Binary Decompositions for High-Order Entropy Coding of Grayscale Images

Steve S. Yu, *Member, IEEE,* and Nikolas P. Galatsanos, *Senior Member, IEEE*

***Abstract*—High-order entropy coding (HOEC) has the potential to provide higher compression ratios than the usually used zero-order entropy coding (ZOEC) approaches. However, serious implementation difficulties severely limit the practical value of HOEC for grayscale image compression. We examine the bit-plane decomposition (BPD) representation as a simple alternative that bypasses some of the implementation difficulties of HOEC. We show, however, that BPD introduces undesired coding overhead when used to represent grayscale images. We therefore propose a new binary image representation called magnitude-based binary decomposition (MBBD) which avoids any coding overhead when used to represent grayscale images. Thus, MBBD both bypasses the implementation difficulties of HOEC and does not have the drawbacks of the BPD. We present numerical experiments that verify the theoretical analysis of the BPD and MBBD representations. In addition, our experiments demonstrate that MBBD-HOEC yields better results than ZOEC for lossy image compression and is also very effective for progressive image transmission.**

I. Introduction

ENTROPY coding is a data compression approach which exploits the nonuniformity of the probability distribution of the data and encodes them using a variable length code. A representative example of this approach is the well-known Huffman code [1]. It is also known that the high-order entropy properties of a source can be exploited effectively using either the joint probability of a group data or their conditional probability [2].

In order to apply high-order entropy coding (HOEC), the data model must describe how the outcome of a datum depends statistically on its neighbors. In an Lth order data model, the redundancy of the data can be exploited using the Lth order conditional probability. To achieve optimal compression of a given data source, two basic problems must be considered. Firstly, from Shannon's basic encoding theorem, we know that for a given data sequence, the minimum average length of a code is equal to the entropy, H, of the sequence [3]. Finding such a code is not a trivial problem. Fortunately, arithmetic coding which was recently developed fulfills this expectation [4], [5]. Secondly, the entropy, H, is determined by the statistics of the source, i.e., the probability density function (pdf) of the source. Generating optimal codes for a source model requires knowledge of the pdf. Thus, estimating the pdf of the data for a given source model at the receiving side is another difficult problem. One approach of estimating the pdf of the source is to adaptively update it at both the transmitter and the receiver. Universal modeling is an adaptive technique which generates the statistics during the actual coding operation by adapting on-line [6]. Langdon and Rissanen successfully used tenth order source model and arithmetic coding for binary image compression [7].

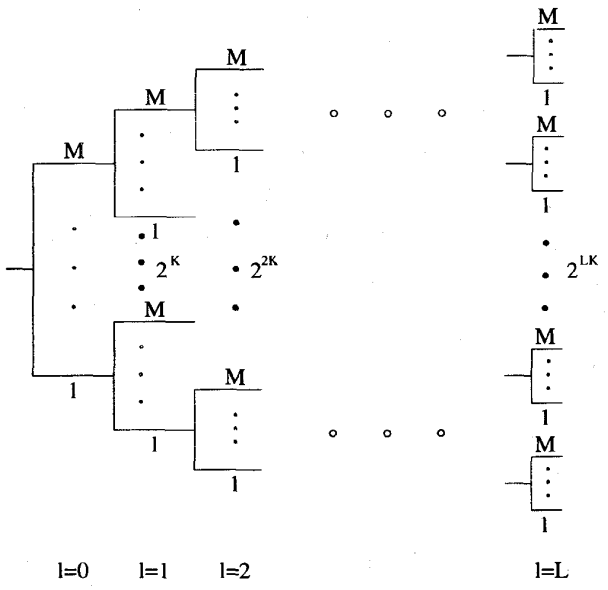


Fig. 1. Tree structure of Lth order Markov model.

The application of HOEC to high quality grayscale images with many gray levels requires the estimation of a formidably large number of conditional pdf's. This results in a huge implementation complexity. To better explain this, the conditional model of a source will be represented as a tree. In Fig. 1 such a tree is shown for an Lth order model with $M = 2^K$ symbols. It is clear that a Lth order model will contain M^L entries. In addition, the severe inaccuracies of estimating the high-order statistics would diminish the gains through HOEC.

HOEC has been used in many studies for both lossy and lossless image compression. Gharavi [8] suggested that the simplest solution to the large number of pdf's was to reduce the number of grayscale values for conditioning. For example, in a first order source model, if a pixel (8 b, 256 grayscale values) is counted as a previous condition, we can quantize this pixel to 3 b so that only eight conditioning states are required.

Manuscript received September 12, 1994; revised June 3, 1995. This paper was recommended by Associate Editor K.-H. Tzou.

S. S. Yu is with AT&T Bell Laboratories, Naperville, IL 60566 USA.

N. P. Galatsanos is with Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA.

Publisher Item Identifier S 1051-8215(96)01319-5.

Using a second order source model, his experiments indicated that when the grayscale values for the conditioning pixels is reduced to 3 b, the HOEC can no longer preserve its superiority over zero-order entropy coding (ZOEC) approaches. However, even in this case, 64 tables with 256 probabilities in each table are required. It has been reported that conditional pdf's can be eliminated or merged without significant effect on the coding results [9], [10]. These approaches start with a first-order conditional tree and then extend the tree iteratively by growing only those branches which yield maximum entropy reduction. The tree extension process continues until the desired number of leaves is obtained. This method was further developed to achieve more reduction [11] by limiting the number of code tables and the size of the tables using vector quantization [12]. Although these reduction techniques do not require knowledge of entire tree, they still require a large number of conditional probabilities to start the process of reduction. Another alternative is to grow the tree, from the root, with the coding operation. Generally, the probabilities gradually generated by this process will be closer to the image statistical characteristics than those obtained by assuming that each state contains the same initial probabilities. However, one needs to assign suitable probabilities to code data which have "brand new" conditioning states, i.e., states which are encountered for the first time. To solve this *zero frequency* problem, Bar–Ness and Choi [13] employed *flags* to signify nonexisting states. This not only makes coding more complex but also increases the coding cost because the flags introduce coding overhead.

In this paper, we examine the well-known bit-plane decomposition (BPD) representation as an alternative approach to bypass the implementation difficulties of HOEC. However, we found that this representation introduces coding overhead. We, therefore, introduce a new binary image representation called magnitude-based binary decomposition (MBBD). The advantages of MBBD are: first, it significantly reduces the complexity and the storage requirements of the HOEC implementation. Second, it does not introduce any coding overhead as BPD representation. Finally, it is very simple to implement. The rest of this paper is organized as follows: in Section II, we present the necessary background for this study and discuss the practical restrictions that apply to the implementation of HOEC. In Section III, we present the theoretical analysis of the BPD and MBBD representations. In Section IV, we give some implementation details for the application of MBBD-HOEC to lossy transform-based image compression. In Section V, we present experimental results that verify the theoretical analysis and the value of this approach for lossy image compression applications. Finally, in Section VI we present our conclusions from this work.

