subband is decimated and coded separately. The recent interest in subband coding is motivated by the feasibility of incorporating the HVS model to image coding algorithms [2]–[4], [20]–[22]. A subband coding scheme that used the concept of JND profile for image compression was proposed in [21], where the base sensitivity of each subband is first derived empirically from measuring

Manuscript received October 4, 1993; revised March 21, 1995. This work was supported by the National Science Council, Taiwan, ROC, by Grant NSC 82-0404-E-036-023. This paper was recommended by Associate Editor D. Anastassiou.

The authors are with the Department of Electrical Engineering, Tatung Institute of Technology, Taipei, Taiwan, 10451, R.O.C.

IEEE Log Number 9415607.

the energy of noticeable noises at the midgrey background. The base sensitivity of each subband is then adjusted by local brightness and texture energy to obtain the perceptual threshold for setting the stepsize of DPCM quantizer.

Even though a very high percentage of total signal energy is contained in the lowest frequency subband, the truncation or undercoding of high-band signals will result in the perception of the distortion due to aliasing effects. On the other hand, unless the significant signals are cautiously encoded, the overcoding of high-band signals is the price to pay for gaining good image quality. Consequently, the problem to be solved for optimizing the subband coding scheme is how to locate the perceptually important signals in every frequency subband, and how to encode these signals with the lowest possible bit rate without exceeding the error visibility threshold.

In this paper, a simple yet effective perceptual model is proposed to estimate the JND/MND profile of the grey-level image. A subband coding algorithm is then designed to remove the perceptual redundancy as quantified by the JND/MND profile. In Section II, important psychovisual properties of the HVS are reviewed, and the perceptual redundancies inherent in a still image are investigated. In Section III, the proposed model that incorporates two fundamental perceptual effects is described. To verify the validity of the proposed perceptual model in obtaining effectual JND profiles, a subjective test that compares the perceptual quality of the JND-contaminated image with the original image is conducted. In Section IV, the proposed subband coding scheme is described. Considering the fact that the conventional PSNR cannot sufficiently reflect the real perceptual quality of the reproduced image, a new fidelity assessment to measure the perceptible distortion energy remaining in the reconstructed image is defined in Section V. Simulation results are presented in terms of PSNR, PSPNR and the required bit rate. For the purpose of comparison, the coding results obtained from applying the ISO-JPEG coder to the same test images are also presented. A summary of the paper is finally given in Section VI.

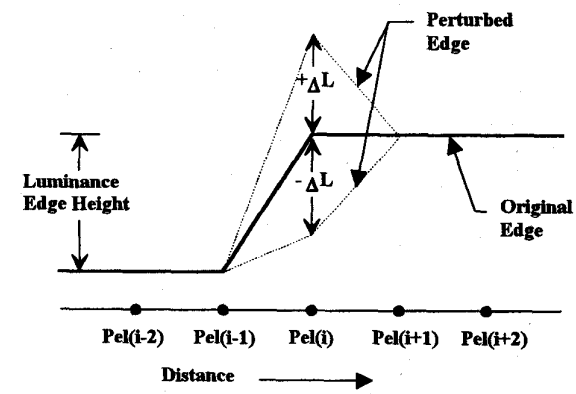
II. Perceptual Redundancies

The perceptual redundancies inherent in a still image are basically due to the inconsistency in sensitivity of the HVS to stimuli of varying levels of contrast and luminance changes in the spatial domain. Applying psychophysics to image coding, the visibility threshold of coding impairment is the most important measure for quantifying perceptual redundancy. It has been defined as the magnitude of the stimulus at which it becomes just visible or just invisible [15]. The visibility threshold of a particular stimulus depends on many factors [1]. Considering only achromatic images in the spatial domain, there are primarily two factors affecting the error visibility threshold of each pixel. One is the average background luminance behind the pixel to be tested. The other is the spatial nonuniformity of the background luminance.

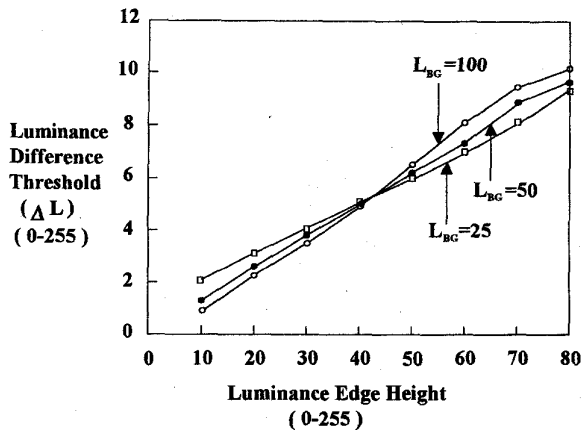
In experiments investigating the effect of the first factor, it was found that human visual perception is sensitive to luminance contrast rather than the absolute luminance value [23], [24]. As indicated by Weber's law [15], [25], if the luminance of a test stimulus is just noticeable from the surrounding luminance, then the ratio of just noticeable luminance difference to stimulus' luminance, known as Weber fraction, is approximately constant. In fact, due to the presence of ambient illumination surrounding the display, the noise in very dark areas tends to be less visible than that occurring in regions of higher luminance. Therefore, the modification has been made that, as the background luminance is low, the Weber fraction increases as the background luminance decreases [15]. On the other hand, if the background luminance is high, the Weber fraction remains constant as the background luminance is increased. This modification has been applied to perceptually tuned image coder [21], where high visibility thresholds are assumed in either very dark or bright regions, and low thresholds in regions of medium grey levels around 127.

