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**

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

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 subbands exhibit quite different characteristics from one to another 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

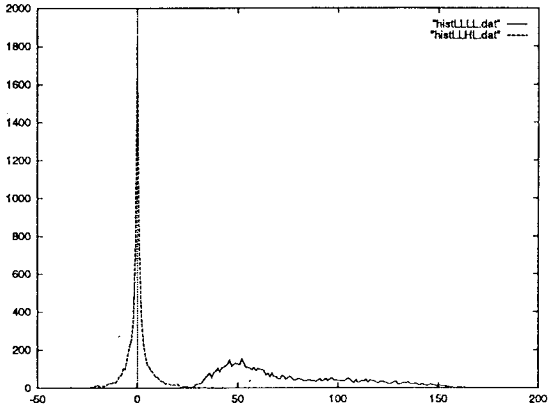


Fig. 3. Typical histograms of the subbands (_____ the lowpass band, - - - a highpass band). The horizontal axis is the intensity axis, and the vertical axis is the histogram count axis.

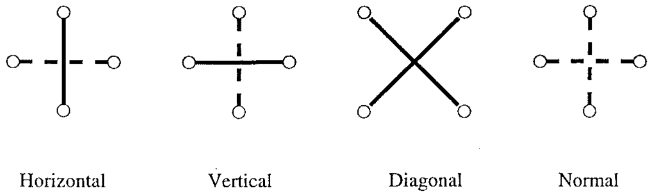


Fig. 4. Cliques for subbands with different preferential directions.

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

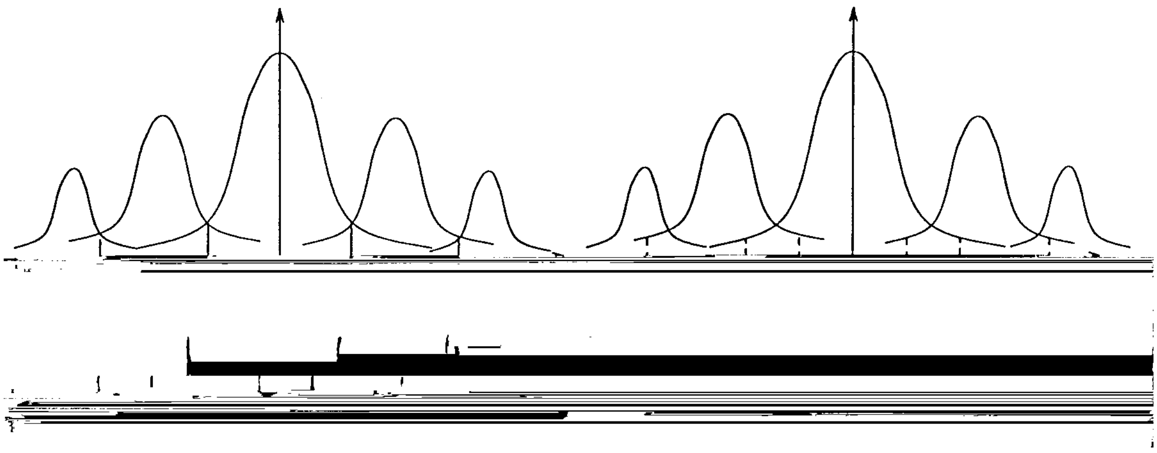


Fig. 6. Dead zone effect.

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 c4ave 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. B EYOND QUANTIZATION

A. Coding of the Quantized High Frequency Subbands

Coding of an image generally includes two distinct operations: quantization 663is429(symbol)-428(coding.)-431(The)-432(adaptive)-430(quan-)]TJfl T*fl [(tization)-487(with)-480(spatial)-485(cons word length coder to code the labels of the nonzero value 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 valuecan be better exploited to improve the runlength coding efficiency