II. Background

Based on Shannon's fundamental theorem [2], the minimum bit rate required for the representation of a source is determined by the entropy of the source. Two commonly used source models are the zero-order (statistical independent or memoryless) and the high-order source. They can be used, respectively, to characterize the amount of information and thus the low bound limits of the coding cost for each case.

A. Memoryless Entropy and High-Order Entropy

For an independent source, the entropy is given by

$$H(X) = -\sum_{x} P(x) \log_2 P(x) \tag{1}$$

where X is a random variable, x is a possible realization with probability $P(x)$, and the units of $H(X)$ are b/pel. The entropy $H(X)$ establishes a low bound on the average bit rate for a memoryless source and gives the amount of information carried by random variable X.

Two commonly used measures of the information for a correlated source are the block entropy and the conditional entropy. Block entropy is used to characterize the source when a block of data is combined. For an L-dimensional vector, the block entropy is defined as

$$\begin{aligned} H_L(\mathbf{X}) &\triangleq H_L(X_1, X_2, \cdots, X_L) \\ &= -\sum_{\mathbf{x}} P(\mathbf{x}) \log_2 P(\mathbf{x}) \end{aligned} \tag{2}$$

where $\mathbf{X}$ is a random vector and $\mathbf{x}$ is a possible realization vector with probability $P(\mathbf{x})$. The summation is over all the 2^{LK} possible L-tuples for each pixel with K b grayscale.

For conditional entropy, we assume that $L-1$ components from L-dimensional vector $\mathbf{x}$, $x_1, x_2, \cdots, x_L - 1$, have already been received by the receiver. Component x_L can then be coded according to the conditional entropy as

$$\begin{aligned} &H(X_L|X_1, X_2, \cdots, X_{L-1}) = \\ &\quad -\sum_{\mathbf{x}} P(\mathbf{x}) \log_2 P(x_L|x_1, x_2, \cdots, x_{L-1}). \end{aligned} \tag{3}$$

It is well known [14] that for block entropies we have

$$\begin{aligned} \frac{1}{L} H_L(\mathbf{X}) &\le \frac{1}{L-1} H_{L-1}(\mathbf{X}) \\ &\le \cdots \\ &\le H_1(\mathbf{X}) \\ &\triangleq H(X) \end{aligned} \tag{4}$$

and for the conditional entropies

$$\begin{aligned} &H(X_L|X_1, X_2, \cdots, X_{L-1}) \\ &\quad \le H(X_{L-1}|X_1, X_2, \cdots, X_{L-2}) \\ &\quad \le \cdots \\ &\quad \le H(X_2|X_1) \\ &\quad \le H(X_1) \\ &\quad \triangleq H(X). \end{aligned} \tag{5}$$

The previous results indicate that the higher the order of the probability used, the lower the resulting entropy. These equations are satisfied with the equality if and only if the pixels are statistically independent. Nevertheless, most images of practical interest are spatially correlated even after application of a decorrelating transform.

B. Practical Difficulties in Implementation of HOEC

From our earlier discussion, it is clear that HOEC can theoretically improve coding efficiency. However, the complexity of HOEC implementation increases exponentially with the order L. Therefore, the practical value of this approach is much smaller than that theoretically expected.

The implementation complexity of high-order conditional coding to image data can be easily understood by examining the tree structure of the conditional probability model. As mentioned in the introduction, an Lth order source model with a K b representation levels requires 2^{KL} codes tables with $M = 2^K$ probabilities in each table. As an example, a typical image pixel is usually quantized to 8 b, $K = 8$. If we consider tenth order source model, $L = 10$, then we need 2^{80} (about 10^{24}) tables with 256 probabilities in each table. This number is too large to even consider implementing with present computer technology.

Another problem of using HOEC is the inaccuracy that occurs when estimating the conditional probabilities. In most practical applications, the source model is not known and is incrementally built at the receiver as data arrives. Two strategies can be used to estimate the conditional probabilities. The first one starts with the full tree using certain initial pdf's, for example uniform. This method immediately requires a large tree structure and yields unacceptable transient characteristics of the estimated statistics. The second one grows the tree from the root, simultaneously with the coding operation. In this situation, the initial pdf tables are all empty (zero value). Generally, the probabilities gradually generated by this process will be closer to the statistical characteristics of the image. However, the drawback in both approaches is that they use estimated statistics. This leads to inaccurate estimates of the average code lengths and, hence, the coding advantages of using HOEC implied by (4) or (5) cannot be materialized. It is clear that as the order of the source model increases, since the number of the data remains the same, fewer data can be used to estimate the required conditional probabilities. Thus, in practice there is a point after which the use of a higher order source model does not reduce the coding cost.

III. Binary Image Representations

In order to use HOEC successfully, two critical issues must be addressed: first, the complexity of the implementation, and second the inaccuracy in estimating the conditional probabilities. Both problems are inherent to HOEC implementations which require 2^{KL} code-tables (or states) and total $2^{K(1+L)}$ conditional probabilities.

One way to solve these problems is to reduce K. Since for grayscale images K determines the quantization levels of the pixels, it cannot be reduced infinitely. However, if a multilevel image ($K > 1$) can be represented using a group of binary subimages ($K = 1$), the representation levels of the original pixels can be determined by the number of the subimages. Thus, if HOEC coding is applied to the subimages instead of the original image, the implementation complexity will be significantly reduced and the accuracy of estimating the statistical model will also be increased.

However, there are some concerns for any new image representation. First, an exact replica of the original image should be reconstructible from the new representation. This means whatever information is contained in the old representation must be included in the new representation without any loss. Second, the new representation should still maintain the statistical redundancy of adjacent pixels so it can be further exploited by utilizing HOEC. Finally, the new representation must avoid introducing any kind of redundant information that will increase the coding cost.

A. Bit-Plane Decomposition Image Coding

The bit-plane decomposition (BPD) is an image representation that was originally designed for coding certain types of telemetric data [15]. The potential applications of BPD extend beyond telemetric data and the merits of its technique are summarized in [16] and [17]. For BPD, each pixel of a grayscale image is represented by a K bit binary code so that the grayscale image is decomposed into K images, each having only two levels denoted by zero or one. Each two-level or 1 b image is referred to as a bit-plane, and one can think of an ordering of these bit-plane images from the most significant to the least significant. For an $N \times N$ image, each bit-plane contains $N \times N$ picture elements to be coded. The advantage of this decomposition is that each bit-plane can then be coded using a binary image compression technique. Thus, HOEC can be used to code each bit-plane separately. It is clear that if the BPD representation is used, the implementation of HOEC is greatly facilitated since the number of possible states in each bit-plane is only 2^L.