The second factor reflects the fact that the reduction in the visibility of stimuli is caused by the increase in the spatial nonuniformity of the background luminance. This fact is known as spatial masking. Several efforts have been made to utilize some forms of spatial masking to improve the coding efficiency [12]–[16], [26]–[31]. However, masking effect is a very complicated process that no single theoretical formulation has been able to justify various forms of masking. Many investigations on this effect have been conducted to determine the luminance difference threshold for a small spot or line of stimulus at positions close to the luminance edge [13], [15], [27]–[31]. Spatial masking has been widely used to optimize the differential pulse-code modulation (DPCM) coders, where the quantizer of variable stepsizes is designed in such a way that quantization errors can be masked by the luminance transitions [12], [26]–[31]. In many approaches, the visibility thresholds of luminance difference are defined as a function of the amplitude of luminance edge, in which the perturbation is varied until it become just noticeable [12], [15]. Fig. 1 shows one of the experimental results that relates the luminance difference threshold to the luminance edge height [15]. These results also indicate that the visibility threshold depends not only on the edge height but also on the background luminance. In [12], a masking function of spatial activity is calculated at each pixel as the weighted sum of the horizontal and vertical luminance slopes at the neighboring pixels, where the weight decreases as the distance between the neighboring pixel and the central pixel increases. A noise visibility function is then defined to measure the subjective magnitude of the test stimulus as the masking function exceeds a given threshold. The experimental outcomes showed that high values of masking function generally lead to low values of visibility function, and vice versa. These results also verified that the decrease in noise visibility occurs near the spatial details and fewer pixels have high spatial details in most pictures. The approach of Musmann and Erdmann [13], [31] can be considered as a simplified version of the above masking model, which keeps the quantization error below a visibility threshold. The visibility threshold in this approach is associated with the masking function defined at each pixel as the maximum prediction error from the four neighbouring pixels.

In addition to the investigation on the threshold sensitivity in the spatial domain, many psychovisual studies have shown



(a)



(b)

Fig. 1. (a) Stimuli for determining the visibility thresholds due to luminance change, (b) difference visibility thresholds at different background luminances [15].

that the perception of distortion depends on its frequency distribution [5]–[10]. The contrast sensitivity at different spatial frequencies has been measured and modelled by the modulation transfer function [5]. From the response curve of the MTF shown in Fig. 2, it indicates that the HVS generally has a quasi-bandpass characteristic. The sensitivity of distortion at higher spatial frequencies is lower than that at lower spatial frequencies. In Section IV, the property of the MTF is employed to derive an energy weighting function for decomposing the JND signals into components of different frequency bands.

III. The Proposed Perceptual Model

The elemental work of this paper is to incorporate the properties of the HVS into the estimation of the JND profile for measuring the perceptual redundancies inherent in an image. Since no unified HVS model has been established for this purpose, only well-known properties of the HVS, as mentioned in the previous section, are employed in developing a perceptual model that estimates, from data in the spatial domain, the JND value associated with each pixel of the image. In real-life images, the visibility threshold of JND could be a very complicated function of the two factors

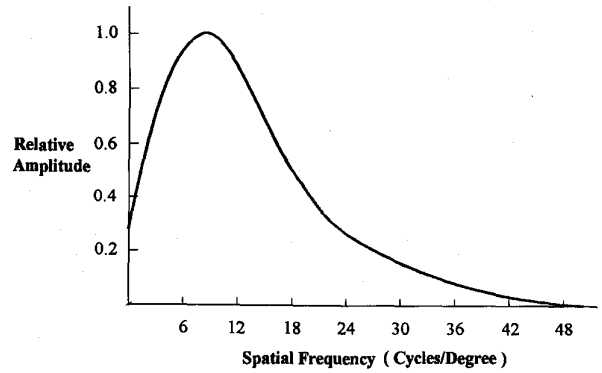


Fig. 2. A typical modulation transfer function of the HVS.

mentioned above. However, for simplicity, the proposed model is obtained by simplifying the inter-relevance of the two factors, and assuming that the JND value is determined by the dominant effect of the two factors. The perceptual model for estimating the full-band JND profile is described by the following expressions

$$\mathrm{JND}_{fb}(x,y) = \max\{f_1(bg(x,y), mg(x,y)), f_2(bg(x,y))\} \tag{1}$$

$$f_1(bg(x,y), mg(x,y)) = mg(x,y)\alpha(bg(x,y)) + \beta(bg(x,y)) \tag{2}$$

$$f_2(bg(x,y)) = \begin{cases} T_0 \cdot (1 - (bg(x,y)/127)^{1/2})) + 3 \\ \quad \text{for } bg(x,y) \le 127 \\ \gamma \cdot (bg(x,y) - 127) + 3 \\ \quad \text{for } bg(x,y) > 127 \end{cases} \tag{3}$$

$$\alpha(bg(x,y)) = bg(x,y) \cdot 0.0001 + 0.115 \tag{4}$$

$$\beta(bg(x,y)) = \lambda - bg(x,y) \cdot 0.01 \quad \text{for} \quad 0 \le x < H \quad 0 \le y < W \tag{5}$$

where $bg(x,y)$ and $mg(x,y)$ are the average background luminance and the maximum weighted average of luminance differences around the pixel at (x,y), respectively. H and W denote respectively the height and width of the image. The spatial masking effect is modelled by function, $f_1(x,y)$, the linear behaviour of which is obtained from the tests similar to those conducted in [12] and [15], and is close to the relationship shown in Fig. 1. As can be noted in Fig. 1, the approximated slope of the line that relates the visibility threshold to the luminance difference tends to increase slightly as the background luminance increases. The parameters $\alpha(x,y)$ and $\beta(x,y)$ are the background-luminance dependent functions that specify the slope of the line and the intersection with the visibility threshold axis. The visibility threshold due to background luminance is given by function $f_2(x,y)$, in which the relationship between noise sensitivity and the background luminance is verified by a subjective test. In the experiment, a small square area, 32 × 32 pixels, is located in the center of a flat field of constant grey level. For each possible grey level of the flat field, the noises of fixed amplitude are randomly added to or subtracted from the pixels within the square area. Through varying the amplitude of the

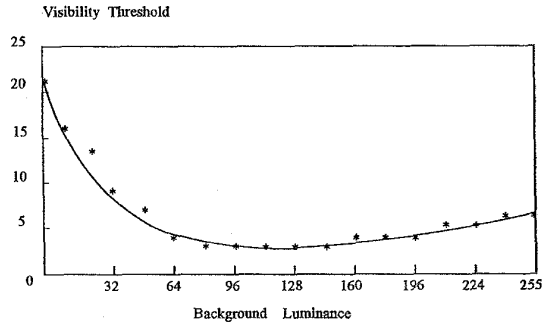


Fig. 3. Visibility thresholds due to background luminance.

