A Scene Adaptive and Signal Adaptive Quantization for Subband Image and Video Compression Using Wavelets

Jiebo Luo, *Member, IEEE*, Chang Wen Chen, *Member, IEEE*, Kevin J. Parker, *Fellow, IEEE*, and Thomas S. Huang, *Fellow, IEEE*

Abstract—Discrete wavelet transform (DWT) provides an advantageous framework of multiresolution space-frequency representation with promising applications in image processing. The challenge as well as the opportunity in wavelet-based compression is to exploit the characteristics of the subband coefficients with respect to both spectral and spatial localities. A common problem with many existing quantization methods is that the inherent image structures are severely distorted with coarse quantization. Observation shows that subband coefficients with the same magnitude generally do not have the same perceptual importance; this depends on whether or not they belong to clustered scene structures. We propose in this paper a novel *scene* adaptive and *signal* adaptive quantization scheme capable of exploiting both the spectral and spatial localization properties resulting from wavelet transform. The proposed quantization is implemented as a maximum *a posteriori* probability (MAP) estimation-based clustering process in which subband coefficients are quantized to their cluster means, subject to local spatial constraints. The intensity distribution of each cluster within a subband is modeled by an optimal Laplacian source to achieve the signal adaptivity, while spatial constraints are enforced by appropriate Gibbs random fields (GRF) to achieve the scene adaptivity. Consequently, with spatially isolated coefficients removed and clustered coefficients retained at the same time, the available bits are allocated to visually important scene structures so that the information loss is least perceptible. Furthermore, the reconstruction noise in the decompressed image can be suppressed using another GRF-based enhancement algorithm. Experimental results have shown the potentials of this quantization scheme for low bit-rate image and video compression.

Index Terms— Adaptive quantization, image and video compression, Gibbs random field, spatial constraints, subband coding, wavelet coding.

I. INTRODUCTION

THE rapid development of high performance computing and communication has opened up tremendous opportunities for various computer-based applications with image and video communication capability. However, the data required to represent the image and video signal in digital form would continue to overwhelm the capacity of many communication and storage systems. Therefore, a well designed compression element is often the most important component in such visual communication systems.

Over the years, various frameworks have been proposed to deal with image and video compression at different bit rates. JPEG is a discrete cosine transform (DCT)-based coding standard for still images [1]. Video coding standards, such as H.261 [2] and MPEG [3], are also DCT-based coding schemes with block-based motion estimation and motion compensation capabilities. However, at low bit rates, such block DCT-based standard coding schemes generally suffer from visually annoying "blocking effect" originated from the simple but unnatural rectangular block partition. They are also limited by the performance and the complexity of motion estimation and motion compensation in video coding. Therefore, an alternative coding scheme free of the "blocking artifact," and without or with less demanding motion estimation and compensation requirements for video coding, is desired at low bit rates. A multidimensional subband coding scheme has been proposed [4], [5] which generally employs no estimation or compensation of interframe motions. Subband or wavelet coding is especially advantageous at low bit rates when the "blocking artifacts" resulting from DCT-based coding or vector quantization have become noticeable and visually objectionable. Two-dimensional (2-D) subband image coding has been investigated with much success [6]. The application of three-dimensional (3-D) subband decomposition to video coding has also recently been attempted with initial success [5], [7], [8]. In addition, the architecture for real-time implementation of 3-D subband video coding has been proposed with comparable computational complexity as the motion-compensated DCT scheme and more complicated data storage and movement procedures [9].

Image and video coding schemes based on subband decomposition exploit the difference in perceptual response so that the compression strategies can be adjusted to an individual subband. Pyramid subband coding is equivalent to wavelet transform coding [10]. Wavelet transform coding resembles the human visual system (HVS) in that an image is decomposed

into multiscale representations. Moreover, wavelets have good localization properties both in space and frequency domains [11]. These two features provide excellent opportunities to incorporate the properties of the HVS and devise appropriate coding strategies to achieve high performance image and video compression. In general, for a target bit rate, higher compression ratio in high frequency subbands, where the distortion becomes less visible, allows the low frequency subbands to be coded with high fidelity. Although this is not unique to subband schemes, prioritized coding is limited in a DCT-based scheme because of the sole use of frequency representation. Decomposed subbands provide a joint space-frequency representation of the signal. Therefore, one can devise a coding scheme to take advantage of both the frequency and spatial characteristics of the subbands. In other words, one can determine the perceptual importance of the subband coefficients based on not only the frequency content, but also the spatial content, or scene structures. The combination of high compression ratio for perceptually insignificant coefficients and high fidelity for perceptually significant coefficients provides a promising alternative to high quality image and video coding at low bit rates.

For high frequency subbands, where the correlation has already been reduced by subband decomposition, various scalar and vector quantization schemes have been proposed, including: PCM (scalar quantization) [5], finite state scalar quantization [12], vector quantization [13], edge-based vector quantization technique [14], geometric vector quantization (GVQ) based on constrained sparse codebooks [8], and a scalar quantization that utilizes a local activity measure in the base band to predict the amplitude range of the pixels in the upper bands [15], etc. All these schemes have been proposed to take advantage of the characteristics of the high frequency subbands in order to increase the coding efficiency.

However, a common problem with many existing quantization methods is that the inherent image structures are severely distorted with coarse quantization. An apparent drawback of the conventional scalar quantization schemes is the inefficiency in approaching the entropy limit. Therefore, image fidelity cannot be properly maintained when the quantization becomes very coarse at low bit rates. Vector quantization (VQ), on the other hand, would generally achieve better coding efficiency. In general, VQ is performed by approximating the signal to be coded by a vector from a codebook generated from a set of training images based on minimizing the mean square error (MSE) [13]. In the case of GVQ, the structure and sparseness of the high frequency data is exploited by constraining the number of quantization levels for a given block size. The number of levels and block size determine the bit rate, and the levels and shape adapt for each block [8], [16], [17]. In general, the creation of a universal codebook for any image is impossible. The performance of vector quantization applied to a particular image largely depends on a codebook generated in advance and is not adaptive to a given signal. This inability of signal dependent adaptation will limit the exploitation of the individualized correlation in an arbitrarily given image. Moreover, to form vectors, rectangular block partitioning of images is usually adopted. At low bit rates, such block partitioning often destroys the inherent scene structure of a given image, since the approximation of a given block

by a vector from the codebook could alter the position, orientation, and the strength of the structures within the block, such as edge segments. As a result, at low bit rates, vector quantization often produces visible blocking artifacts which severely degrade the image quality. Moreover, the codebook generation and the searching against the codebook in vector quantization are usually computationally expensive. Some suboptimal implementations are often adopted in practice mainly to reduce the computational complexity [18]. These problems would adversely affect the coding performance.

Therefore, both conventional vector quantization and traditional scalar quantization schemes have very limited adaptivity with respect to the scene structure and the characteristics of an individual decomposed subband. It is also difficult for VQ schemes to accommodate the properties of the HVS, which is vital in evaluating the reconstructed images, in particular at low bit rates. In other words, these schemes have limitations in exploiting the unique spatial and spectral localities of wavelet decomposition as well as the psychovisual redundancies in the subbands are therefore not amenable to achieve high performance coding.

The proposed adaptive quantization with spatial constraints is intended to resolve the aforementioned problems. The incorporation of Gibbs random fields as spatial constraints in a clustering process enables the quantization to be both signal adaptive and scene adaptive. Such a quantization constitutes the major distinction of this scheme from the existing ones because it is designed to exploit both the spectral *and* spatial localities *simultaneously*. In this scheme, an adaptive clustering with spatial constraints is applied to the sparse and highly structured high-frequency bands to accomplish the quantization. The incorporation of localized spatial constraints is justified and facilitated by the existence of good spatial locality in the subbands decomposed using wavelets. Upon clustering, the representation of each pixel by its cluster mean is equivalent to a quantization process. However, such quantization enables us to preserve the important scene structures and eliminate most isolated nonprominent impulsive noises which have negligible perceptual significance. The compression ratio of these quantized high-frequency subbands can be greatly increased because the entropy has been reduced due to the smoother spatial distribution of each cluster within these subbands. In addition, the reconstructed images from these quantized high-frequency subbands can also be enhanced in the postprocessing stage using an enhancement algorithm based again on a Gibbs random field so that the reconstruction noise can be suppressed while the image details are well preserved.

We implement the clustering as a Bayesian estimation through optimal modeling of the intensity distributions and efficient enforcement of various spatial constraints in different subbands. We use the terminology of "scene adaptive" and "signal adaptive" to emphasize two different aspects of our algorithm. First, the signal adaptive property refers to the modeling and exploitation of the intensity distribution of the coefficients. This is accomplished by using a Laplacian model to model the intensity distribution of each cluster in the Bayesian estimation framework. Second, the scene adaptive property refers to the modeling and exploitation of the spatial redundancies in a given high frequency subband. This is
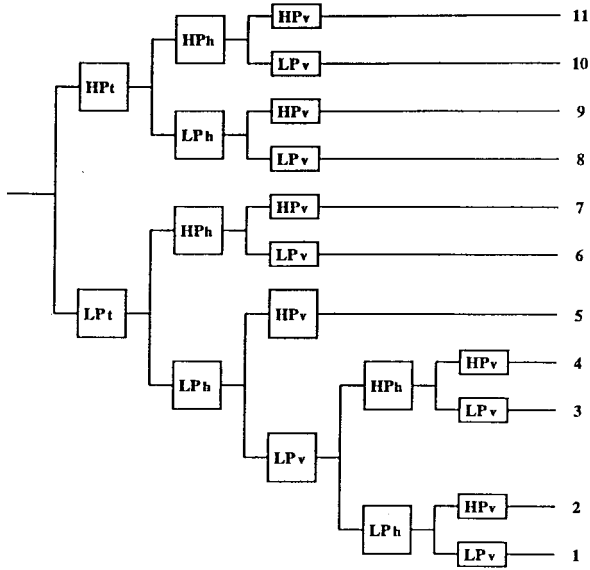
Fig. 1. The 11-band tree-structured decomposition for video signal.