Although BPD-based image coding is a well-known compression technique, to the best of our knowledge, no one in the past looked into analyzing rigorously the theoretical coding cost of this representation. We believe that it is important to analyze the coding cost of the BPD for a zeroth order source model before using it with a higher-order source models. This analysis could give an indication whether BPD can at least in theory yield a reduction in the coding cost when it is used in conjunction with HOEC.

Letting H_{BPD} represent the memoryless entropy of the BPD of an image, and H represent the memoryless entropy of original image, we show in Appendix A that:

Theorem 1—$H_{BPD} \geq H$: This theorem implies that the BPD representation of a grayscale image requires a larger code than that of the original image. This is not surprising since dependencies among the bit-planes are not used by the BPD during coding. Indeed, in our proof of Theorem 1 in Appendix A we show that the entropy of the original data is equal to the conditional entropy of the bit-planes of the BPD, $H(\mathbf{X}) = H_{BPD}(\mathbf{X}|\, BP)$.

The conditional entropy of the BPD for an Lth-order source model is $H_{BPD}(\mathbf{X}|\, S^L)$, where S^L indicates the conditional state defined by L components. Based on (5) we have $H_{BPD}(\mathbf{X}|\, S^L) \leq H_{BPD}$. However, since $H \leq H_{BPD}$, no conclusion can be drawn from the relation of $H_{BPD}(\mathbf{X}|\, S^L)$ and H. This indicates that the BPD decomposition *cannot theoretically guarantee* a reduction of the memoryless coding

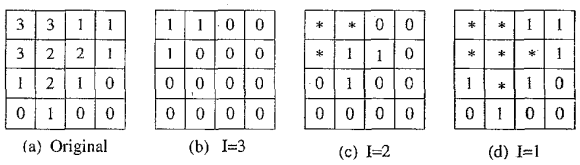


Fig. 2. An example of binary decomposed representation of image.

cost when HOEC is applied. In fact, if we consider both within a bit-plane and among bit-planes correlations, we will get $H_{BPD}(\mathbf{X}|\, BP,\, S^L) \leq H$. However, the implementation of a HOEC scheme that would capitalize on such a 3-D model would be quite complicated. Nevertheless, because of its simplicity and progressive transmission capabilities, BPD representation is still a rather promising technique that can be used with HOEC for compressing some types of images.

B. Magnitude-Based Binary Decomposition Image Coding

Motivated by the coding cost analysis of the BPD and the desire to use HOEC, we propose a new binary representation of grayscale images called MBBD. An image with K b/pixel, $M = 2^K$ representation levels is decomposed into $M - 1$ binary subimages. In each subimage, a "1" denotes the corresponding representation level and a "0" all the other *uncoded* outcomes. These subimages are arranged in the order from the one containing the largest pixel value to the one containing the smallest. It is important to notice that the effective number of "pixels" reduces as we move from the subimage with the largest pixel value to the subimage with the smallest one.

One of the examples for the MBBD representation is illustrated in Fig. 2. The original image is a four-level grayscale ($0 \leq I \leq 3$) image of size 4×4. The original four-level image in Fig. 2(a) is decomposed into three two-level subimages, as shown in Fig. 2(b)–(d). For each subimage, "1" represents the *information* pixel in the subimage, "0" represents the *noninformation* pixel and "*" represents the position with no symbol to be coded. In the first subimage, Fig. 2(b), the "1" indicates the position of the pixels with value *three* in the original image and the "0" the position of the pixels with smaller values. In the second subimage, Fig. 2(c), the "1" indicates the position of the pixels with value *two* and the "0" the pixels with values *one* and *zero*. The "*" indicates the positions of the image that we already know their pixel value and thus they can be omitted in this bit plane. It is obvious that the last subimage contains all the information about pixel values of *one* and *zero*.

For the similar reason as for the BPD case, the memoryless entropy of the MBBD representation was compared to the memoryless entropy of the original image data. Let H and H_{MBBD} denote the memoryless entropy of the original data and the MBBD representation, respectively. In Appendix B we show the following theorem.

Theorem 2—$H_{MBBD} = H$: The implication of this theorem is that when the MBBD representation is used, ZOEC of the data yields the same coding cost as ZOEC of the original representation. Assuming that an Lth order source model is used to independently code each subimage of the MBBD representation. We shall indicate the entropy of this representation by $H_{MBBD}(\mathbf{X}|\, S^L)$. Then, it follows directly from Theorem 2 and the properties of high-order entropy, that $H_{MBBD}(\mathbf{X}|\, S^L) \leq H$. This *guarantees*, at least in theory, if the exact probability model of the source is known, that HOEC of the MBBD representation yields a reduction of the coding cost compared to the use of ZOEC on the original image. In practice, however, the reduction of the coding cost through the use of MBBD-HOEC depends on the accuracy of the estimation of the probability model.

MBBD is a method that provides an efficient platform for exploiting the statistical redundancy in images for compression purposes. Moreover, this representation provides a natural approach for progressive image transmission [18]–[20] when transform-based image coding is used. In this case it is logical to transmit the largest coefficients first followed by the smaller ones.

A similar method for progressive transmission of images was proposed for DCT-based image coding [21]. That method was based on partition priority coding (PPC) which allows the transmission of data in magnitude order without overhead due to prioritization. In fact, MBBD, which provides complete magnitude prioritization, can be viewed as a special case of the PPC. However, there are some significant differences between the two approaches. Firstly, MBBD is designed to be used with HOEC. Thus, MBBD not only provides prioritization capabilities without coding overhead but also provides the capability of lowering bit rate of a memoryless entropy based approach. Secondly, the MBBD yields binary images so that the techniques available for compressing binary images can be used for MBBD image coding. Finally, MBBD is not specifically designed for progressive image transmission, which is usually used for lossy image compression. MBBD-based image coding is effective for lossless image compression as well [22].

IV. Lossy Image Compression Using MBBD-HOEC

MBBD high-order entropy coding is applicable to a wide range of image compression applications. One such application is low bit rate lossy transform-based image compression. In this paper, we used the MBBD decomposition with the discrete cosine transform (DCT) and the subband/wavelet transform (ST/WT). A block diagram of the MBBD-based image codec is shown in Fig. 3.