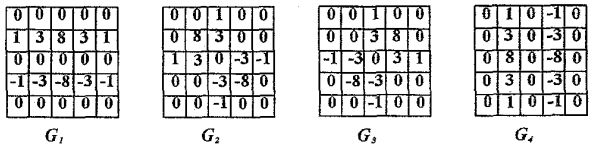


Fig. 4. Operators for calculating the weighted average of luminance changes in four directions.

noise, the visibility threshold for each grey level is determined when the square region contaminated by the noises is just noticeable. The experimental outcome is found close to the test results obtained in [21], and is depicted in Fig. 3 where the relationship corresponding to low background luminance (below 127) is modelled by a root equation while the other part (over 127) is approximated by a linear function. Both of the above experiments are conducted in a dark room with a 19–in SPARC-workstation monitor, against which a viewing distance of approximately 6 times image height is taken. In $f_2(x,y), T_0$ and γ denote, respectively, the visibility threshold when background grey level is 0, and the slope of the line that models the function at higher background luminance. The λ in $\beta(x,y)$ affects the average amplitude of visibility threshold due to spatial masking effect. As can be found in the experiments, the values of these three variables increases as the viewing distance increases. Hence, the JND/MND profile can be modelled as a monotonically increasing function of viewing distance. For the viewing distance taken in the above experiments, T_0, γ and λ are found to be 17, 3/128, and 1/2, respectively. In the case where the bit-rate budget is tight, the coding distortion is expected to be uniformly distributed over the reconstructed image, such that the impairment at different locations have the same visibility and the image quality will not degrade abruptly with the creation of any obvious spatially localized distortion. The MND profiles of different distortion levels are hence required to optimize the coding efficiency at different bit rate budget. In this paper, the MND profile is simply obtained from upraising the JND profile by multiplying every element of JND profile by a constant scale factor as a distortion index. Thus, an MND profile with a distortion index, $\boldsymbol{d}$, can be expressed as

$$\mathrm{MND}_{d,fb}(x,y) = \mathrm{JND}_{fb}(x,y) \cdot \boldsymbol{d} \quad \text{for} \quad 0 \le x < H, \quad 0 \le y < W, \tag{6}$$

where the value of $\boldsymbol{d}$ ranges from 1.0 to 4.0.

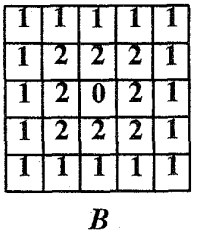


Fig. 5. The operator for calculating the average background luminance.

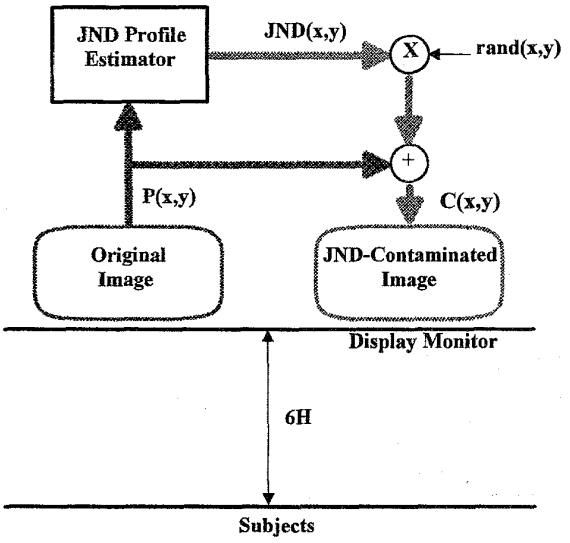


Fig. 6. A subjective test that compares the JND-contaminated image to the original image for refining the proposed perceptual model and verifying its validity.

The $mg(x,y)$ across the pixel at (x,y) is determined by calculating the weighted average of luminance changes around the pixel in four directions. As shown in Fig. 4, four operators, $G_k(i,j)$, for $k = 1, \cdots, 4$, and $i,j = 1, \cdots, 5$, are employed to perform the calculation, where the weighting coefficient decreases as the distance away from the central pixel increases.

$$mg(x,y) = \max_{k=1,2,3,4} \{|\mathrm{grad}_k\,(x,y)|\} \tag{7}$$

$$\mathrm{grad}_k\,(x,y) = \frac{1}{16} \sum_{i=1}^{5} \sum_{j=1}^{5} p(x-3+i, y-3+j) \cdot G_k(i,j) \quad \text{for} \quad 0 \le x < H, \quad 0 \le y < W, \tag{8}$$

where $p(x,y)$ denotes the pixel at (x,y). The average background luminance, $bg(x,y)$, is calculated by a weighted low-pass operator, $B(i,j), i,j = 1, \cdots, 5$, shown in Fig. 5.

$$bg(x,y) = \frac{1}{32} \sum_{i=1}^{5} \sum_{j=1}^{5} p(x-3+i, y-3+j) \cdot B(i,j). \tag{9}$$

To justify the validity of the proposed perceptual model, a subjective test that compares the perceptual quality of the image contaminated by the noises of JND profile with the original image is conducted (Fig. 6). If the JND profiles obtained from the proposed perceptual model are accurate, the perceptual quality of the corresponding JND-contaminated

(a) (b)

(c)

Fig. 7. (a) Original "*Lenna*" image, (b) the "*Lenna*" image contaminanted by the JND profile, and (c) the JND profile of "*Lenna*" image in which the maximum value, 20, is represented by the brightest pixel and the minimum value, 3, by the darkest pixel.

image should be as good as the original image while the PSNR of the contaminated image should be as low as possible. Fig. 7 shows the estimated JND profile of a test image, with which the original image is contaminated by randomly adding to or subtracting from each pixel with its corresponding JND value. A contaminated image $c(x, y)$ can thus be obtained as

$$c(x,y) = p(x,y) + \text{rand}\,(x,y) \cdot \text{JND}_{fb}\,(x,y)$$
$$\text{rand}\,(x,y) = 1 \text{ or } -1, \quad \text{for} \quad 0 \le x < H, \quad 0 \le y < W. \tag{10}$$

As the images shown in Fig. 7 are displayed on the above-mentioned monitor in a dark room, we can hardly discern the difference between the original and the contaminated image at a viewing distance of about 6 times the image height. The PSNR of the contaminated image is 33.1 dB. This measure not only can be used to quantify the amount of imperceptible distortion allowed for transparent coding of images at a specified viewing distance, but also to refine the perceptual model to gain a more accurate estimation of the JND profile at a given viewing distance. The test image contaminated by the MND profile of $\boldsymbol{d} = 3.0$ is shown in Fig. 8.