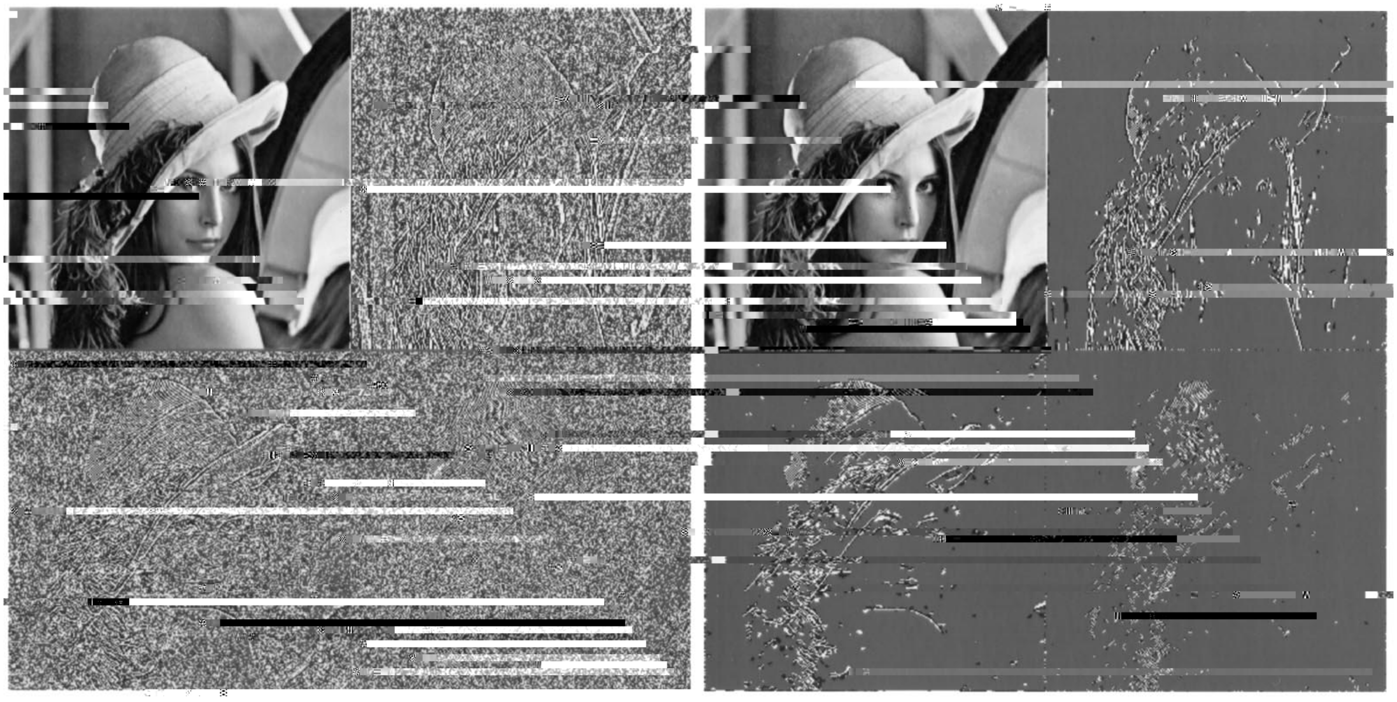
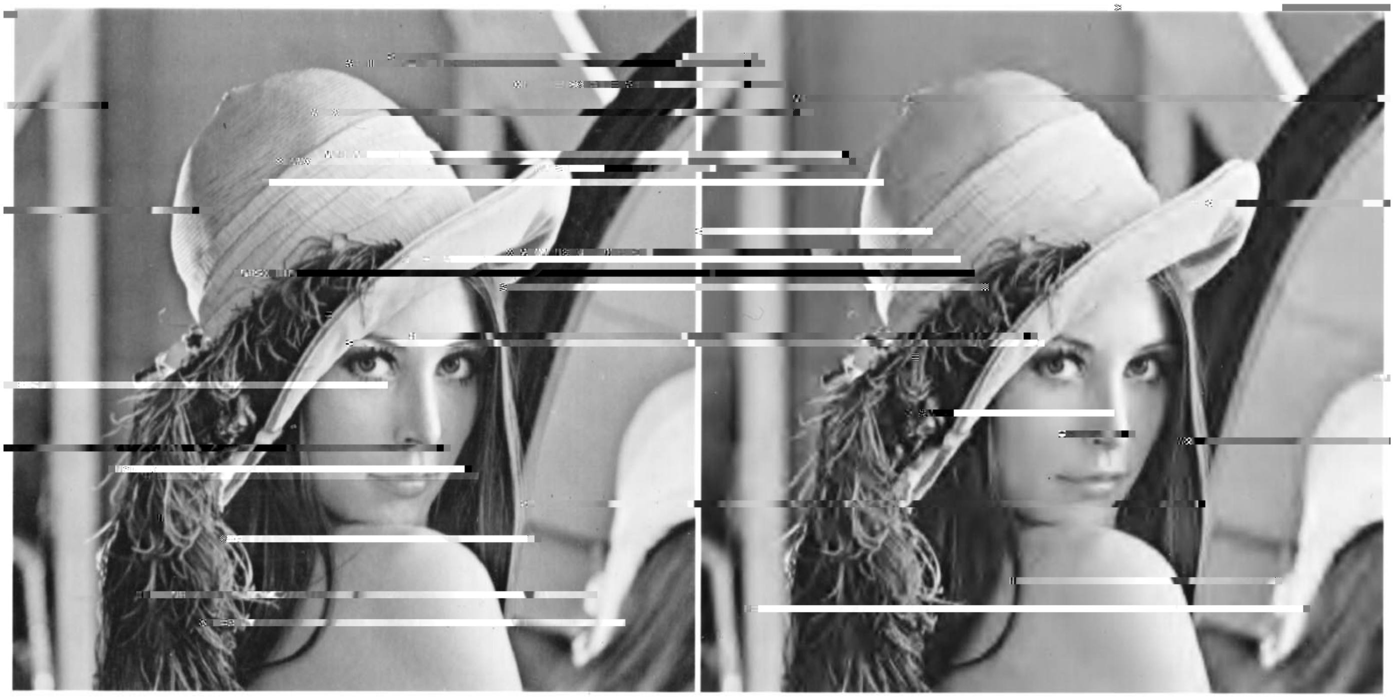