Fig. 2. Template for displaying the 11-band decomposition scheme.

accomplished by using Gibbs random fields tuned according to the orientation and the resolution of each subband. The scene adaptivity and signal adaptivity are generally related to the exploitation of the psychovisual redundancies within the framework of wavelet decomposition.

This paper is organized as follows. Section II briefly describes the subband analysis and synthesis scheme for image and video coding. In particular, we discuss the spatio-temporal decomposition of video signals and the characteristics of each subband and the corresponding coding strategies. Section III introduces the adaptive quantization algorithm and its implementations. In particular, detailed discussions are devoted to optimal Laplacian modeling of the cluster distributions, the effective enforcement of various spatial constraints using Gibbs random fields (GRF), and an efficient noniterative implementation of the clustering-based quantization. In Section IV, we discuss issues beyond quantization, including an enhancement technique for the postprocessing of the reconstructed images from the quantized subbands. Experimental results are presented in Section V. Section VI concludes this paper with some discussions.

II. SUBBAND SCHEMES FOR IMAGE AND VIDEO CODING

Subband coding was initially developed for speech coding by Crochiere in 1976 [19] and has since proved to be a powerful technique for both speech and image compression. The extension of the subband coding to multidimensional signal processing was introduced in [4], and the application to image and video compression has been attempted with much success [6], [20], [21]. In image compression, the subband decomposition is accomplished by passing the image data through a bank of analysis filters. Since the bandwidth of each filtered subband image is reduced, they can be subsampled at its new Nyquist frequency, resulting in a series of reduced size subband images. These subbands are more tractable than the original signal in that each subband image may be coded separately, transmitted over the communication system, and decoded at the destination. These received subband images
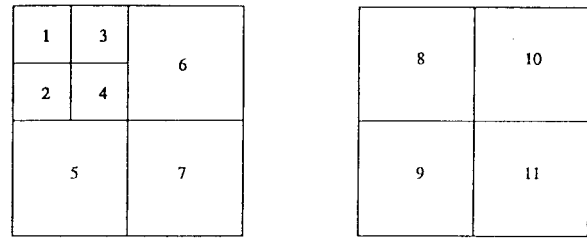
are then upsampled to form images of original size and passed through the corresponding bank of synthesis filters, where they are interpolated and added to obtain the reconstructed image.

Three-dimensional subband coding was originally proposed in [5] as a promising technique for video compression. It has shown comparable performance to other methods, such as transform coding and vector quantization. The video signal is decomposed into temporal and spatial frequency subbands using temporal and spatial analysis filterbanks. In the following, the procedure for spatio-temporal decomposition and reconstruction of video signal using the 3-D subband scheme is described. A 2-D scheme for image signal is a special case in that only spatial analysis and synthesis are involved. To recognize the difference between the temporal frequency response and the spatial frequency response of the HVS, the filterbanks used for temporal decomposition are often different from those for spatial decompositions. After subband decomposition, each subband would exhibit certain distinct features corresponding to the characteristics of the filterbanks. These features are utilized in the design of compression strategies in order to fully exploit the redundancy in the decomposed subbands.

A. Three-Dimensinal Subband Spatio-Temporal Decomposition

To minimize the computational burden of the temporal filtering in decomposing the video signal, temporal decomposition is based upon the two-tap Haar filterbank [5], [8], [17]. This also minimizes the number of frames that need to be stored and the delay caused by the analysis and synthesis procedures. The temporal decomposition results in two subbands: the highpass temporal (HPT) band, i.e., frame difference (FD), and the lowpass temporal (LPT) band.

In the case of spatial analysis and synthesis, longer length filters can be applied since these filters can be operated in parallel and the storage requirements are not affected by the filter length. Therefore, spatial decomposition, both horizontal and vertical, is often based on multitap filterbanks. With separable filters, multidimensional analysis and synthesis can be carried out in stages of directional filtering. To achieve high compression, the lowest frequency band can be further decomposed in a tree structure fashion. The high frequency subbands contain structures approximately aligned along horizontal, vertical, or diagonal direction. Fig. 1 shows an 11-band tree-structured decomposition scheme for video signals. The template for displaying the decomposed 11-band subband images is shown in Fig. 2.

In this research, wavelet filterbanks, namely, the Daubechies' 9/7 biorthogonal wavelets [13], are employed

to decompose and reconstruct the signal. The regularity and orthogonality of the wavelet filterbanks ensure the reconstruction of image and video signals with high perceptual quality. Moreover, it has been shown [10], [13], [22] that the wavelet transform corresponds well to the human psychovisual mechanism because of its localization characteristics in both space and frequency domains. Note that the choice of wavelets also corresponds well to the proposed quantization scheme. First, the good localization of wavelet decomposition in frequency domain offers good frequency separation that facilitates efficient compression. Second, and more important, the good localization of wavelet decomposition in spatial domain justifies and facilitates the incorporation of spatial constraints in the quantization. Appropriate spatial constraints can then be efficiently enforced to identify and preserve perceptually important components in the process of quantization.

B. Characteristics of Subbands and Corresponding Coding Strategy

After the spatio-temporal decomposition, the resultant sub-bands exhibit quite different characteristics from one to another and have different perceptual responses. The quantization strategies need to be designed to suit individualized subbands to achieve optimal representation with minimum visible distortions. For the 11-band decomposition of the video signal, the characteristics of each band are summarized in the following. For the decomposition of 2-D still images, similar characteristics exist.

1) Band 1 is a low resolution representation of the original image and has similar histogram characteristics, but with much smoother spatial distribution. It can be efficiently coded using DPCM.
2) Bands 2–7 contain spatial high frequency components of the LPT band. They consist different amount of "edges" and "impulses" corresponding to different directions and resolution levels. Since the signal power and the perceptual importance in general decreases as the resolution level increases, the bit allocation should be adjusted accordingly.
3) Band 8 is the low pass spatial band of the HPT band and contains most motion energy. It needs more bits or finer quantization when the motion activity is high.
4) Bands 9–11 represent spatial high-frequency components of the FD. They usually contain very low signal energy and are of low perceptual sensitivity.

The quantization and coding algorithm should be developed based on these characteristics. In general, the strategies are summarized as follows. Subbands at the lower resolution levels (with smaller index in Fig. 2) contain most signal energy and are of higher visual significance. They require higher quality coding and hence, finer quantization. Subbands at the higher levels are quantized coarsely or may be discarded.

III. ADAPTIVE QUANTIZATION OF HIGH-FREQUENCY SUBBANDS

Many attempts in low bit-rate subband coding have been concentrated in the study of characteristics of the high frequency subbands so that the features of these subbands can

be incorporated in the design of coding algorithms [8], [14], [15], [16], [22]. One characteristic of the high frequency subbands is their less significant perceptual responses. They can often afford coarse representations that would result in fewer bits needed to code the image without introducing much visible distortion in the reconstructed images. Another important characteristic of the high-frequency subbands is the spatial structures in these subbands. These structures appear as sparse "edges" and "impulses" that correspond mainly to a few strong intensity discontinuities in temporal or spatial domains. In general, strong and clustered "edges" and "impulses" are of significant visual importance and need to be preserved in the quantization. On the other hand, there are some nonstructural weak impulses corresponding to the noise that has much less visual importance but would need considerable amount of bits to code. Removal of the noise would lead to significant coding gain with perceptually negligible distortion in the reconstructed image. In addition, these sparse "edges" and "impulses" exhibit well defined directional arrangement in accordance with the filtering direction in the subband analysis.

To achieve the desired simultaneous scene adaptivity and signal adaptivity, we propose a novel quantization scheme for high frequency subbands based on the concept of adaptive clustering with spatial constraints. This scheme utilizes Gibbs random fields to enforce neighborhood constraints in order to remove those isolated "impulses" and weak local variations whose contributions to the reconstruction are negligible. The smoothing of the perceptually insignificant pixels is accomplished in a scale-dependent way similar to the perception of the HVS. As visual psychophysics states, the HVS is sensitive to not only the frequency contents, but also several spatially localized characteristics, including the background luminance and contrast, the proximity to edges, texture masking, and scale [22]–[24]. As will be shown, the entropy of the subband images after the proposed adaptive quantization is reduced without significant perceptual distortions in the reconstructed images. It is the principle of scene adaptive and signal adaptive quantization as the result of the exploitation of the HVS, and the spatial and spectral localities of wavelet transform, that constitutes the fundamental difference between this quantization scheme and the existing ones.