Fig. 8. The "*Lenna*" image contaminated by the MND profile of $\boldsymbol{d} = 3.0$.

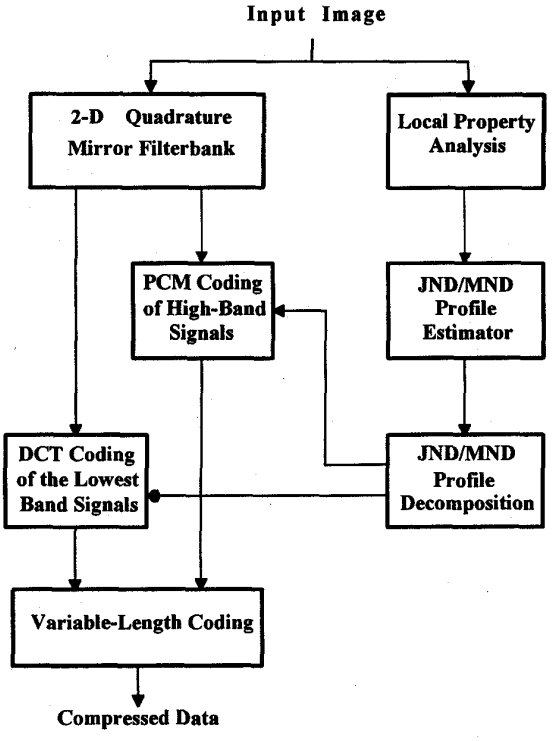


Fig. 9. The functional block diagram of the proposed subband image coder.

IV. The Proposed Coding Scheme

The functional block diagram of the proposed subband coder is shown in Fig. 9, where the whole system is divided into three parts. The first part is the filterbank for subband decomposition. The second part includes the proposed perceptual model for estimating the JND/MND profile of the input image and a mechanism for decomposing the full-band JND/MND profile into component JND/MND profiles of different frequency subbands. The third part is the coder that encodes the perceptually important subsignals with appropriate coding parameters determined by the corresponding JND/MND profile.

A. Subband Coding

Subband image coding is an efficient method for achieving high-quality image compression at low bit rates. It has the advantage of avoiding blocky type artefacts over the coding schemes based on discrete cosine transform (DCT) [15], [19], [20]. Subband coding is a type of hierarchical coding scheme that provides the compatibility among different coding systems by splitting the source image into layers of subimages with different resolutions [32], [33]. To reach a near-perfect reconstruction, the analysis and synthesis of image signals are usually accomplished by 2-D quadrature mirror filterbanks (QMFs) which permit an alias-free reconstruction of the input signal [19], [34]. In the simulation of the proposed subband coder, a 16-tap QMF [36] is used to decompose the input image into 16 equal subbands (Fig. 10), where the filtering is performed by circular convolution to suppress the errors near the margin of the image [35].

In many subband coding schemes, compression is achieved by allocating a proper number of bits to each subband on a predefined criterion [19], [20]. In this paper, subband coding is optimized by allocating bits to perceptually important signals such that coding errors are confined within the JND or MND profile. In order that the perceptually important signals can be effectually located and encoded, the perceptual significance of each signal must be evaluated.

B. The Decomposition of the JND/MND Profile

In order to evaluate the perceptual significance of each signal in a subband, the JND/MND profile corresponding to each subband must be found. In the spatial frequency domain, the HVS acts as a quasi-bandpass system. The distortion of high spatial frequencies requires higher energy to be visible than the distortion of mid-spatial frequencies. According to the sensitivity of the HVS to spatial frequencies, a weighting function to dispense the full-band JND/MND energy to different subbands is derived from computing the average MTF value of each subband. For the decomposition of 16 equal subbands, the relationship between the full-band JND (or MND) energy profile and the component JND (or MND) energy profiles can be given by

$$\mathrm{JND}_q^2(m,n) = \left[\sum_{i=0}^{3}\sum_{j=0}^{3} \mathrm{JND}_{fb}^2(i+m\cdot 4, j+n\cdot 4,\right]\cdot\omega_q \quad \text{for } q=0,1,\cdots,15, \quad \text{and} \quad 0\le m<H/4, \quad 0\le n<w/4, \tag{11}$$

where $\mathrm{JND}_q(m,n)$ denotes the magnitude of JND at (m,n) of the qth subband. ω_q represents the weight for the qth subband and is defined as the normalized spatial-frequency desensitivity,

$$\omega_q = \left(S_q\cdot\sum_{k=0}^{15} S_k^{-1}\right)^{-1}, \quad \text{for} \quad q=0,1,\cdots,15, \tag{12}$$

where S_k denotes, in terms of signal energy, the average sensitivity of the HVS to spatial frequencies in the kth subband. As can be comprehended from (11) and (12), the distortion visibility threshold of each subband is inversely proportional to the relative sensitivity of the corresponding spatial frequencies. The average sensitivity S_k is calculated by

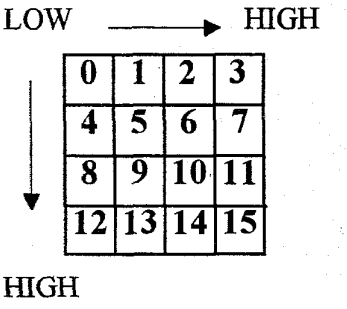


Fig. 10. The numbering of 16 equal subbands.