Successful compression of images at low bit-rates has two requirements: first, the removal of the redundancies in the image data, and second, the elimination of the irrelevant information. A typical low bit rate image compression algorithm contains three basic blocks: transform, quantization, and entropy encoding. During the transform, the image is subjected to an invertible transform with aim to reduce the statistical dependencies of the image elements. The transform coefficients are then quantized to reduce the size of the data. Thus, irrelevant information is discarded through the quantization process. However, this process results in a loss of information which can never be recovered. The entropy

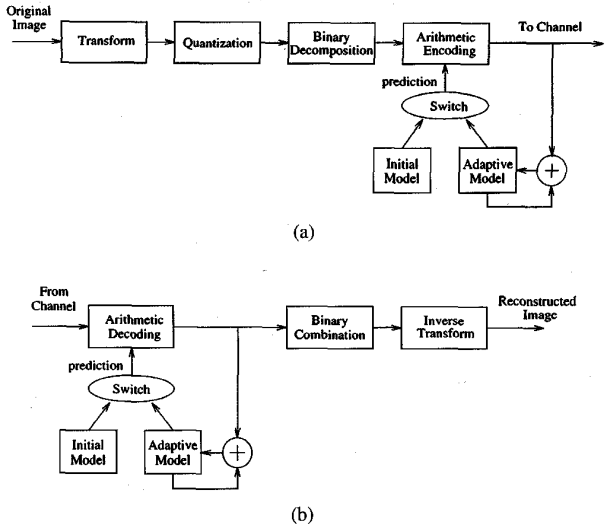


Fig. 3. Block diagram of the MBBD codec for low bit rate image compression.

encoding process takes the quantized data and produces an efficient code for them. The transformed coefficients, although less correlated than the original image pixels, still exhibit some correlation. Moreover, the quantization of the transform coefficients may actually increase spatial correlation. Further exploitation of the remaining correlations is achieved through HOEC.

The success of the MBBD approach is based on the accuracy that the probabilities of the high-order source models can be estimated. We assume that the entire MBBD output is a nonstationary source, but each subimage of the MBBD representation is a stationary sub-source. Therefore, the probability model for each subimage is estimated using only the data from that subimage.

In the codec shown, Fig. 3, the conditional pdf tables are built incrementally by including the currently encoded pixel. When starting with limited information, the accuracy of estimation is poor. Therefore, another model is maintained simultaneously, and the currently best performing model is used to make predictions. An initial model based on an average statistical model obtained from a group of training images is used when the coding begins. The algorithm for switching from the initial model to the adaptive model is based on two criteria: switch the initial model to the adaptive model if $n > T$ and $H_{ad}^{l} < H_{in}^{l}$, or $n > 2T$, where n is the number of the encoded data and T is the threshold. T defines the minimum number of data required to generate a credible adaptive model. H_{ad}^{l} is the entropy of the adaptive model for the l-most recently encoded data starting from each subimage, and H_{in}^{l} is the entropy of the initial model for the same l data. T was chosen experimentally and differs from subimage to subimage. A rule of thumb is to set T to be 1/5–1/3 of the total number of data in the entire data stream of that subimage source. l is selected to satisfy $T < l = n < 2T$.

When applying MBBD-HOEC, the selection of the scanning order of the transform coefficients results in a difference in performances. For stationary Markov models, the conditional states of a 2-d Markov source are usually defined by neighborhood templates. A description of different templates for modeling a stationary binary Markov model appears in [7]. For MBBD-HOEC coding, since the positions of the binary pixels vary unpredictably in each subimage, there is no fixed template to fit the conditional states. Therefore, instead of seeking a fixed 2-D template, we search for a 1-D scan. This scan is selected to maximize the correlation between the current pixel and its predecessors. For the block DCT data, as an example, we used a zig-zag 1-D scan within the block and a head-to-head and tail-to-tail connection between blocks [22]. For the wavelet/subband transform, the scan was determine based on the correlation direction of each band.

V. Experimental Results

A. Experiment 1

The purpose of this experiment is to verify the theoretical results obtained from BPD and MBBD representations. In this experiment, MBBD-HOEC was applied to the subband transform (ST) of a group of eight grayscale images. The size of each image is 256 × 256. The images were decomposed into 16 equal subbands. The filters used to compute the subbands were the 16-tap quadrature-mirror filter bank [23]. The transformed coefficients were quantized by a 5-b midstep uniform quantizer.

The resulting entropies from this experiment (not including the lowest frequency subband data) are summarized in Tables I and II. In Table I, we calculated the memoryless entropy and the tenth order conditional entropy of BPD representation. As expected, the memoryless entropy of the BPD representation is higher than the entropy of the original representation. Although the conditional entropy of BPD is lower than the memoryless one, no conclusion can be derived when comparing the conditional entropy to the memoryless entropy of original representation. In Table II, we calculated the memoryless entropy and tenth order conditional entropy of the MBBD representation. As expected, its memoryless entropy is exactly same as the memoryless entropy of the original representation, and the conditional entropy is lower. The entropy of a tenth order source models for the tested image data was from 12.46 to 28.26% with an average of 18.44% lower than their memoryless counterparts.

If we knew exactly the pdf's of the data, we could code them at bit-rates as low as their entropies using arithmetic coding. However, since we must estimate the pdf's during the coding process, the real coding costs are expected to be higher than the corresponding entropies. The real coding rates for the same group of test image data using a tenth order source model and a memoryless model are listed in Table III. The results show that, in practice, the coding gains obtained by using a tenth order source model are only from 7.58 to 22.22%, with an average of 13.22% lower than their memoryless counterparts. As expected, the real coding costs for both memoryless and high-order source models are higher than their theoretical counterparts. Taking the data for the Lena image as an example, we see that the practical coding cost of a zero-

TABLE I
HIGH-ORDER ENTROPY VERSUS MEMORYLESS ENTROPY FOR BPD REPRESENTATION

rate \ image	lena	scene	pepper	area	crown	couple	skull	karen
h (b/p)	1.2892	2.0200	1.2543	1.1696	0.9323	0.9256	0.4385	0.9643
H (b/p)	0.9966	1.7042	0.9920	0.9364	0.6231	0.6899	0.2954	0.7293

h: memoryless entropy of BPD source

H: conditional entropy of 10th order BPD Markov source

TABLE II
HIGH-ORDER ENTROPY VERSUS MEMORYLESS ENTROPY FOR MBBD REPRESENTATION

rate \ image	lena	scene	pepper	area	crown	couple	skull	karen	average
h (b/p)	0.9594	1.5625	0.9483	0.8688	0.6733	0.6809	0.3093	0.7049	
H (b/p)	0.7690	1.3678	0.8125	0.7293	0.4829	0.5490	0.2465	0.5856	
D (%)	19.85	12.46	14.32	16.06	28.26	19.35	20.31	16.92	18.44

h: memoryless entropy of MBBD source (bits/pel)
H: conditional entropy of 10th order MBBD Markov source (bits/pel)
D: the percentage of difference [%: 100(h-H)/h]