A. Adaptive Quantization Algorithm

The proposed adaptive quantization of high frequency subbands is accomplished through an adaptive clustering process. In this clustering-based quantization, each pixel is quantized to its cluster mean according to its intensity and its neighborhood constraints modeled by a Gibbs random field. Such a clustering process results in an adaptive quantization in two aspects. First, the quantization is *signal adaptive* since the number of quantization levels needed and the value of these quantization levels are determined according to the statistical characteristics and the perceptual frequency response of each subband. Second, through enforcing spatial constraints, isolated pixels or pixels representing local noisy variations are quantized to the mean of the cluster to whom majority of their neighbors belong and therefore, are absorbed by the neighborhood. With such a constrained clustering, the spatial distribution of the subband, especially the noisy background, becomes rather smooth. However, the prominent structures and details with

significant perceptual importance are preserved mimicking the HVS perception. In the perceptual literature, the Gestalt psychologists of the 1920's and 1930's investigated questions of how the human visual system groups together simple visual patterns. More recently in computer vision literature [25], these Gestalt investigations have inspired work in perceptual grouping, an area championed by Lowe [26] and Witkin and Tenenbaum [27]. In particular, Lowe [26] defines perceptual grouping as a basic capability of the human visual system to derive relevant grouping and spatial structures from an image *without* prior knowledge of its contents. As expected and will be shown later, the adaptive quantization is able to group together the subband coefficients likely to have come from intrinsic objects in the original scene, without requiring specific object models [28]. The quantization depends on the local scene structure and is therefore *scene adaptive*. Upon the completion of such an adaptive clustering and quantization, the highpass subbands contain mainly refined "edges" or "clumps" over a much cleaned background. Since the "noise" is largely removed and the "edges" are redefined using only a few levels, the images are significantly less busy with greatly reduced entropy.

We have tailored the clustering algorithm proposed in [29] and [30] to develop an enhanced adaptive clustering algorithm. It has been shown in [29] and [31]–[33] that images can be modeled by a Gibbs random field and image clustering can be accomplished through a maximum *a posteriori* probability (MAP) estimation. Using Bayes' theorem and the log likelihood function, the Bayesian estimation that yields MAP of the clustering x given the image y can be expressed as

$$\hat{x} = \arg\max_x p(x \mid y)$$
$$= \arg\max_x \{\log p(y \mid x) + \log p(x)\} \qquad (1)$$

where $p(x)$ is the *a priori* probability of the clustering x, and $p(y \mid x)$ represents the conditional probability of the image data y given the clustering x. There are two components in the overall probability function. The conditional probability corresponds to the adaptive capability that forces the clustering to be consistent with intensity distribution of the corresponding cluster. The prior probability corresponds to the spatial smoothness constraints which will be characterized by a Gibbs random field. There are several distinctions between our adaptive quantization algorithm and the GRF-based clustering algorithms in [29] and [30]. First, we have different models for the *a priori* probability $p(x)$. The Gibbs random field, i.e., the parameter β, is adjusted according to the orientation and the resolution of each subband in our algorithm. Second, we have a different model for the conditional probability $p(y \mid x)$. We use a Laplacian model, as opposed to a Gaussian model, to model the intensity distribution of each cluster. Finally, we develop a noniterative implementation suitable for quantization, where the means are not obtained iteratively, but obtained in advance using a Lloyd–Max quantizer. These aspects of differences will be elaborated in detail in the following.

1) Modeling of Spatial Constraints: Gibbs random fields, the practical equivalences of Markov random fields, have been widely used to represent various types of spatial dependency in images [29], [31]. A Gibbs random field can be characterized by a neighborhood system and a Gibbs potential function. A

Gibbs distribution can then be defined as

$$p(x) \propto \exp\left\{-\sum_c V_c(x)\right\} \qquad (2)$$

where $V_c(x)$ is called the clique potential. Associated with the neighborhood system are cliques and their potentials. A clique c is a set of sites where all elements are neighbors [34]. In this study, we consider that a 2-D image is defined on the Cartesian grid, and the neighborhood of a pixel consists of its four nearest pixels.

Image clustering constrained by Gibbs random fields is accomplished by assigning labels to each pixel in the given image according to its own intensity value and the properties of its neighbors. A label $x_s = i$ indicates that the pixel s belongs to the ith class of the K classes. According to the essential property of a Markov random field, the conditional probability $p(y \mid x)$, and thus the clustering, depends only on the local neighborhood constraints. A two-point clique potential function suitable for clustering can be defined as

$$V_c(x) = \begin{cases} -\beta, & \text{if } x_s = x_t \text{ and } s, t \in c \\ +\beta, & \text{if } x_s \neq x_t \text{ and } s, t \in c. \end{cases} \qquad (3)$$

Note that the maximization of the overall posterior probability implies the pursuit of the lowest potential state. Therefore, by penalizing inhomogeneous clustering with positive potential β and rewarding homogeneous clustering with negative potential $-\beta$ within local neighborhoods, this potential function can be used to enforce desired spatial constraints to achieve homogeneous clustering if an appropriate neighborhood system c and a proper parameter β are selected.

We have developed four types of clique for the parameterization of the Gibbs random field according to the characteristics of the high frequency subbands. In Fig. 4, the solid lines indicate strong connections with large β along the nonpreferential directions to enforce strong smoothness constraints, while the dashed lines represent weak constraints with small β along the preferential directions. The preferential direction of a subband is defined as the direction along which the structures are aligned, and is perpendicular to the filtering direction. Within each subband, the image details along the dashed line direction can be preserved and the smoothing is done mainly along the nonpreferential direction. The cliques shown in Fig. 4, from left to right, are suitable for the horizontal, vertical, diagonal high frequency subbands, and the lowest frequency subband, respectively. Note that the proposed adaptive quantization may not be suitable for the lowest frequency subband in the cases where the letter which needs to be coded with high fidelity to ensure overall high quality reconstruction. In this band of the lowest resolution, each pixel corresponds to manifold pixels in the original image and small quantization error will be magnified in the reconstruction. Only when it is necessary to apply the adaptive quantization to the baseband at a very low bit rate, relatively weaker spatial constraints can be enforced using a normal clique as illustrated in Fig. 4. Furthermore, β can be adjusted to the resolution level of each subband to reflect different neighborhood constraints on the grid at different scales. In general, larger β is used for the subbands at the higher resolution levels in accordance with the increase in resolution and scale. For example, β can be doubled every time moving to the next higher resolution
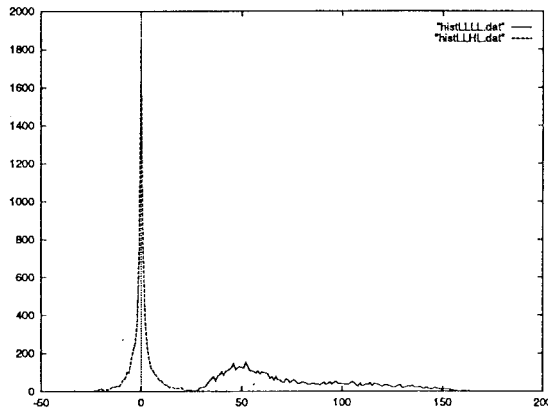
Fig. 3.   Typical histograms of the subbands (_____ the lowpass band, - - - a highpass band). The horizontal axis is the intensity axis, and the vertical axis is the histogram count axis.
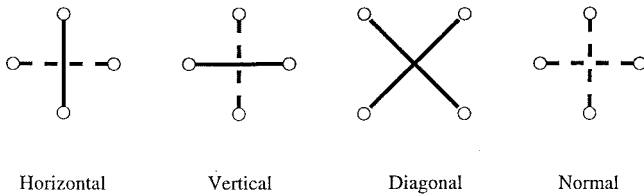

Horizontal   Vertical   Diagonal   Normal

Fig. 4.   Cliques for subbands with different preferential directions.

level. β can also be related to bit allocation in progressive coding in that larger β is used for the subbands on higher levels to reduce the bit stream when bits are running out. Such flexible parameterization of the Gibbs random field allows us to preserve the most significant structures in a given subband under the bit rates constraints.

2) Modeling of the Cluster Intensity Distribution: It has been shown that the overall distribution of a high frequency subband, as shown in Fig. 3, can be optimally modeled by a Laplacian with zero mean. Such modeling yields the best coding performance under optimal bit allocation [10]. Within each high frequency subband, nonzero coefficients are basically clustered into "edges," i.e., oscillating positive or negative "strips" over the fairly uniform zero background, or appear as isolated "impulses." For a quantization scheme that is scene adaptive, it needs to preserve those critical positive, negative, and zero values which are of perpetual significance in the reconstruction. PCM was first introduced to quantize these subbands and a "dead zone" technique [35] was proposed to suppress visually insignificant noise around zero by setting a relatively larger quantization interval around zero. This technique allows finer quantization of the tails of the Laplacian distribution because the pixels of larger amplitude are often of greater visual importance [5], [8], [35]. However, the noise suppression using this technique is limited to smoothing only the noise close to the zero background and leaves noises in the rest of the range of the intensity distribution unaffected.