$$S_k = \frac{16}{H\cdot W}\sum_{u=\varepsilon_k\cdot h}^{(\varepsilon_k+1)h-1}\ \sum_{v=\rho_k\cdot w}^{(\rho_k+1)w-1}\xi(u,v), \quad \text{for} \quad k=0,1,\cdots,15, \tag{13}$$

where

$$h=H/4, \quad w=W/4, \quad \varepsilon_k=\lfloor k/4\rfloor, \quad \rho_k=k-\varepsilon_k\cdot 4$$

and $\xi(u,v)$ denotes the response curve of the MTF for $0\le u<H, 0\le v<W$. A generalized formula for fitting the response curve of the MTF can be represented by

$$\xi(u,v)=a\cdot\left[b+\left(\frac{\Omega(u,v)}{\Omega_0}\right)\right]\cdot\exp\left[-\left(\frac{\Omega(u,v)}{\Omega_0}\right)^c\right] \tag{14}$$

where

$$\Omega(u,v)=\left[\left(\frac{32v}{W}\right)^2+\left(\frac{24u}{H}\right)^2\right]^{1/2} \quad \text{for} \quad 0\le u<H, \quad 0\le v<W \tag{15}$$

is the spatial frequency in cycles per degree. The calculation of spatial frequency is based on a viewing distance of about 6 times the image height and the aspect ratio of 4:3. It is assumed that, at this viewing distance, 64 pixels in horizontal direction subtend approximately one degree of vision [9], [15]. The constants a, b, c, and Ω_0 are parameters that shape the MTF curve [24]. It has been found that the MTF curve modelled by $a=2.6, b=0.0192, c=1.1$, and $\Omega_0=8.772$ is useful to typical applications of image compression [5]. As listed in Table I are the JND energy weights derived from this MTF curve for the 16 equal subbands of the 512 × 512 image.

C. The Coding of the Lowest Band

The information contained in the lowest band is the dominant part of the whole image. The quality of the reconstructed lowest band thus substantially affects the overall quality of the reconstructed image. Since the pixel-to-pixel correlation of this band is still very high, a 2-D DCT algorithm is employed to encode this band. The coding algorithm based on DCT has been widely recognized as a near optimum approach in terms of energy compaction and decorrelation between pixels [7]–[9]. However, to effectually remove the perceptual

TABLE I
THE JND-ENERGY WEIGHTS OF 16 EQUAL SUBBANDS FOR THE 512 × 512 IMAGE

0.0211	**0.0197**	**0.0273**	**0.0445**
0.0221	**0.0261**	**0.0365**	**0.0581**
0.0396	**0.0464**	**0.0626**	**0.0954**
0.0865	**0.0990**	**0.1286**	**0.1865**

redundancy from signals in the transform domain, the perceptual significance of each transform coefficient needs to be evaluated, and the significant coefficient must be quantized with the proper stepsize.

In this paper, the lowest band is first divided into nonoverlapping blocks of dimension 8 × 8, to each of which the 2-D DCT transform used in [37] is applied. To locate the perceptually significant transform coefficients in a block, it is first assumed that the low-sequency ac coefficients with the magnitude less than 0.5 and the high-sequency ac coefficients with magnitude less than 1.0 are considered as insignificant. Second, as conformed to Parseval's theorem, the total coding-error energy resulting from quantizing all ac coefficients is assumed equal to the distortion energy given by the lowest-band JND/MND profile of the block. This indicates that no error tolerance is allowed in encoding dc coefficients. Further assumption is made that the block's distortion energy excluding the part of low-magnitude ac coefficients is dispensed uniformly by the ac coefficients with larger magnitude. Hence, the local threshold, $\Gamma(f,g)$, for assessing the perceptual significance of the large-magnitude coefficients in block (f,g) can be obtained by the following equations

$$\Gamma(f,g) = \sqrt{\frac{R(f,g)}{N(f,g)}}, \tag{16}$$

$$R(f,g) = \sum_{i=0}^{7}\sum_{j=0}^{7} \mathrm{JND}_0^2\,(f\cdot 8+i, g\cdot 8+j) - \sum_{(u,v)\in A_{f,g}} F_{f,g}^2(u,v) - \sum_{(u,v)\in B_{f,g}} F_{f,g}^2(u,v) \tag{17}$$

$$N(f,g) = 63 - \sum_{(u,v)\in A_{fg}} 1 - \sum_{(u,v)\in B_{f,g}} 1 \tag{18}$$

$$A_{f,g} = \{(u,v)|F_{f,g}(u,v) < 0.5, \quad \text{for} \quad 0 \le u \le 3, \quad 0 \le v \le 3 \quad \text{and} \quad (u,v) \ne (0,0)\} \tag{19}$$

$$B_{f,g} = \{(u,v)|F_{fg}(u,v) < 1.0, \quad \text{for} \quad 4 \le u \le 7, \quad 4 \le v \le 7\} \quad \text{for} \quad 0 \le f < H/8, \quad 0 \le g < W/8 \tag{20}$$

where $\{F_{f,g}(u,v)\}$ denotes the DCT transform coefficients of the block (f,g). The coefficient with a magnitude less than the local threshold will also be regarded as perceptually insignificant and discarded. On the other hand, the coefficient with a magnitude larger than the local threshold is considered as perceptually significant, and is coded by an uniform quantizer with a stepsize set to the local threshold. The smallest stepsize of the uniform quantizer is set to 1.0 to avoid the overcoding occurred in the case of low JND/MND energy. Since the correlation between the average luminance of neighbouring blocks is high, the dc coefficients of the lowest band are DPCM encoded with the variable-length code. The pixel configuration for 2-D prediction is the same as that proposed in the ISO-JPEG [38]. The positions of perceptually important ac coefficients are located through a zigzag scan. The run-length between any two nonzero quantized coefficients and the quantization indices of ac coefficients are variable-length encoded to achieve the best compression.