TABLE III
EXPERIMENTAL CODING RESULTS

rate \ image	lena	scene	pepper	area	crown	couple	skull	karen	average
r (b/p)	0.9701	1.5957	0.9700	0.8884	0.6818	0.6905	0.3127	0.7164	
R (b/p)	0.8711	1.4745	0.8715	0.7816	0.5303	0.5887	0.2633	0.6224	
D (%)	10.19	7.58	10.14	12.02	22.22	14.74	15.76	13.12	13.22

r: coding cost of memoryless MBBD source with adaptive modeling
R: coding cost of 10th order MBBD Markov source with adaptive modeling
D: the percentage of difference [%: 100(r-R)/r]

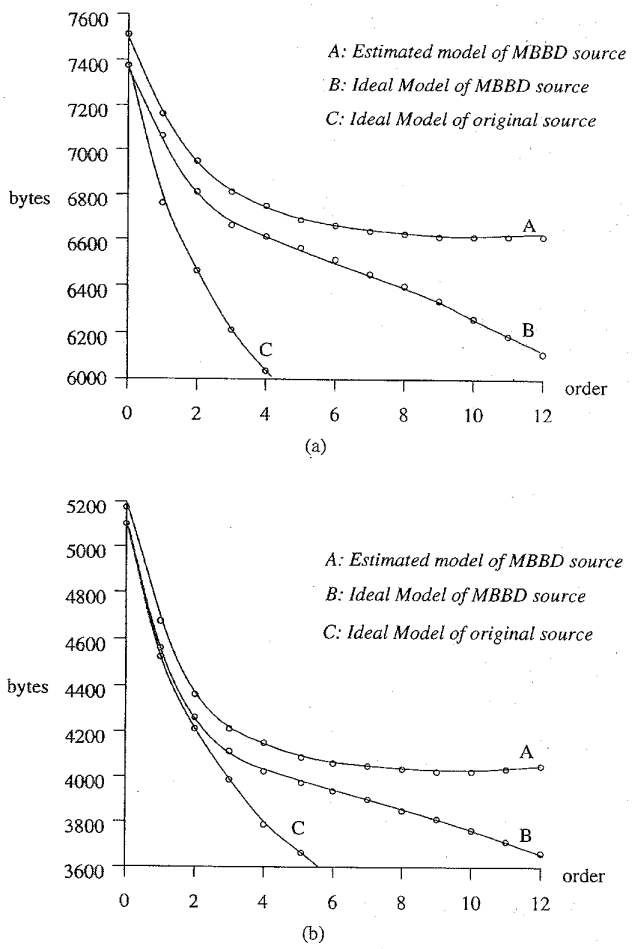


Fig. 4. Coding cost (bytes) versus order of source model. (a) Lena, (b) Crown.

order source model is 0.9701 b/pel, just slightly higher (about 1%) than its entropy 0.9594 b/pel. In contrast, the practical coding cost of tenth-order source model is 0.8711 b/pel, and it is more than 10% higher than its tenth order entropy which is 0.7690 b/pel. This indicates that the estimation of probabilities for higher order models is less accurate than those for lower ones. In fact, this is an expected result since more data can be used for the estimation of lower order probabilities.

Fig. 4 shows the coding costs for the subband data of two images, (a) Lena and (b) Crown, as functions of the different orders of the source models. Curve A is the coding cost for an estimated model of the MBBD representation. Curve B is the coding cost with an ideal model assuming that the statistics of the MBBD representation are exactly known. Curve C is the coding cost using the original representation of the coefficients and assuming that its statistics are exactly known. These results clearly show that the higher the order, the lower the coding cost, especially when the probabilities are known exactly. However, because of the inaccuracies in estimating the probabilities as the order of the model increases, the practical coding gain cannot infinitely increase with the order. Our experimental results indicate that for the Lena image, the optimum coding gain is obtained with a tenth order conditional probability model. For the Crown image, the optimum coding gain occurs with a ninth-order model. The comparison of curves C and B indicates that the data in their original representation have higher spatial correlations compared to the MBBD representation. This result supports our intuition that the MBBD representation reduces the correlation among the neighboring pixels. However, this is the penalty that has to be paid for the simplification that the MBBD offers.

From curves B and C in Fig. 4, we can make some interesting observations. For the Lena image, the second-order HOEC of the original data results in the same coding performance as the seventh-order MBBD-HOEC. For the second order HOEC of the original data, since a 5-b quantizer is used, we have $2^{2\times5}$ states each with 2^5 probabilities, thus a total of approximately 2^{15} probabilities must be estimated. In contrast, for the seventh order MBBD-HOEC we have 2^7 states each with two probabilities. We also have $2^5 - 1$ subimage planes for which we compute the probabilities separately. Thus, the total number of probabilities that must be estimated are 2^{12}. Therefore, 2^3 times more probabilities have to be estimated for potentially equal performance using the original representation. Moreover, for the MBBD-HOEC implementation the 2^{12} probabilities are estimated from $2^5 - 1$

subimage planes while in the other case 2^{15} probabilities are estimated from only one image plane. Therefore, apart from the simplification in the implementation, MBBD also allows more accurate estimation of the conditional probabilities. For the "Crown" image, the third order HOEC of the original data yields the same coding performance as the fifth order MBBD-HOEC. Using a similar analysis, the implementation MBBD-HOEC requires the estimation of approximately 2^{10} probabilities. In contrast, the implementation of HOEC of the original source requires 2^{20} probabilities. From these two examples the advantages of MBBD-HOEC become obvious.

B. Experiment 2

The purpose of this experiment is to test MBBD-HOEC for lossy transform-based image compression and progressive image transmission. We tested two compression approaches: a DCT and a WT based one. For both cases the standard 256 × 256 and 512 × 512 grayscale Lena images were used.

Case 1—WT-based Image Compression: A four-level and a five-level hierarchical wavelet-based decomposition was used for the 256 × 256 and the 512 × 512 images, respectively. The biorthogonal spline variant filters [24] were used to implement this decomposition. A uniform scalar 5-b quantizer was initially used for the WT coefficients. Then, zero-tree quantization [25] was applied to the transformed coefficients. The zero-tree quantization was aimed to achieve optimal results at a bit rate around 1.0 b/pel. All other bit-rates were obtained using progressive transmission. MBBD-HOEC was applied using different order models for every scale. A second-order model was used at the coarsest scale and an eighth-order at the finest scale level. The reason for using a relatively low order Markov model for the coefficients at the coarser scale (lower frequency) is that the number of coefficients at coarser scale are much fewer than those in finer scales. Since limited data can cause severe inaccuracies in estimating conditional probabilities, we use relatively low order models for these coefficients.