There are several possible models for the individual intensity distribution $p(y \mid x)$ of each cluster x in (1), including Gaussian, generalized Gaussian, and Laplacian probability density functions (PDF's). In the case of clustering [30], the conditional density is typically modeled as a Gaussian process

with mean μ and variance σ

$$p(y \mid x) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\sum_s \frac{1}{2(\sigma)^2}(y-\mu)^2\right\}. \quad (4)$$

With a Gaussian model, we can derive the overall probability density as

$$P(x \mid y) \propto \sum_s \left\{\ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2(\sigma)^2}(y-\mu)^2\right\} - \sum_c V_c(x). \quad (5)$$

However, for a clustering-based quantization, such an assumption would not lead to an optimal modeling. It is natural to model the individual cluster conditional density as a Laplacian process considering that the overall intensity distribution can be optimally approximated by a Laplacian source. For a given cluster, if we assume

$$p(y \mid x) \propto \frac{1}{\sqrt{2}\sigma} \exp\left\{-\sum_s \frac{\sqrt{2}}{\sigma}|y-\mu|\right\} \quad (6)$$

then the overall probability density becomes

$$p(x \mid y) \propto \sum_s \left\{\ln \frac{1}{\sqrt{2}\sigma} - \frac{\sqrt{2}}{\sigma}|y-\mu|\right\} - \sum_c V_c(x). \quad (7)$$

To examine the validity of the modeling of the cluster distribution $p(y \mid x)$, we construct the overall distribution based on cluster distributions such that the actual distribution of the coefficients in a given subband is modeled as the superposition of individual cluster distributions whose statistical parameters are obtained from the optimal clustering. As clearly shown in Fig. 5, the superposition of multiple Gaussian distributions is unable to yield a satisfactory approximation to the overall histogram. Not only can individual Gaussian modes be identified, but the characteristics of the distribution, e.g., first-order and second-order derivatives of the distribution, are also quite different. This is due to the fact that the exponent term in a Gaussian distribution is quadratic, while in a Laplacian distribution it is essentially linear. On the other hand, the composite distribution of multiple Laplacian distributions is very consistent with the overall Laplacian distribution, especially in the tail parts where perceptually important information usually resides. Goodness-of-fit tests can also show the superiority of the multimodal Laplacian modeling of the cluster distribution with less fitting error [36]. In the case of clustering-based adaptive quantization, as a result of the optimal modeling of the cluster distribution, we are able to obtain optimal quantization and therefore achieve optimal reconstruction from the quantized high-frequency subbands. Comparison of the quantized subbands using different modeling is given in Fig. 12.

Note that the reconstruction levels are global for the entire subband, and therefore, only labels need to be coded and transmitted. However, this quantization scheme is indeed *adaptive* for the following reasons. The local neighborhood of each pixel site s changes from one location to another, therefore, two coefficients with the same intensity value are not necessarily quantized to the same level. Depending on the quantization (or clustering) of the local neighboring coefficients, a coefficient is quantized according to a local Bayesian estimation based on 1) its own intensity value, 2)
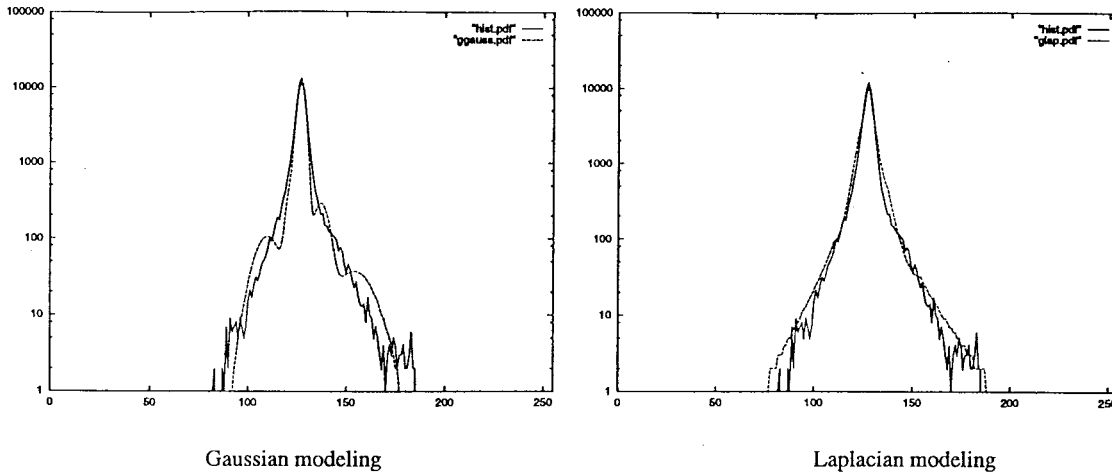
Fig. 5. Modeling of the intensity distribution in high frequency subbands (blowup in log scale). The horizontal axis is the intensity axis (shifted by 127), and the vertical axis is the histogram count axis.

its neighboring coefficients, and 3) the orientation and the resolution of the subband it belongs to, as if spatially adaptive "local quantization tables" were utilized. Virtually, there exits a "local quantization table" according to the local spatial configuration of each site. Therefore, using just one set of cluster means, we are able to achieve a spatially adaptive quantization under the framework of Bayesian estimation.

With the GRF-based spatial constraints, how a pixel is quantized is not only determined by its intensity, but also by its neighborhood spatial constraints. It is noteworthy that this adaptive quantization scheme cannot be achieved by the combination of scalar quantization and noise filtering, such as median filtering. Seemingly, median filtering can be used to remove impulsive noise while preserving edges. However, median filtering is appropriate for normal images containing regions. It is the existence of regions that generates the necessary majority votes so that edges of the region can be preserved. The subband images are essentially composed of thin "edges" and isolated "impulses," with literally *no* regions, over the zero background. While median filtering can remove "impulses," it would also remove those thin and long structures, such as meandering edge segments. A prominent spike of large amplitude would also be removed by median filtering, but it can be preserved by the adaptive quantization because the first energy term in (7) would be large enough so that it is not absorbed by the neighborhood. Furthermore, the spatial information within a median filtering window is not preserved in the median filtering. The reason being, median filters (and other order-statistics filters) seek to obtain only one good representative among the N neighboring pixels, and this median can be any of the N values. Therefore, the spatial localization of the thin edges can be altered, though within a local window, during the median filtering process.

The incorporation of a Gibbs random field in the MAP estimation allows us to achieve a similar but better "dead zone" effect, originally proposed in [35]. Unlike the original approach which generates "dead zone" simply by intensity thresholding, we achieved an improved "dead zone" which suppresses noises according to both the intensity and local spatial constraints. Moreover, the adaptive quantization is

capable of suppressing noise in the entire range of the intensity distribution, instead of being limited to the zone around zero. As illustrated in Fig. 6, without spatial constraint, the partition of clusters is such that the zones of clusters are separated. With the incorporation of the spatial constraints, the zones of clusters are actually overlapped with each other. This overlapped partition allows us to achieve an overlapped quantization, which is fundamentally different from all existing quantization schemes. Therefore, it enables the suppression of noise in the entire range of the distribution. The actual quantization intervals are essentially enlarged, not just for the central zone around zero, but for all the quantization intervals.

3) Implementations: The original adaptive K-mean clustering based on (7) can be implemented using a local optimization technique called ICM [37]. The ICM is efficient to enforce local spatial constraints [38]. At first, an initial clustering x is obtained through the simple K-mean algorithm. In this study, an odd K is chosen for the total number of levels for each subband since the histograms of the high frequency subbands are approximately symmetric around zero. K can be assigned according to the perceptual importance of each subband, i.e., the characteristics of the HVS [22], and the principle of optimal bit allocation. The subbands of lower resolution often have larger dynamic range. Therefore, they are assigned more levels since they carry more perceptually important information. Then, the overall probability function is maximized in a site-by-site fashion, with the mean μ and the variance σ of each cluster being updated after each iteration. The optimization is accomplished through alternating between the MAP estimation of the clustering and the iterative update of the cluster means and variances. Such a alternating process is repeated until no pixels change classes. The result is the adaptive clustering of the given high-frequency subband. Finally, the quantized subband is obtained by replacing each pixel with its cluster mean.