D. The Coding of High-Frequency Bands

In high frequency subbands, only signals with magnitude larger than the corresponding JND/MND values are considered as perceptually significant and encoded. To encode these perceptually significant signals, each subband is first partitioned into nonoverlapping 4 × 4 subblocks. Depending upon the number of significant signals in a subblock, all subblocks are classified into important and unimportant types. The subblock that contains less than two significant points is regarded as unimportant. This is based on the observation that few isolated points do not affect the perceptual quality. The significant signals are therefore located by performing two layers of scanning. The first-layer scanning is performed on a subblock-by-subblock basis to locate important subblocks. Only signals within the important subblocks are scanned in the second layer and PCM encoded. The scanning schemes of both layers in a subband is determined by the band location in the spatial frequency space. The signals in subbands #1, #2, #3 are scanned horizontally, and those in subbands #4, #8, #12 are scanned vertically. The signals of the remaining high frequency subbands are zigzag scanned. The perceptually significant signals in a subblock are quantized with the same stepsize which is set to the minimum of the selected JND/MND values, but not less than 1.0 to avoid overcoding. The location of important subblocks in each subband is run-length encoded, in which the runs of important subblocks are variable-length encoded. The runs of significant points in each subblock, quantization indices of significant signals, and the quantizer stepsize of each important subblock are all variable-length encoded by the Huffman code.

V. SIMULATION RESULTS AND DISCUSSION

To assess the perceptual quality of the reconstructed image and to rate the performance of the proposed coding algorithm, an effective fidelity criterion is needed. The peak signal-to-noise ratio

$$\mathrm{PSNR} = 20\ \log_{10}\ \frac{255}{\sqrt{\mathrm{MSE}}} \tag{21}$$

(a)

(b)

(c)

Fig. 11. The reconstructed images of (a) "*Lenna*" coded at 0.34 bpp and (b) "*Pepper*" coded at 0.38 bpp. (c) Original "*Pepper*" image.

is a widely used measure of image quality. It however cannot accurately reflect the real perceptual quality of the reconstructed image, particularly at low bit rates [5], [39]. Consequently, a quantitative fidelity measure that is conformable to the subjective evaluation of image quality is desired [40]. In this paper, a computable fidelity measure is defined to evaluate the quality of the compressed image in terms of the perceptible distortion energy. In this measure, the perceptible distortion is referred to the part by which the distortion is noticeable, or the part of distortion that exceeds the threshold given by the JND profile. The new fidelity criterion, as termed by peak signal-to-perceptible-noise ratio (PSPNR), is defined as

$$\mathrm{PSPNR} = 20 \cdot \log_{10} \frac{255}{\sqrt{E\{[|p(x,y) - \hat{p}(x,y)| - \mathrm{JND}_{fb}(x,y)]^2 \cdot \delta(x,y)\}}} \tag{22}$$

$$\delta(x,y) = \begin{cases} 1, & \text{if } |p(x,y) - \hat{p}(x,y)| > \mathrm{JND}_{fb}(x,y) \\ 0, & \text{if } |p(x,y) - \hat{p}(x,y)| \le \mathrm{JND}_{fb}(x,y) \end{cases} \quad \text{for} \quad 0 \le x < H, \quad 0 \le y < W \tag{23}$$

where $\hat{p}(x,y)$ denotes the reconstructed pixel at (x,y). The PSPNR is obviously dependent upon the JND profile, and thus a function of viewing distance. Therefore, for a reconstructed image of fixed PSNR, the average level of JND profile increases as the viewing distance increases, and so does the associated PSPNR.

TABLE II
PSNRs, PSPNRs, AND BIT RATES FOR ACHIEVING NEAR-TRANSPARENT CODING OF TEST IMAGES

	Bit Rate	PSNR	PSPNR
LENNA	0.34	34.50	44.92
PEPPER	0.38	33.96	42.18

A variety of grey-level images, ranged from simple to complex, have been encoded by the proposed algorithm with the corresponding JND/MND profiles. The coding results show that, as viewed at a distance of at least 6 times the image height, the transparent quality can be achieved at the bit rates ranging from 0.1 to 0.5 bpp. In average, very high perceptual quality or near-transparent quality can be obtained at the bit rate less than 0.4 bpp for typical images. In Fig. 11, the reconstructed images of "*Lenna*" and "*Pepper*" are shown, which are encoded with the corresponding JND profiles. Viewing these reconstructed images on the display monitor at the specified distance, we can hardly discern the differences between the original and the reconstructed image. As indicated in Table II, the near-transparent coding of these images is achieved at the bit rates of less than 0.4 bpp, and the difference between PSNR and PSPNR quantifies the amount of perceptible distortion removed by the proposed coding algorithm.

For the purpose of comparison, the reconstructed images of different bit rates are compared with those encoded by the ISO-JPEG standard. The bit rate required by the proposed coding algorithm can be controlled by varying the distortion index of the MND profile, while the bit rate required by the JPEG algorithm can be controlled by varying the scale factor of the quantization table [38]. For instance, the MND profiles of $d = 3.6, 1.6$, and 1.25 are employed to compress the image "*Pepper*" at the bit rates of 0.15, 0.23, and 0.33 bpp, respectively. The reconstructed images encoded at low bit rates are shown in Figs. 12 and 13. We can notice that the perceptual quality of the images compressed by the proposed algorithm is more acceptable than that of the images compressed by the JPEG algorithm. The decrease in the distortion index of MND profile increases the fidelity of the compressed image. A higher bit rate is required to remove more perceptible distortion, and thus resulting in higher values of PSPNR and PSNR. The values of PSPNR and PSNR versus the required bit rates for both the proposed algorithm and JPEG algorithm are plotted in Figs. 14 and 15, which indicate that the performance of the proposed algorithm is superior to that of the ISO-JPEG standard. The amount of perceptual redundancy, (PSPNR-PSNR), removed by the proposed algorithm is larger than that removed by the JPEG algorithm, and increases with the increasing bit rate. At higher bit rates, the perceptual redundancy removed by the JPEG algorithm is rather limited. This is especially the case in encoding the image "*Pepper*".