Coding performance is measured using the PSNR fidelity criterion which is defined by

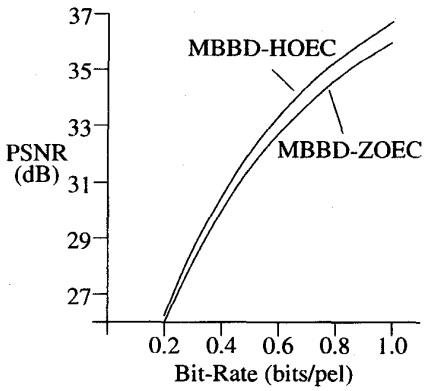
$$\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{\mathrm{MSE}} \tag{6}$$

where MSE is the mean squared error. The PSNR results at different bit-rates are plotted in Fig. 6(a) for 256 × 256 Lena image and Fig. 7(a) for 512 × 512 Lena image. The simulation result shows that hierarchical WT with zero-tree quantization can achieve good PSNR's. Most importantly, the simulations indicate that MBBD-HOEC outperforms ZOEC approach by over half a dB on average.

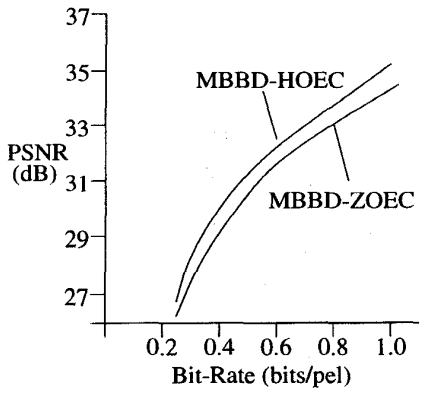
Case 2—DCT Based Image Compression: We applied 8-by-8 block DCT transform on our test images and then used the quantization matrix, shown in Fig. 5, to obtain the quantized DCT coefficients. An adaptive bit-map based on energy was used to achieve optimal results at around 1.0 b/pel. Other bit-rates were obtained using progressive transmission. PSNR results at different bit-rates of MBBD-HOEC and ZOEC are shown in Figs. 6(b) and 7(b) for the 256 × 256 and 512 × 512 Lena images, respectively. It is clear that MBBD-HOEC provides better PSNR results than those of ZOEC.

1	2	2	3	3	4	4	5
2	2	3	3	4	4	5	5
2	3	3	4	4	5	5	6
3	3	4	4	5	5	6	6
3	4	4	5	5	6	6	7
4	4	5	5	6	6	7	7
4	5	5	6	6	7	7	8
5	5	6	6	7	7	8	8

Fig. 5. DCT quantization matrix.



(a)



(b)

Fig. 6. Coding results for 256 × 256 Lena image. (a) WT-based compression and (b) DCT-based compression.

A comparison of the progressive performances using the PSNR fidelity criterion for the hierarchical WT, uniform ST from Experiment 1, and DCT based coding approaches for the 256 × 256 Lena image is shown in Fig. 8. The superior performance of the hierarchical WT coding scheme is primarily attributed to zero-tree quantization. The results in Fig. 8 are also compared with PPC based priority DCT compression [21]. This comparison shows that MBBD-HOEC based DCT outperformed PPC-based DCT progressive transmission by more than 1 dB at low bit-rates.

From this experiment we observed some interesting properties of MBBD-HOEC. We observed that the improvements

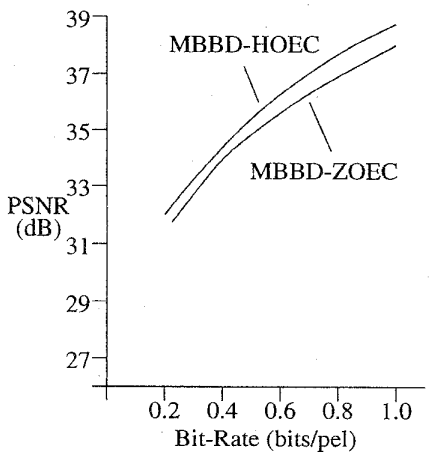


Fig. 7. Coding results for 512 × 512 Lena image. (a) WT-based compression and (b) DCT-based compression.

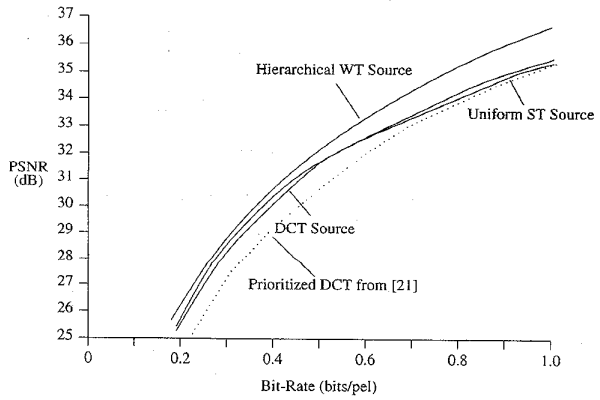


Fig. 8. MBBD-HOEC using different transformed sources of Lena image.

obtained when using the MBBD-HOEC algorithm on larger images are slightly higher than those with smaller images. The explanation of this phenomenon is that in dealing with larger images, more pixels are used to build the adaptive model. The more data used, the less zero frequency problems are encountered and the more accurate probabilities can be predicted. Another observation from these experiments is that the improvement in rate of MBBD-HOEC over MBBD-ZOEC grows as the progressive transmission continues, which means that higher improvements are achieved at higher bit-rates. This is expected since our progressive transmission process transmits the subimage represented by the largest coefficient first and then the subimages with the smaller ones. The subimages with the large coefficients are relatively less redundant, and are described and coded with relatively lower order models than those with the small coefficients. Thus, MBBD-HOEC can only achieve less improvements for these subimages.

VI. Conclusions

In this paper, motivated by the potential gains of HOEC when applied to image compression, we first examined the entropy properties of BPD decomposition. We showed that independent coding of the bit-planes increases the coding cost of the original representation. Motivated by this result, we introduced MBBD. This decomposition is a simple new binary representation that significantly reduces the complexity of implementing the HOEC, and improves the accuracy of estimating the statistical models that are required. The penalty paid for this simplification is a loss in the spatial correlation of the data. However, our experiments indicate that even with this loss MBBD-HOEC is a more efficient approach for exploiting the statistical redundancy of the sources than its ZOEC counterpart. The experiments also verified that MBBD-HOEC can be used for both lossy compression and progressive transmission of images.