There are still some problems with the ICM implementation of this clustering-based adaptive quantization. First, the intensity distribution and the spatial constraints are coupled in an iterative process in the ICM process. Even a small parameter β can impose very strong constraints of the Gibbs random field
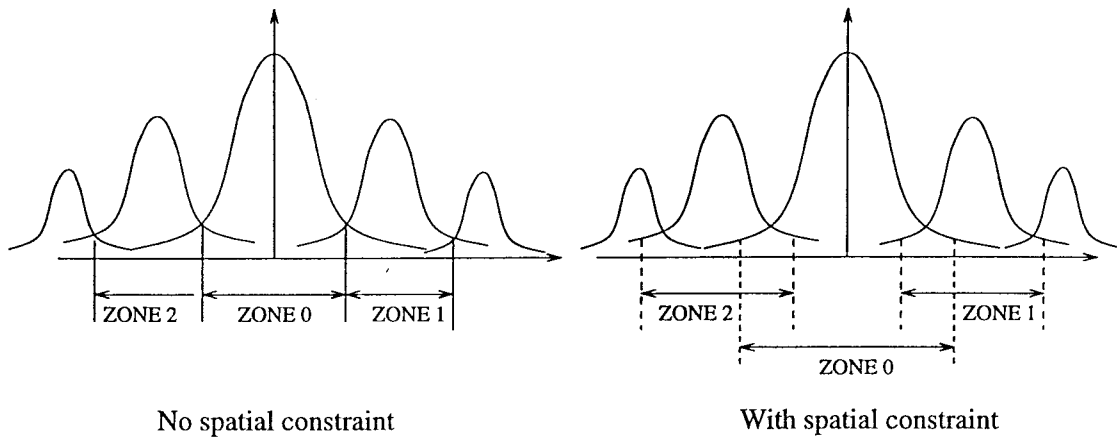
Fig. 6. Dead zone effect.

over large distances through clique interactions in successive iterative processes. Therefore, some edge enhancing effect can occur, which is not desired in the case of quantization if image fidelity is the concern. Second, the iterative implementation is still considered time-consuming although the ICM is one of the computationally least expensive optimization techniques [29]. In the case of video communication where large amounts of subbands are generated in the spatio-temporal decomposition, it cannot afford an expensive computation since real-time processing is often required.

For the clustering-based adaptive quantization, we developed a two-step noniterative implementation. At first, a Lloyd–Max scalar quantizer is found whose optimal reconstruction levels are used as the means of clusters. MAP estimation of the clustering is then accomplished in virtually one iteration because the cluster means and variances have been predetermined. The spatial constraints are only used to eliminate those nonprominent impulsive pixels while preserving the important structures. In our experiments, the cluster means (i.e., the reconstruction levels in quantization) obtained using iterative implementation turned out to be very close to those obtained using a Lloyd–Max quantizer. This observation is not surprising because both implementations optimize similar objective functions. However, the noniterative implementation not only is computationally efficient, but more importantly, produces better reconstruction results because the local spatial constraints are more appropriately enforced.

IV. BEYOND QUANTIZATION

A. Coding of the Quantized High Frequency Subbands

Coding of an image generally includes two distinct operations: quantization and symbol coding. The adaptive quantization with spatial constraints is capable of removing the "noise" of low perceptual significance, which would otherwise need considerable bits to code. The quantized high frequency subbands are then coded by a symbol coder, which generally includes an entropy coder. With the reduction of entropy upon the adaptive quantization, a lower bit rate is expected from the entropy coding. The entropy coder consists of a variable word length coder to code the labels of the nonzero values

of the clustered subbands and a runlength coder to code their corresponding locations [20]. Different scanning schemes can be used for individual subband to increase the runlength since these clustered high-frequency subbands are composed of well defined "edges" whose directions correspond to the direction of the highpass filtering used to obtain the decomposition. Because of the smoother background in the quantized subbands, a Hilbert–Peano scan [39] can also be very effective. Another scheme of increasing the runlength is to partition the subbands into nonoverlapping blocks [35]. Through such partitioning, local area of zero values can be better exploited to improve the runlength coding efficiency.

In our experiments, we will use the directional scan schemes followed by a runlength coding. The horizontal and vertical subbands are scanned accordingly. We also use horizontal scan for diagonal subbands for simplicity since we found that the gain margin is rather small by using a diagonal zigzag scan. Recently, zerotree-based coding algorithms have achieved great success in wavelet-based coding [40], [41] due to the efficient symbol coding techniques which exploit the intrinsic parent-descendent dependencies in the wavelet decomposition. We will adopt the zerotree coding technique proposed in [40] to code the quantization level indexes in the experiments where more levels of wavelet decomposition are selected. Upon the adaptive quantization, the subband contains very few "clustered edges," which consist of nonzero coefficients, over a very clean zero background. Therefore, the zerotree coding can be very efficient.

B. Enhancement Algorithm

There are some artifacts in the reconstructed image due to the quantization in different frequency subbands. These artifacts generally appear as ringing effect around sharp edges, loss of fine details, and blotchiness in the slowly-varying regions. While the loss of fine details is difficult to recover, the other typical artifacts in wavelet-based coding are not as visually annoying as the blocking effect, and some of them can be removed or reduced. A Gibbs random field is again applicable as a spatial constraint to remove these artifacts and enhance the reconstructed image. The enhancement is also formed as a MAP estimation.
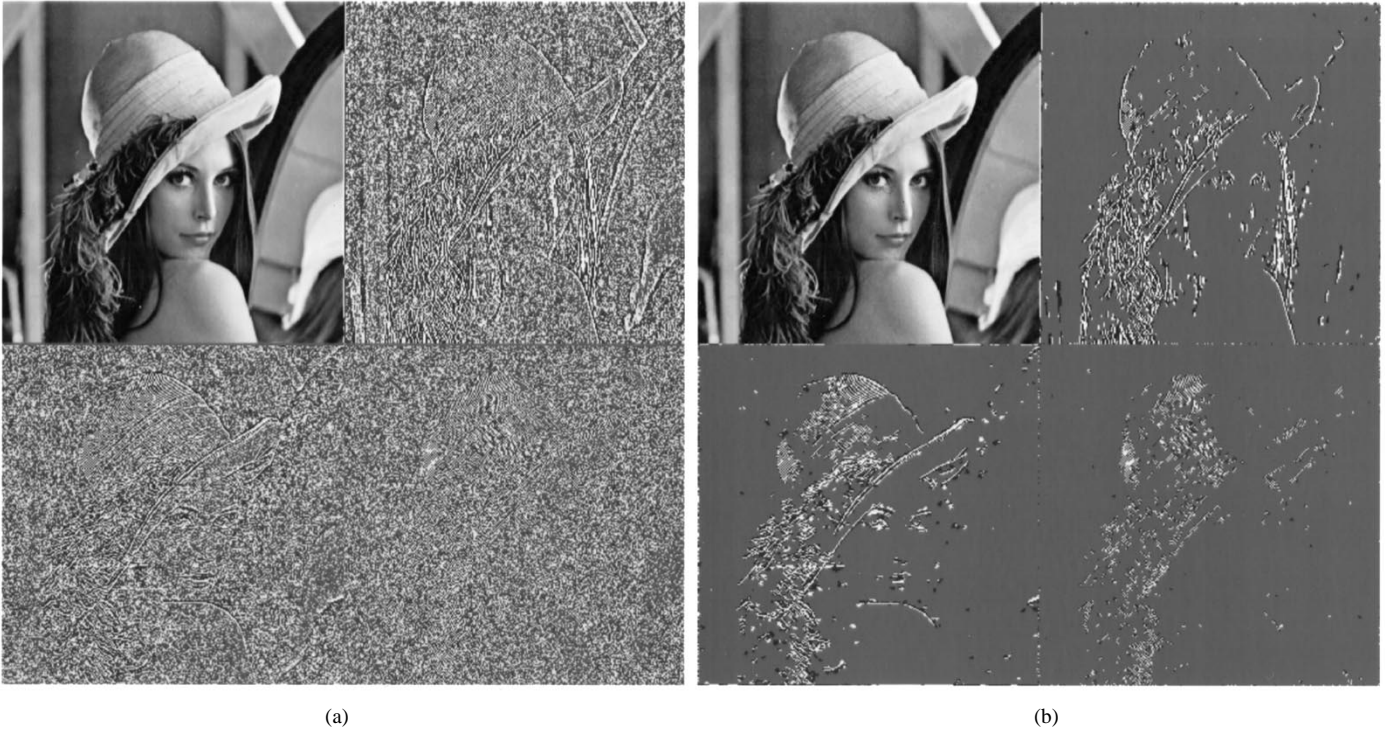
Fig. 7. A four-band decomposition of the "Lena" image: (a) original subbands and (b) quantized high frequency subbands.

The conditional probability of the quantization y given the original data x can be written as

$$p(y \mid x) = \begin{cases} 1, & y = \mathbf{Q}[\mathbf{W}(x)] \\ 0, & y \neq \mathbf{Q}[\mathbf{W}(x)] \end{cases} \quad (8)$$

where $\mathbf{Q}$ stands for the quantization and $\mathbf{W}$ denotes the wavelet transform. The conditional probability states that the estimated image should conform to the quantized data. This constraint can be enforced by projecting the estimated image back to the transform domain, i.e., decomposing the image in the same way as before, and adjusting the pixels so that the same quantized subband image is maintained.