Although high perceptual quality can be achieved at low bit rates in encoding the test images by the proposed algorithm with JND profiles, a difference between the PSNR of the

(a) (b)

Fig. 12. Reconstructed images of "*Lenna*" coded at 0.14 bpp by (a) the proposed algorithm, and by (b) the JPEG algorithm.

(a) (b)

Fig. 13. Reconstructed images of "*Pepper*" coded at 0.15 bpp by (a) the proposed algorithm, and by (b) the JPEG algorithm.

compressed image and that of the JND-contaminated image exists. This indicates that a certain amount of perceptual redundancy is still remaining in the reconstructed image, and implies that some parts of the signals are overcoded. The bit rate for transparent coding is therefore likely to be further reduced if the JND profile can be effectually utilized. This can be achieved by optimizing the allocation of JND energy to signals of different perceptual significance. On the other hand, some overcoding may be due to the use of the minimum stepsize in encoding the significant signals of high frequency bands. This overcoding can be minimized through a statistical approach to determine a larger stepsize rather than the minimum of the selected JND values. For the DCT coding of the lowest-band signals, a more complicated approach to dispense the block JND energy to transform coefficients as well as to determine the adequate quantizer stepsize is currently being investigated.

VI. Summary

The intent of this paper is to optimize the subband coding scheme by exploiting the psychovisual properties of the human visual system, such that high perceptual quality of the compressed image can be maintained at very low bit rates. In order to reach the optimality where the coding impairment is invisible or minimally noticeable at a given viewing distance, the approach of JND/MND profile is conducted in two aspects. One is to develop an effective HVS model for obtaining an accurate JND/MND profile by which the perceptual redundancy is quantified. The other is to design a subband coding

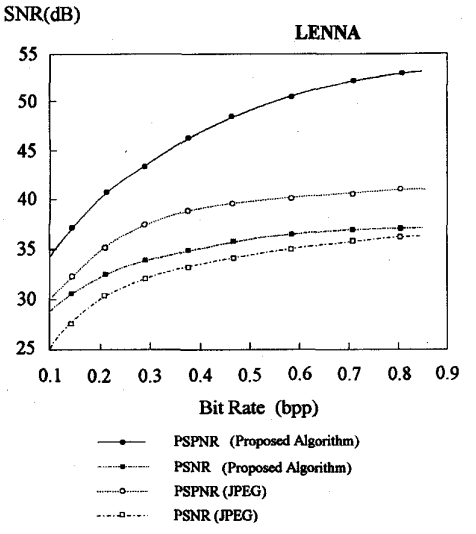


Fig. 14. Plots of PSPNR and PSNR versus bit rate for "*Lenna*" image.

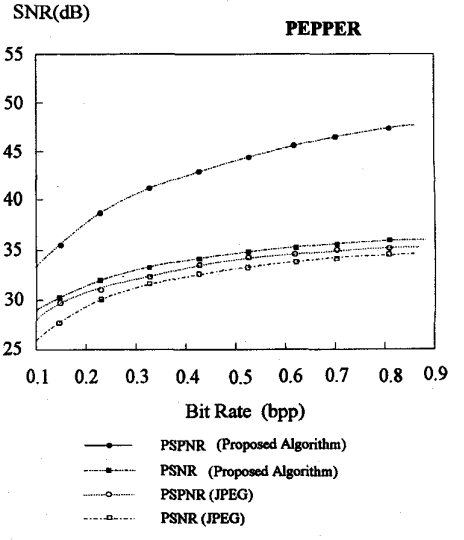


Fig. 15. Plots of PSPNR and PSNR versus bit rate for "*Pepper*" image.

algorithm for removing the perceptual redundancy. Since no unified HVS model is available for this purpose, only well-known psychovisual properties of the HVS are embodied in the proposed perceptual model. Even though the proposed model is somewhat simple and might not be the optimum, it is found feasible in quantifying perceptual redundancies and effective in improving coding efficiency for most of the real-life images. The estimation of JND profiles is performed in the spatial domain through analyzing local properties of image signals. The major part of the coding algorithm is, however, based on the analysis in the frequency domain. The JND/MND profile is accordingly decomposed by the MTF of the HVS into component profiles for screening out perceptually insignificant signals. A new fidelity measure is proposed in this paper to evaluate the quality of the compressed image, in which only the part of coding errors exceeding the JND profile is taken into account. The simulation results show that, as

compared to the ISO-JPEG standard, the proposed algorithm can effectually remove more perceptual redundancy and reach a higher perceptual quality at lower bit rates. For typical real-life images, the transparent or near-transparent quality can be obtained at the bit rates less than 0.4 bpp.

REFERENCES

[1] K. H. Tzou, T. R. Hsing, and J. G. Dunham, "Applications of physiological human visual system model to image compression," in *Proc. SPIE Conf. Applicat. Digital Image Process. VII,* vol. 504, 1984, pp. 419–424.

[2] R. Forchheimer and T. Kromamder, "Image coding—From waveforms to animation," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 37, pp. 2008–2023, Dec. 1989.

[3] N. Jayant, "Signal compression: Technology targets and research directions," *IEEE J. Select. Areas Commun.,* vol. 10, pp. 314–323, June 1992.

[4] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE,* vol. 81, pp. 1385–1422, Oct. 1993.

[5] J. L. Mannos and D. J. Sakrison, "The effect of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory,* vol. IT-20, pp. 525–536, July 1974.

[6] C. F. Hall and E. L. Hall, "A nonlinear models for the spatial characteristics of the human visual system," *IEEE Trans. Syst., Man, Cybernet.,* vol. SMC-7, pp. 161–170, Mar. 1977.

[7] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.,* vol. COM-33, pp. 551–557, June 1985.

[8] K. N. Ngan, K. S. Leong, and H. Singh, "Adaptive cosine transform coding of images in perceptual domain," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 37, pp. 1743–1749, Nov. 1989.

[9] D. L. McLaren and D. T. Nguyen, "Removal of subjective redundancy from DCT-coded images," *IEE Proc.—I1,* vol. 138, pp. 345–350, Oct. 1991.