Appendix A

Proof of Theorem 1: For a k b, 2^k grayscale image, its memoryless entropy is defined as

$$H(X) = -\sum_{m=0}^{2^k-1} p_m \log_2 p_m \tag{A1}$$

where p_m is the probability of pixels having value m. The entropy of BPD of this grayscale image is defined by

$$\begin{aligned} H_{BPD}(X) &= \sum_{i=0}^{k-1} H(X_i) \\ &= -\sum_{i=0}^{k-1}\sum_{j=0}^{1} P(X_i = j) \log_2 P(X_i = j) \end{aligned} \tag{A2}$$

where $H(X_i)$ is the entropy of the ith bit-plane and the probabilities for the bit-plane i $(0 \le i \le k-1)$ are given by

$$\begin{aligned} P(X_i = 0) &= \sum_{m=0}^{2^{k-1-i}-1}\sum_{n=0}^{2^i-1} p_{m2^{i+1}+n}, \\ P(X_i = 1) &= \sum_{m=0}^{2^{k-1-i}-1}\sum_{n=2^i}^{2^{i+1}-1} p_{m2^{i+1}+n}. \end{aligned} \tag{A3}$$

The previous definition of $H_{BPD}(X)$ assumes that bit-planes are statistically independent. However, if we consider the statistical dependencies among the bit-plane to be transmitted and bit-planes being transmitted, we can have the following conditional entropy

$$H_{BPD}(X|BP) = \sum_{i=0}^{k-1} H(X_i|X_{i-1}\cdots X_0). \tag{A4}$$

For each bit-plane, from information theory [14], we know

$$H(X_i) \geq H(X_i|X_{i-1}\cdots X_0) \tag{A5}$$

therefore, from (A2) and (A4) it is clear that

$$H_{BPD}(X) \geq H_{BPD}(X|BP). \tag{A6}$$

For bit-plane i, we can write

$$\begin{aligned}
&H(X_i|X_{i-1}\cdots X_0)\\
&\quad = -\sum_{X_i\cdots X_0} P(X_iX_{i-1}\cdots X_0)\\
&\qquad \cdot \log_2 P(X_i|X_{i-1}\cdots X_0)\\
&\quad = -P(X_i=0\cdots X_0=0)\\
&\qquad \cdot \log_2 P(X_i=0|X_{i-1}=0\cdots X_0=0)\cdots\\
&\qquad -P(X_i=1\cdots X_0=1)\\
&\qquad \cdot \log_2 P(X_i=1|X_{i-1}=1\cdots X_0=1).
\end{aligned} \tag{A7}$$

Using the definition of conditional probability, $P(A|B) = P(AB)/P(B)$, we get

$$\begin{aligned}
&H(X_i|X_{i-1}\cdots X_0)\\
&\quad = -\left(\sum_{n=0}^{2^{k-1-i}-1} p_{n2^{i+1}}\right)\\
&\qquad \cdot \log_2 \frac{\left(\sum_{n=0}^{2^{k-1-i}-1} p_{n2^{i+1}}\right)}{\left(\sum_{n=0}^{2^{k-i}-1} p_{n2^i}\right)}\cdots\\
&\qquad -\left(\sum_{n=0}^{2^{k-1-i}-1} p_{2^{i+1}-1+n2^{i+1}}\right)\\
&\qquad \cdot \log_2 \frac{\left(\sum_{n=0}^{2^{k-1-i}-1} p_{2^{i+1}-1+n2^{i+1}}\right)}{\left(\sum_{n=0}^{2^{k-i}-1} p_{2^i-1+n2^i}\right)}\\
&\quad = -\sum_{m=0}^{2^{i+1}-1}\left(\sum_{n=0}^{2^{k-1-i}-1} p_{m+n2^{i+1}}\right)\\
&\qquad \cdot \log_2\left(\sum_{n=0}^{2^{k-1-i}-1} p_{m+n2^{i+1}}\right)\\
&\qquad + \sum_{m=0}^{2^i-1}\left(\sum_{n=0}^{2^{k-i}-1} p_{m+n2^i}\right)\\
&\qquad \cdot \log_2\left(\sum_{n=0}^{2^{k-i}-1} p_{m+n2^i}\right).
\end{aligned} \tag{A8}$$

Combining (A4) and (A7), we have

$$\begin{aligned}
&H_{BPD}(X|BP)\\
&\quad = H(X_{k-1}|X_{k-2}\cdots X_0)\\
&\qquad + H(X_{k-2}|X_{k-3}\cdots X_0)+\cdots+H(X_0)\\
&\quad = -\sum_{m=0}^{2^k-1} p_m \log_2 p_m\\
&\qquad + \sum_{m=0}^{2^{k-1}-1} (p_m+p_{m+2^{k-1}}) \log_2 (p_m+p_{m+2^{k-1}})\\
&\qquad - \sum_{m=0}^{2^{k-1}-1} (p_m+p_{m+2^{k-1}})\\
&\qquad \cdot \log_2 (p_m+p_{m+2^{k-1}})+\cdots\\
&\qquad + \sum_{m=0}^{1}\left(\sum_{n=0}^{2^{k-1}} p_{m+2n}\right)\log_2\left(\sum_{n=0}^{2^{k-1}} p_{m+2n}\right)\\
&\qquad - \sum_{m=0}^{1}\left(\sum_{n=0}^{2^{k-1}} p_{m+2n}\right)\log_2\left(\sum_{n=0}^{2^{k-1}} p_{m+2n}\right)\\
&\qquad + \left(\sum_{n=0}^{2^k-1} p_n\right)\log_2\left(\sum_{n=0}^{2^k-1} p_n\right).
\end{aligned} \tag{A9}$$

Since most of the terms in (A7) cancel each other and since $\sum_{n=0}^{2^k-1} p_n \triangleq 1$, we have

$$\begin{aligned}
H_{BPD}(X|BP) &= -\sum_{m=0}^{2^k-1} p_m \log_2 p_m\\
&= H(X).
\end{aligned} \tag{A10}$$

From (A6) and (A10), we have $H_{BPD}(X) \geq H(X)$, which proves Theorem 1.