We use a specific Gibbs random field, the Huber–Markov random field model to model the *a priori* probability. Its potential function $V_{c,T}(x)$ is in the form of (9). The Huber minimax function has been successfully applied to the removal of block effect in low bit rate transform coding [42], [43]. It can be written as

$$V_{c,T}(x) = \begin{cases} x^2, & |x| \leq T \\ T^2 + 2T(|x| - T), & |x| > T. \end{cases} \quad (9)$$

The desirable property of this minimax function is its ability to smooth the artifacts in the image while still preserving the image detail, such as edges and regions of textures. If we define the gray level differences between the current pixel $x_{m,n}$ and the pixels within its neighborhood $N_{m,n}$ as

$$\{x_{m,n} - x_{k,l}\}_{k,l \in N_{m,n}}, \quad 1 \leq m, \quad n \leq N \quad (10)$$

then these differences can be used as the argument of the Huber minimax function. The quadratic segment of the minimax

function imposes least mean square smoothing of the artifacts when the local variation is below T. On the other hand, the linear segment of the Huber minimax function enables the preservation of the image detail by allowing large discontinuities in the image with a much lighter penalty. The overall enhanced image is given by

$$\hat{x} = \arg\min_{x \in \mathcal{X}} \sum_{k,l \in N_{m,n}} V_{c,T}(x_{m,n} - x_{k,l}), \quad 1 \leq m, \quad n \leq N. \quad (11)$$

Since the projection to the constraint space $\mathcal{X} = \{x : y = \mathbf{Q}[Hx]\}$ requires a full cycle of subband analysis and synthesis, a suboptimal solution with the least computation would be the unconstrained noniterative estimation of (11). The Huber–Markov random field model also results in very low computational complexity. To compute the derivative of the function for performing local gradient-descent in an ICM-like scheme, only linear operations are involved.

V. EXPERIMENTAL RESULTS

Experimental results have been obtained using the test image "Lena" and the test video sequence "Salesman." As discussed previously, the temporal filterbank is the two-tap Haar filterbank. Daubechies' wavelet 9/7 biorthogonal filterbank [10] is selected for the spatial analysis and synthesis. The decomposition, quantization, reconstruction, and enhancement of "Lena" and a typical frame of "Salesman" sequence are shown in Figs. 7–13. To examine the quantization results, the quantized subbands are displayed with midgray cluster corresponding to the zero value, darker clusters to the negative values, and brighter clusters to the positive values, as shown in Figs. 7 and 11. The spatial distribution of the quantized

Fig. 8. Reconstruction of "Lena" using the EZW algorithm [40]: (a) the original "official" "Lena" image and (b) the reconstructed image.

Fig. 9. Reconstruction of "Lena" with the adaptive quantization and the EZW algorithm: (a) the reconstructed image and (b) the enhanced image.

subband is made much smoother because of the incorporation of spatial constraints. Using the adaptive quantization, we remove those perceptually negligible noisy contents and only preserve those visually important components in the high frequency subbands (see Fig. 7). To boost the contrast and emphasize the effect of the adaptive quantization for display purpose, histogram equalization has been performed on those subband images. The numerical results on entropy reduction are presented in Tables I and II.

In terms of the modeling of the intensity distribution, multiple Laplacian modeling is able to produce the most coherent quantization. In terms of the implementation, the noniterative
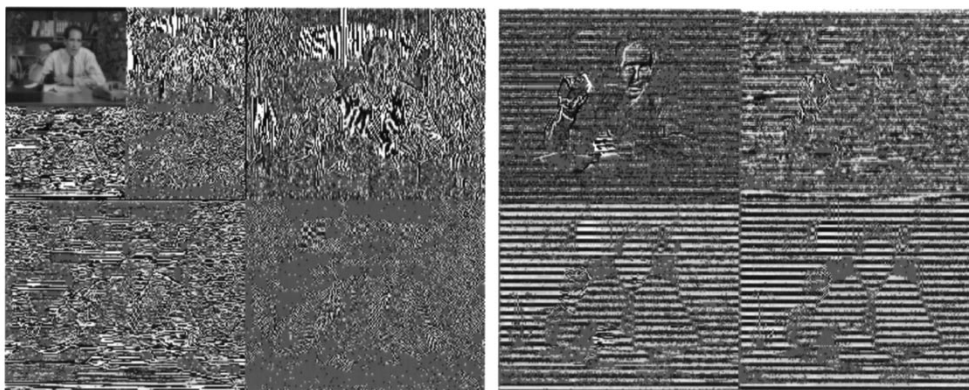
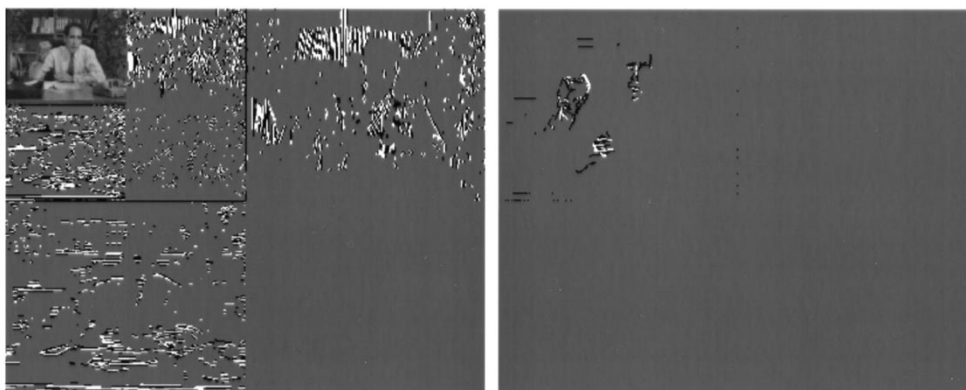Fig. 10. An 11-band decomposition of LPT band of a typical "Salesman" frame.

Fig. 11. The quantized subbands of the "Salesman" frames.

quantization enforces the spatial constraints more rigorously than the quantization through the ICM. This may need some explanation. It is clear that the interaction between the spatial constraints and the image force is desired for a image segmentation problem. However, in the adaptive quantization, we consider that both the elimination of nonstructural coefficients and the preservation of original image scene structures are important. We found that the adaptive quantization through the ICM sometimes altered the original image structure or created some nonexistent structures. The reason is that the high-frequency subbands contain mostly thin *edge-like structures* rather than *regions* contained in normal images. Overall, the noniterative implementation (NICM) with Laplacian modeling outperforms the other two combinations. This is clear from Fig. 12(d), where the adaptive quantization eliminates the noises without altering those important scene structures. Such a combination also yields the highest PSNR as is shown in Table I. In Table II, the entropy of the "Salesman" subband images before and after the quantization shows significant entropy reduction in the high-frequency bands. Because we use directional scan and runlength coding, the entropy of these high-frequency subbands is calculated using the first-order entropy of appropriate direction instead of the zero-order entropy, which is solely based on the histogram. In general, the first-order entropy is smaller than the zero-order entropy. Some insignificant subbands are discarded and therefore, are not listed in this table. The entropy of the lowest frequency

subband is obtained using DPCM and is included in the overall entropy.

To demonstrate the effectiveness of the proposed adaptive quantization, Shapiro's state-of-the-art embedded zerotree wavelet (EZW) coding algorithm [40] is adopted. Compression results are obtained using six-level wavelet decomposition for the following cases: 1) using the original EZW algorithm and 2) cascading the adaptive quantization with the EZW algorithm. The comparison is done using the 512×512 "official" "Lena" image [40]. Higher peak signal-to-noise ratio (PSNR) and better visual quality is obtained at a low bit rate of 0.25 b/p by combining the proposed quantization with the zerotree coding. Noticeably, the rim of the hat, the shoulder, and the face are reproduced much better in Fig. 9 than in Fig. 8. The reason being, the available bits are concentrated on the scene structures in high-frequency subbands that correspond to these important edges in the original image. If the enhancement technique is applied, visually aesthetic reconstruction can be produced with slight PSNR improvements. The remaining minor ringing artifacts and blotchiness are completely removed while the image details are preserved. The PSNR of the reconstructed image in Fig. 9 obtained using the proposed adaptive quantization is 33.52 dB, compared to 33.17 dB by the EZW reported in Fig. 8. The final PSNR after the enhancement is slightly higher at 33.91 dB because the improvements occur at only a small portion of the image pixels, such as around sharp edges. However, corresponding
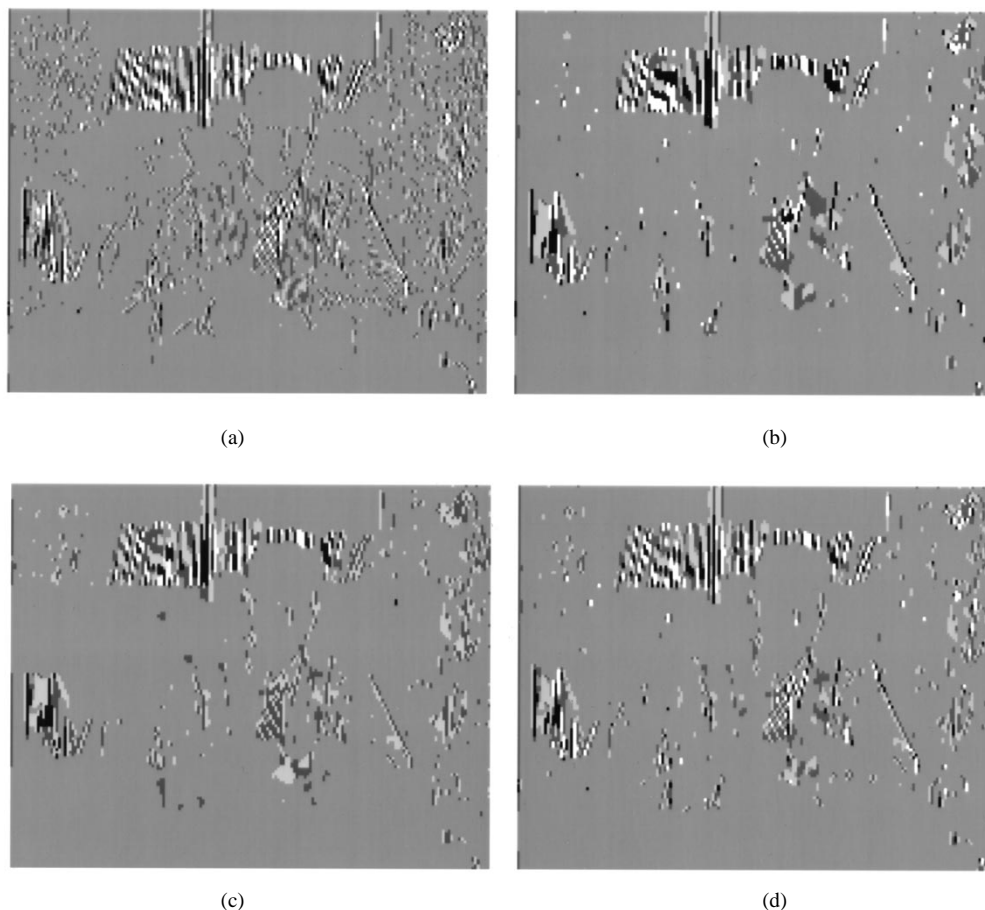
Fig. 12.   Quantization of a high-frequency subband (blowup). (a) Lloyd–Max quantizer without spatial constraints, (b) adaptive quantization with Gaussian modeling, (c) adaptive quantization with Gaussian modeling and ICM, and (d) adaptive quantization with Laplacian modeling and NICM.
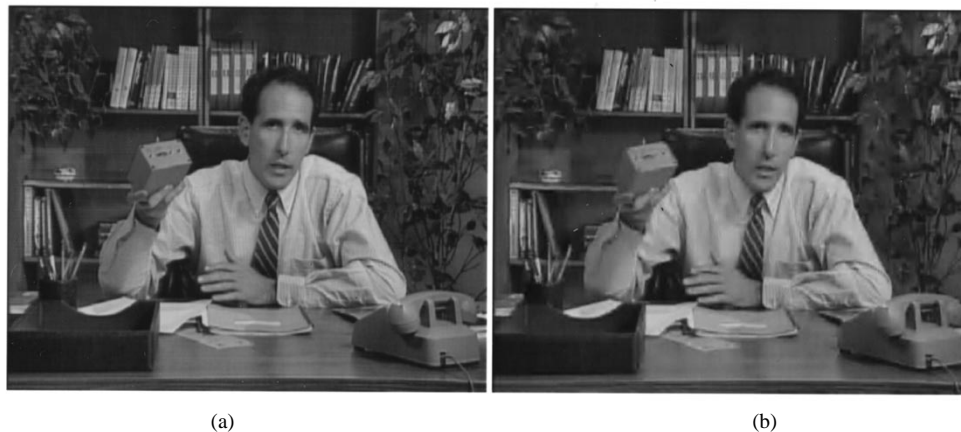