Fig. 7. A four-band decomposition of the “Lena” image: (a) original subbands and (b) quantized high frequency subbands.

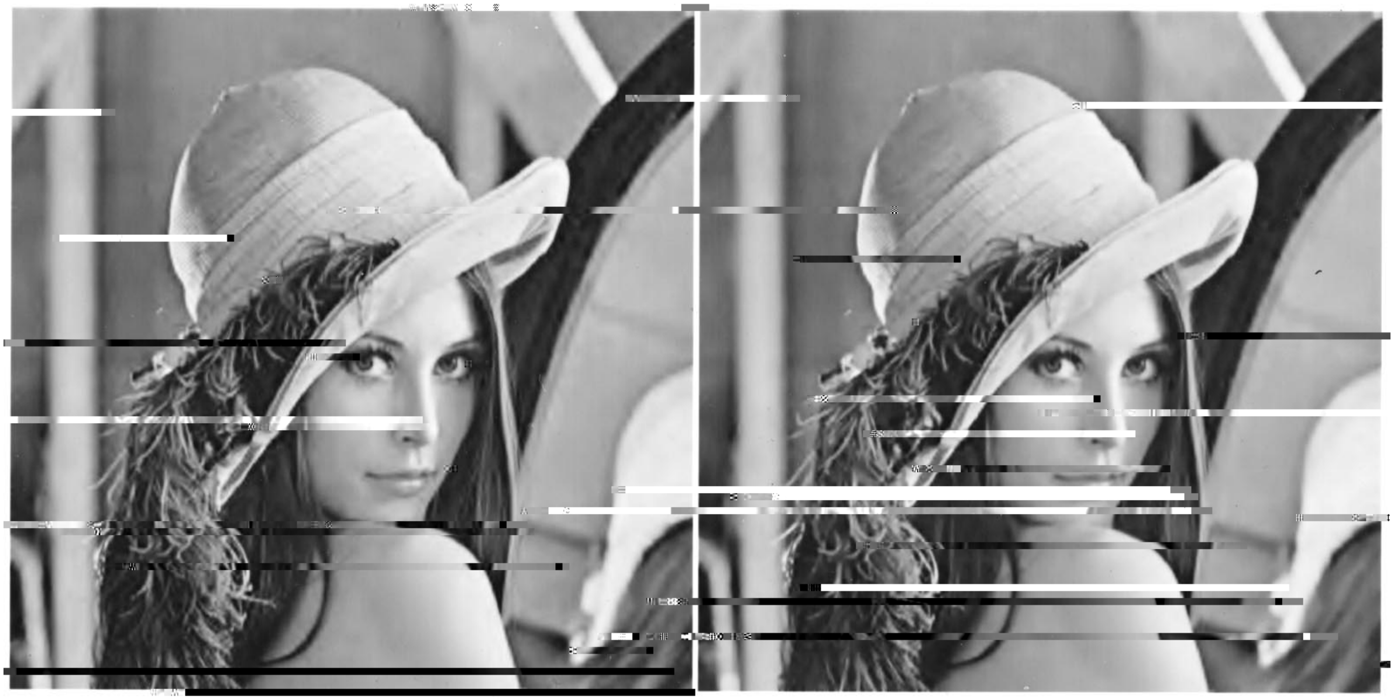
The conditional probability of the quantization y given the original data x can be written as

$$p(y \mid x) = \begin{cases} 1, & y = \end{cases}$$



(a) (b)

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



(a) (b)