[10] C. F. Stromeyer III and B. Julesz, "Spatial-frequency masking in vision: Critical bands and spread of masking," *J. Opt. Soc. Amer.,* vol. 62, pp. 1221–1232, Oct. 1972.

[11] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE Conf.,* vol. 1913, 1993, pp. 202–216.

[12] A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE,* vol. 65, pp. 536–548, Apr. 1977.

[13] H. G. Musmann and W. D. Erdmann, German Patent Appl. No. P 2740945.6., 1977.

[14] J. O. Limb, "On the design of quantizer for DPCM coder: A functional relationship between visibility, probability and masking," *IEEE Trans. Commun.,* vol. COM-26, pp. 573–578, 1978.

[15] A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression.* New York: Plenum, 1988.

[16] D. J. Sakrison, "Image coding applications of vision models," in *Image Transmission Techniques,* W. K. Pratt, Ed. New York: Academic, May 1979, pp. 21–51.

[17] J. B. O. S. Martens and G. M. M. Majoor, "The perceptual relevance of scale-space image coding," *Signal Process.,* vol. 17, pp. 353–364, 1989.

[18] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.,* pp. 314–323, Feb. 1988.

[19] J. W. Woods and S. D. O'eil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 34, pp. 1278–1288, Oct. 1986.

[20] H. Gharavi and A. Tabatabai, "Subband coding of monochrome and color images," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 35, pp. 207–214, Feb. 1988.

[21] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image dependent quantization and post-quantization data compression," *Proc. IEEE Int. Conf., Acoust., Speech, Signal Process.,* vol. 3, pp. 1945–1948, 1989.

[22] J. T. Kim, H. J. Lee, and J. S. Choi, "Subband coding using human visual characteristics for image signals," *IEEE J. Select. Areas Commun.,* vol. 11, pp. 59–64, Jan. 1993.

[23] P. Moon and D. E. Spencer, "The visual effect of nonuniform surrounds," *J. Opt. Soc. Amer.,* vol. 35, pp. 233–248, Mar. 1945.

[24] A. K. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[25] E. C. Carterette and M. P. Friedman, Eds., *Handbook of Perception,* vol. 5. New York: Academic, 1975.

[26] J. Pandel, "Variable bit-rate image sequence coding with adaptive quantization," *Signal Processing: Image Commun.,* vol. 3, pp. 123–128, 1991.

[27] B. Girod, "Psychovisual aspects of image communication," *Signal Process.,* vol. 28, no. 3, pp. 239–251, Sept. 1992.

[28] B. Girod, H. Almer, L. Bengsson, B. Christensson, and P. Weiss, "A subjective evaluation of noise shaping quantization for adaptive intra interframe DPCM coding of color television signals," *IEEE Trans. Commun.,* vol. COM-36, no. 3, pp. 332–346, Mar. 1988.

[29] P. Pirsch, "Design of DPCM quantizers for video signals using subjective tests," *IEEE Trans. Commun.,* vol. COM-29, no. 7, pp. 996–1000, July 1981.

[30] R. Schafer, "Design of adaptive and nonadaptive quantizers using subjective criteria," *Signal Processing,* vol. 5, no. 4, pp. 333–345, July 1983.

[31] H. G. Musmann, "Predictive image coding," in *Image Transmission Techniques,* W. K. Pratt, Ed. New York: Acdemic, May 1979, pp. 81–97.

[32] G. Karlsson and M. Vetterli, "Subband coding of video signals for packet-switched networks," in *Proc. SPIE Conf. Visual Commun. and Image Process. II,* Oct. 1987, pp. 446–456.

[33] M. Nomura, T. Fujii, and N. Ohta, "Basic characteristics of variable rate video coding in ATM enviroment," *IEEE J. Select. Areas Commun.,* vol. 7, pp. 752–760, June 1989.

[34] M. Vetterli, "Multi-dimensional sub-band coding: Some theory and algorithms," *Signal Processing,* vol. 6, pp. 97–112, Apr. 1984.

[35] G. Karlsson and M. Vetterli, "Extension of finite length signals for subband coding," *Signal Processing,* vol. 17, pp. 161–166, June 1989.

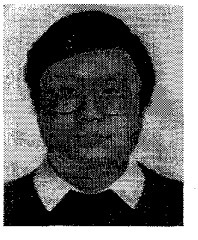
[36] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. 1980 ICASSP,* Apr. 1980, pp. 291–294.

[37] W. H. Chen and W. K. Pratt, "Scene adaptive coder," *IEEE Trans. Commun.,* vol. COM-32, pp. 225–232, Mar. 1984.

[38] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM,* pp. 31–43, Apr. 1991.

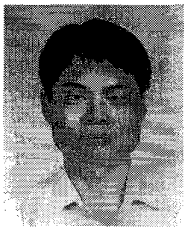
[39] J. O. Limb, "Distortion criterion of the human viewer," *IEEE Trans. Syst., Man, Cybern.,* vol. SMC-9, pp. 778–793, Dec. 1979.

[40] H. de Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments," in *Proc. SPIE Conf.,* vol. 1666, 1992, pp. 16–26.



Chun-Hsien Chou graduated from National Taipei Institute of Technology, Taipei, Taiwan in 1979, and received the M.S. and Ph.D. degrees in electrical engineering from National Tsing Hua University, Hsinchcu, Taiwan, in 1986 and 1990, respectively.

In 1990, he joined the Department of Electrical Engineering at Tatung Institute of Technology, Taipei, Taiwan, as an Assocciate Professor. During the academic year 1991–1992, he was a post-Doctoral research member with AT&T Bell Labs., Murray Hill, NJ. His current research areas include perceptual coding of image and video signals, very low bit rate video coding, and real-time VLSI architectures for multimedia communication.



Yun-Chin Li received the B.S and M.S. degrees in electrical engineering from Tatung Institute of Technology, Taipei, Taiwan, in 1991 and 1993, respectively.

He is currently working towards Ph.D. degree with the Department of Electrical Engineering, National Tsing Hua University, Hsinchcu, Taiwan. His main interests are low bit rate video coding and model-based image coding.