Fig. 13.   Reconstructed frame of the "Salesman" sequence: (a) original frame and (b) overall reconstruction.

visual improvements are of significant importance. For the "Salesman" sequence, we achieved the 40 : 1 compression required for videoconferencing. The compression ratio of 40 : 1 for a common intermediate format (CIF) sequence means the luminance signal is coded at 304 kb/s, which leaves 64 kb/s for the chrominance signal and 16 kb/s for the audio in a 384 kb/s video conferencing application, similar to the scheme adopted in [8], [17]. The PSNR of our results is 33.97 dB and is lower than H.261. However, the perceptual quality of our coded video is better. Note that we only use

the two-tap Haar filter for temporal decomposition and only decompose the LPT band into two levels (we therefore cannot take advantage of the efficient zerotree coding). Nevertheless, the good visual quality of these reconstructed images suggests that the proposed quantization approach is very promising in image and video compression because it is capable of preserving those visually significant components at low bit rates through its signal adaptive and scene adaptive quantization. Recently, Pearlman's group reported a 3-D subband video coding scheme using improved zerotree coding [44]

TABLE I
PSNR OF THE RECONSTRUCTION AND OVERALL ENTROPY REDUCTION IN HIGH-FREQUENCY (HF) SUBBANDS

Quantization scheme	PSNR (dB)	PSNR (dB) after enhancement	Average HF entropy	Average HF entropy after quantization
"lena", Gaussian modeling, ICM	35.53	35.57	3.46	0.320
"lena", Gaussian modeling, NICM	35.54	35.62	3.46	0.318
"lena", Laplacian modeling, NICM	36.29	36.32	3.46	0.316
"salesman", Gaussian modeling, ICM	32.01	32.16	2.58	0.132
"salesman", Gaussian modeling, NICM	31.92	32.07	2.58	0.131
"salesman", Laplacian modeling, NICM	32.97	33.15	2.58	0.129

TABLE II
ENTROPY REDUCTION AFTER QUANTIZATION FOR "SALESMAN" SEQUENCE

Subbands		Before Quantization	After Quantization	Quantization levels
LPT	LLLL	6.66	2.85	stepsize $\Delta=8$
	LLHL	3.98	0.70	7
	LLLH	4.18	0.66	7
	LLHH	2.83	0.16	5
	HL	3.65	0.29	3
	LH	3.61	0.27	3
HPT	LL	1.70	0.07	3

which is able to match the PSNR performance of the motion compensation-based schemes, such as H.261 and H.263.

VI. DISCUSSION AND CONCLUSIONS

It is well known [24] that the HVS tends to be attentive to the major structured discontinuities within an image, rather than intensity changes of individual pixels. Therefore, a desired property for a quantization scheme is the capability of high fidelity representation of major scene structures. Unlike the DCT-based schemes in which spatial information is lost after the transform, the wavelet transform preserves both spatial and frequency information in the decomposed subbands. Since the nature of image scene structures is nonstationary and varies for each individual image, a simple statistical model, as adopted by many existing quantization schemes, is often inadequate for individual scene representation. The combination of a scene structure model and a conventional statistical model will be more appropriate to characterize both the random and deterministic scene distributions within an image. Because scene structures of objects can often be represented by edges, a primitive candidate for scene structure description will be the location, strength, and orientation of edges. In wavelet coding, such edge information is already available in the high-frequency subbands. The issue is how to combine such information with statistical models to achieve a scene adaptive and signal adaptive quantization.

The proposed quantization scheme has provided us an effective way of distinguishing perceptually more important structures from less important ones. Within the high-frequency subbands, those strong and clustered edges correspond to important scene structures and are retained, while those weak and isolated impulses correspond to perceptually negligible components and are discarded. To identify these clustered edges, neighborhood coefficients need to be bound together to determine the presence of scene structures. The binding of scene structures is accomplished by the introduction of naturally defined Gibbs neighborhood systems in the proposed adaptive quantization, while in vector quantization it is accomplished by artificial block partition, which is often inconsistent with the natural boundaries of objects. It is noteworthy that a Gibbs neighborhood system is of dynamic nature since the neighbors of each individual coefficient are different from one location to another. Such a dynamic, individualized neighborhood system is consistent with the natural representation of spatial dependencies and is therefore able to overcome the potential scene distortions caused by any artificial partitioning.

In summary, this novel *scene* adaptive and *signal* adaptive quantization scheme is able to resolve the common problems with some existing quantization methods were designed for wavelet-based compression. The novelty of the proposed quantization lies in the way we exploit the both the spatial and frequency redundancies in the subbands, which are generally related to the psychovisual redundancy of the HVS. The principle of the scene adaptive and signal adaptive quantization is fundamentally different from existing scalar or vector

quantization schemes in that we combine a scene structure model with a conventional statistical model. This quantization scheme has the individuality of scalar quantization in that each coefficient is inspected with regards to its perceptual importance, but in a more efficient way than traditional scalar quantization. It also exploits the local spatial correlation in images as in the case of vector quantization, but in an essentially different way such that it is able to preserve inherent image structures even at low bit rates. Both algorithmic analysis and experimental results have shown that the proposed adaptive quantization provides a promising way of achieving efficient image and video compression at low bit rates. In addition, such adaptive quantization has many refreshing impacts on the subsequent coding and transmission in such aspects as coding efficiency, coding artifacts reduction, transmission loss concealment, and transmission noise reduction, which are currently under investigation.