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

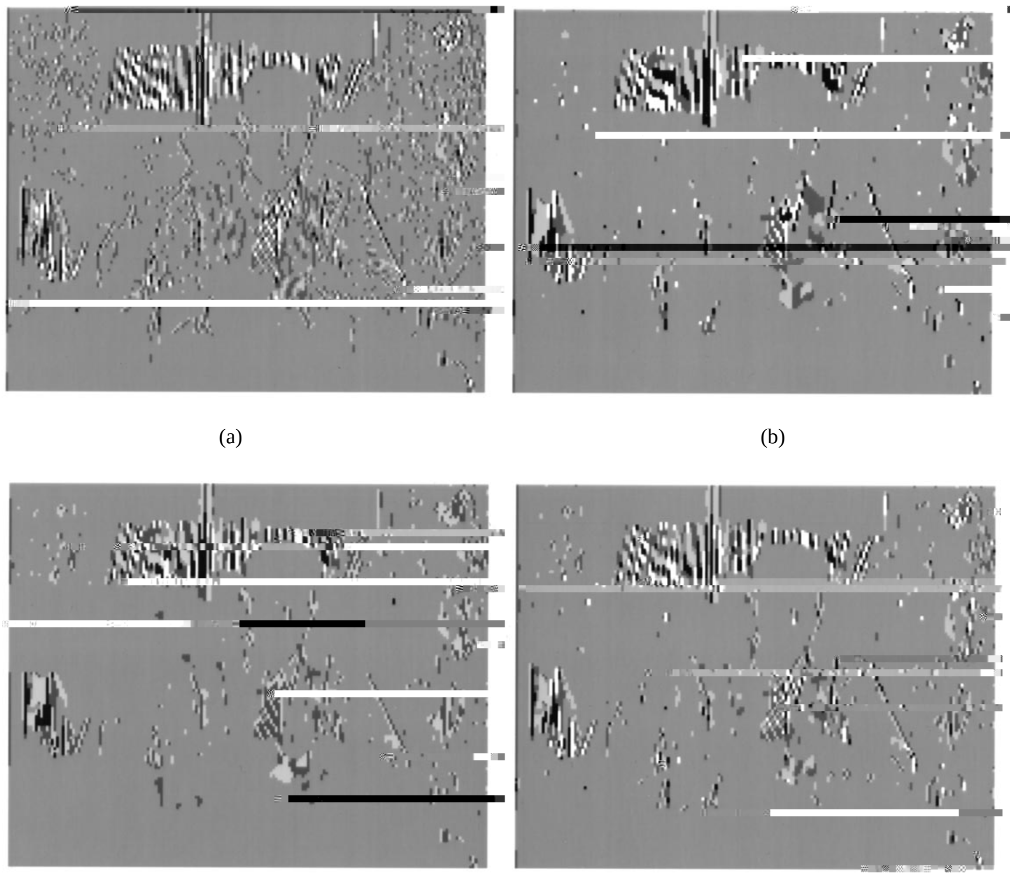


Fig. 12. Quantization of a high-frequency subband (blowup). (a) Lloyd–Max quantizer without spatial constraints, (b) adaptive quantization with Gaussian modeling, (c) adaptive quantization with Gaussian modeling and ICM, and (d) adaptive quantization with Laplacian modeling and NICM.

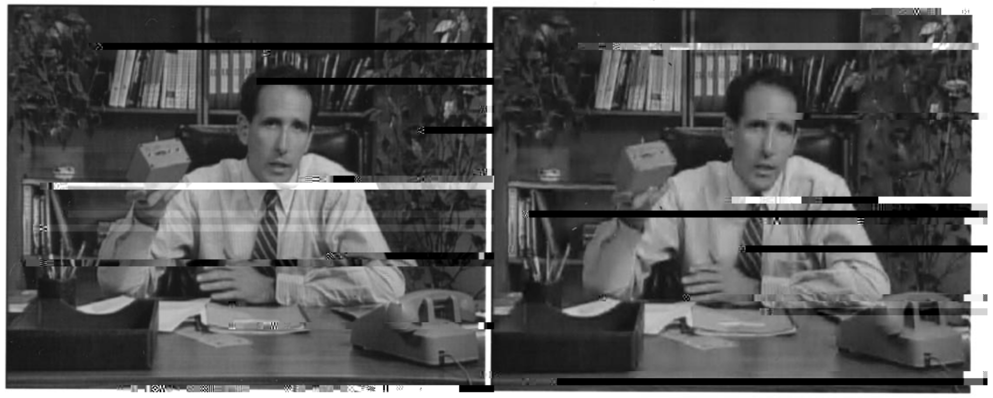


Fig. 13. Reconstructed frame of the "Salesman" sequence: (a) original frame and (b) overall reconstruction.

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(.s (reason) [illegible]

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	36.29	36.32	3.46	0.316

[redacted]

TABLE II
ENTROPY REDUCTION AFTER QUANTIZATION FOR "SALESMAN" SEQUENCE

Subbands		Before Quantization	After Quantization	Quantization levels
LPT	LLLL	6.66	2.85	stepsize Δ=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