ACKNOWLEDGMENT

REFERENCES

[1] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.
[2] M. Liou, "Overview of the px64 kbit/s video coding standard," *Commun. ACM*, vol. 34, pp. 59–63, Apr. 1991.
[3] D. L. Gall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, pp. 46–58, Apr. 1991.
[4] M. Vetterli, "Multi-dimensional subband coding: Some theory and algorithms," *Signal Processing*, vol. 6, pp. 97–112, 1984.
[5] G. Karlsson and M. Vetterli, "Three dimensional sub-band coding of video," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 1988, pp. 1110–1113.
[6] J. W. Woods and S. D. O'Nell, "Sub-band coding of images," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 34, pp. 1278–1288, 1986.
[7] R. H. Bamberger, "New subband decompositions and coders for image and video compression," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, San Francisco, CA, 1992, pp. III–217–220.
[8] C. Podilchuk and A. Jacquin, "Subband video coding with a dynamic bit allocation and geometric vector quantization," in *Human Vision, Visual Processing, and Digital Display III*, vol. SPIE 1666, Boston, MA, pp. 241–252, 1992.
[9] J. Hartung, "Architecture for the real-time implementation of three-dimensional subband video coding," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, San Francisco, CA, 1992, pp. III–225–228.
[10] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using vector quantization in the wavelet transform domain," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 1990, pp. 2297–2300.
[11] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 901–996, 1988.
[12] T. Naveen and J. W. Woods, "Subband finite state scalar quantization," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, 1993, pp. V–613–616.
[13] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, 1992.
[14] N. Mohsenian and N. M. Nasrabadi, "Subband coding of video using edge-based vector quantization technique for compression of the upper bands," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, San Francisco, CA, 1992, pp. III–233–236.
[15] O. Johnsen, O. V. Shentov, and S. K. Mitra, "A technique for the efficient coding of the upper bands in subband coding of images," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Albuquerque, NM, 1990, pp. 2097–2100.
[16] C. Podilchuk, N. S. Jayant, and P. Noll, "Sparse codebooks for the quantization of nondominant subbands in image coding," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Albuquerque, NM, 1990, pp. 2101–2104.
[17] C. Podilchuk, N. S. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Processing*, vol. 2, pp. 125–139, Feb. 1995.
[18] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.
[19] R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital coding of speech in subbands," *Bell Syst. Tech. J.*, vol. 55, pp. 1069–1985, 1976.
[20] H. Gharavi and A. Tabatabai, "Subband coding of monochrome and color images," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 207–214, 1988.
[21] P. H. Westerink, J. Biemond, D. E. Boekee, and J. W. Woods, "Subband coding of images using vector quantization," *IEEE Trans. Commun.*, vol. 36, pp. 713–719, 1988.
[22] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 244–250, Apr. 1992.
[23] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 1989, pp. 1945–1948.
[24] D. Marr, *Vision*. New York: Freeman, 1982.
[25] W. E. L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*. Cambridge, MA: MIT Press, 1990.
[26] D. G. Lowe, *Perceptual Organization and Visual Recognition*. Boston, MA: Kluwer, 1985.
[27] A. P. Witkin and M. Tenenbaum, "On the role of structure in vision," in *Human & Machine Vision*, A. Rosenfeld and J. Beck Eds. New York: Academic, 1983.
[28] A. Sha'ashua and S. Ullman, "Structural saliency: The detection of globally salient structures using a locally connected network," in *Proc. 2nd Int. Conf. Computer Vision*, Tampa, FL, 1988, pp. 321–327.
[29] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, 1984.
[30] T. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Signal Processing*, vol. 40, pp. 901–914, 1992.
[31] J. Luo, C. W. Chen, and K. J. Parker, "On the application of Gibbs random field in image processing: From segmentation to enhancement," *J. Electron. Imaging*, vol. 4, no. 2, pp. 189–198, Apr. 1995.
[32] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 39–55, 1987.
[33] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, 1989.
[34] J. Besag, "Spatial interaction and the statistical analysis of lattice system," *J. Royal Stat. Soc.*, vol. 36, pp. 192–326, 1974.
[35] H. Gharavi, "Subband coding of video signals," in *Subband Image Coding*, J. W. Woods, Ed. Boston, MA: Kluwer, 1991, pp. 229–272.
[36] J. Luo, "Low bit rate wavelet-based image and video compression with adaptive quantization, coding and postprocessing," Ph.D. dissertation, Univ. Rochester, Dept. Elec. Eng., Rochester, NY, 1995.
[37] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc.*, vol. 48, pp. 259–302, 1986.
[38] J. Luo, C. W. Chen, and K. J. Parker, "On the application of Gibbs random field in image processing: From segmentation to enhancement," in *Visual Communication and Image Processing*, A. K. Katsaggelos, Ed. Chicago, IL: SPIE, vol. 2308, pp. 1289–1300, Sept. 1994.
[39] A. Perez, S. Kamata, and E. Kawagushi, "Hilbert scanning arithmetic coding for multispectral image compression," in *Proc. SPIE Conf. Application for Digital Image Processing XIV*, A. G. Tescher, Ed., San Diego, CA: SPIE, July 1991, vol. 1567, pp. 354–361.
[40] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.
[41] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, 1996, to appear.
[42] R. L. Stevenson, "Reduction of coding artifacts in transform image coding," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, 1993, pp. V401–V404.
[43] J. Luo, C. W. Chen, K. J. Parker, and T. S. Huang, "A new method for block effect removal in low bit rate image compression," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Australia, 1994, pp. V341–V344.

[44] Y. Chen and W. A. Pearlman, "Three-dimensional subband coding of video using the zero-tree method," in *Proc. SPIE Symp. Visual Communication and Image Processing*, Orlando, FL, 1996, pp. 1302–1312.

Jiebo Luo (S'92–M'96) was born in Kunming, China, in 1967. He received the B.S. and M.S. degrees in electrical engineering from the University of Science and Technology of China, Hefei, in 1989 and 1992, respectively. In 1995, he received the Ph.D. degree in electrical engineering from the University of Rochester, NY.

He was a Research Assistant in the Image Processing Laboratory and the NSF Center for Electronic Imaging Systems at the University of Rochester from September 1992 to November 1995. In the summer of 1995, he was employed at the Joseph C. Wilson Center for Technology of Xerox Corporation, Webster, NY. He is currently a Senior Research Scientist in the Imaging Science Technology Laboratory of Eastman Kodak Imaging Research and Advanced Development, Rochester, NY. His research interests include image enhancement and manipulation, digital photography, image and video coding, wavelets and applications, medical imaging, pattern recognition, and computer vision.

Dr. Luo was the recipient of the 1994 SPIE Best Student Paper Award for Visual Communication and Image Processing. He is also a member of SPIE.

Chang Wen Chen (S'86–M'90) received the B.S. degree in electrical engineering from University of Science and Technology of China, Hefei, in 1983, the M.S.E.E. degree from University of Southern California, Los Angeles, in 1986, and the Ph.D. degree in electrical engineering from University of Illinois at Urbana-Champaign in 1992.

He was a Research Assistant in the Signal and Image Processing Institute, University of Southern California, Los Angeles, in 1986. From January 1987 to July 1992, he was a Research Assistant at the Coordinated Science Laboratory and the Beckman Institute, University of Illinois at Urbana-Champaign. In the summers of 1989 and 1990, he was employed at National Center for Supercomputing Applications, Champaign, IL, working with the Visualization Service and Development Group. From August 1992 to September 1996, he was on the faculty of the Department of Electrical Engineering, University of Rochester. In 1995, he spent his summer at Kodak working with the Imaging Science Research Labs. Since September 1996, he has been an Assistant Professor of Electrical Engineering at the University of Missouri–Columbia. His current research interests include image and video coding, wireless communication, wavelets, biomedical image processing, graphics, and visualization.
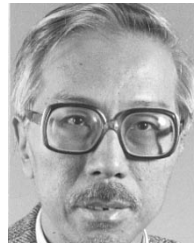
Dr. Chen is a member of SPIE and Tau Beta Pi. He is a senior author of the 1994 SPIE Best Student Paper for Visual Communication and Image Processing.

Kevin J. Parker (S'79–M'81–SM'87–F'95) received the B.S. degree in engineering science, *summa cum laude*, from the State University of New York at Buffalo in 1976. He received the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1978 and 1981.

From 1981 to 1985 he was an Assistant Professor of electrical engineering and radiology. He serves as reviewer and consultant for a number of journals and institutions. His research interests are in medical imaging, linear and nonlinear acoustics, and digital halftoning.

Dr. Parker has received awards from the National Institute of General Medical Sciences (1979), the Lilly Teaching Endowment (1982), the IBM Supercomputing Competition (1989), and the World Federation of Ultrasound in Medicine and Biology (1991). He is a member of the IEEE Sonics and Ultrasonics Symposium Technical Committee. He is also a member of the Acoustical Society of America and the American Institute of Ultrasound in Medicine (AIUM). He has been named a Fellow in both the IEEE and the AIUM for his work in medical imaging. In addition, he was recently named to the Board of Governors of the AIUM.

Thomas S. Huang (S'61–M'63–SM'71–F'79) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, China, and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge.

He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology. During his sabbatical leaves, he has worked at the MIT Lincoln Laboratory, the IBM Thomas J. Watson Research Center, and the Rheinishes Landes Museum in Bonn, West Germany, and he held Visiting Professor positions at the Swiss Institutes of Technology in Zurich and Lausanne, University of Hannover in West Germany, INRS-Telecommunications of the University of Quebec in Montreal, Canada, and University of Tokyo, Japan. He has served as a consultant to numerous industrial firms and government agencies both in the United States and abroad. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 11 books and over 300 papers in network theory, digital filtering, image processing, and computer vision.

Dr. Huang is a Fellow of the International Association of Pattern Recognition and the Optical Society of American and has received a Guggenheim Fellowship, an A. V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Acoustics, Speech, and Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He is a Founding Editor of the *International Journal Computer Vision, Graphics, and Image Processing* and Editor of the "Springer Series in Information Sciences," published by Springer Verlag.