APPENDIX B

Proof of Theorem 2: We start our proof by introducing some notation. Let P_I represents the probability of information pixels in the Ith MBBD image plane. $P_{\overline{I}}$ represents the probability of noninformation pixels in the Ith MBBD image plane. N_I is the number of pixels to be coded in the Ith MBBD image plane. C(MBBD) is the total coding cost, at the low bound, of the MBBD representation. For the Ith MBBD image plane, the probabilities are given by

$$P_I = \frac{p_I}{\sum_{i=I}^{M-1} p_i}$$

and

$$P_{\overline{I}} = \frac{\sum_{i=I+1}^{M-1} p_i}{\sum_{i=I}^{M-1} p_i} \tag{B1}$$

where $P_I + P_{\overline{I}} = 1$. The entropy of the Ith MBBD image plane is given by

$$H_I = -P_I \log_2 P_I - P_{\overline{I}} \log_2 P_{\overline{I}}. \tag{B2}$$

Then the total coding cost at the low bound for the MBBD image representation is

$$\begin{aligned} C(MBBD) &= \sum_{I=0}^{M-1} N_I H_I \\ &= N \sum_{I=0}^{M-1} \left[\left(\sum_{i=I}^{M-1} p_i \right) \right. \\ &\quad \left. \cdot (-P_I \log_2 P_I - P_{\overline{I}} \log_2 P_{\overline{I}}) \right]. \end{aligned} \tag{B3}$$

The entropy of the MBBD representation is given by

$$\begin{aligned} H_{MBBD} &\triangleq \frac{C(MBBD)}{N} \\ &= \sum_{I=0}^{M-1} \left[\left(\sum_{i=I}^{M-1} p_i \right) \left(-\frac{p_I}{\sum_{i=I}^{M-1} p_i} \log_2 \frac{p_I}{\sum_{i=I}^{M-1} p_i} \right. \right. \\ &\quad \left. \left. - \frac{\sum_{i=I+1}^{M-1} p_i}{\sum_{i=I}^{M-1} p_i} \log_2 \frac{\sum_{i=I+1}^{M-1} p_i}{\sum_{i=I}^{M-1} p_i} \right) \right]. \end{aligned} \tag{B4}$$

After some simplification in (B4), we get

$$\begin{aligned} H_{MBBD} = \sum_{I=0}^{M-1} \left[-p_I \log_2 \frac{p_I}{\sum_{i=I}^{M-1} p_i} \right. \\ \left. - \left(\sum_{i=I+1}^{M-1} p_i \right) \log_2 \frac{\sum_{i=I+1}^{M-1} p_i}{\sum_{i=I}^{M-1} p_i} \right]. \end{aligned} \tag{B5}$$

Unfolding this equation by using $\log (A/B) = \log A - \log B$, we have

$$\begin{aligned} H_{MBBD} &= -\sum_{I=0}^{M-1} p_I \log_2 p_I - \sum_{I=1}^{M-1} \\ &\quad \cdot \left[\left(\sum_{i=I}^{M-1} p_i \right) \log_2 \sum_{i=I}^{M-1} p_i \right. \\ &\quad \left. - \left(\sum_{i=I}^{M-1} p_i \right) \log_2 \sum_{i=I}^{M-1} p_i \right] \\ &= -\sum_{i=0}^{M-1} p_i \log_2 p_i \\ &\triangleq H \end{aligned} \tag{B6}$$

which proves Theorem 2.

REFERENCES

[1] D. A. Huffman, "A method for the construction of minimum redundancy codes," *Proc. IRE,* vol. 40, pp. 1098–1101, Sept. 1952.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.,* vol. 27, no. 3, pp. 379–423, July 1948.

[3] R. E. Blahut, *Principles and Practice of Information Theory.* Reading, MA: Addison-Wesley, 1987.

[4] G. Langdon, "An introduction to arithmetic coding," *IBM J. Res. Develop.,* vol. 28, no. 2, pp. 135–149, 1984.

[5] I. H. Witten, R. M. Neal, and J. G. Clearly, "Arithmetic coding for data compression," *Commun. ACM,* vol. 30, pp. 520–540, June 1987.

[6] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory,* vol. IT-24, no. 5, pp. 530–536, Sept. 1978.

[7] G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Trans. Commun.,* vol. COM-29, no. 6, pp. 858–867, June 1981.

[8] H. Gharavi, "Conditional run-length and variable-length coding of digital pictures," *IEEE Trans. Commun.,* vol. COM-35, pp. 671–677, 1987.

[9] S. E. Elnahas, K. H. Tzou, and J. G. Dunham, "Low-rate entropy coding of transformed images," in *Proc. Applications of Digital Image Processing IX,* A. G. Tescher, Ed., vol. SPIE-697, pp. 10–13, Aug. 1986.

[10] K. H. Tzou, "High-order entropy coding for images," *IEEE Trans. Circuits Syst Video Technol.,* vol. 2, no. 1, pp. 87–89, Mar. 1992.

[11] S. M. Lei, M. T. Sun, and K. H. Tzou, "Design and hardware architecture of high-order conditional entropy coding for images," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 2, no. 2, pp. 176–187, June 1992.

[12] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.,* vol. COM-28, no. 1, pp. 84–95, Jan. 1980.

[13] Y. Bar–Ness and S. Choi, "High order Markov arithmetic encoding for data compression," in *25th Allenton Conf. Commun., Contr. Comp.,* Sept. 1987.

[14] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[15] J. W. Schwartz, "Data processing in scientific space probes," Ph.D. dissertation, Dept. of EE, Yale University, New Haven, CT, Sept. 1963.

[16] J. W. Schwartz and R. C. Barker, "Bit-plane encoding: A technique for source encoding," *IEEE Trans. Aerosp. Electron. Syst.,* vol. AES-2, no. 4, pp. 385–392, July 1966.

[17] Rabani and Jones, *Digital Image Compression Techniques,* vol. SPIE-TT7, 1991.

[18] S. L. Tanimoto, "Image transmission with gross information first," *Comput. Graphic Image Process.,* vol. 9, pp. 72–74, Jan. 1979.

[19] K. Knowlton, "Progressive transmission of grey scale and binary pictures by simple, efficient, and lossless encoding scheme," *Proc. IEEE,* vol. 68, pp. 885–896, July 1980.

[20] K. H. Tzou, "Progressive image transmission: A review and comparison of techniques," *Optical Eng.,* vol. 26, no. 7, pp. 581–589, July 1987.

[21] Y. Huang, H. M. Dreizen, and N. P. Galatsanos, "Prioritized DCT for compression and processive transmission of images," *IEEE Trans. Image Processing,* vol. 1, no. 4, pp. 477–487, Oct. 1992.

[22] S. S. Yu, "Adaptive compression of images using high-order entropy coding," Ph.D. thesis, IIT, 1994.

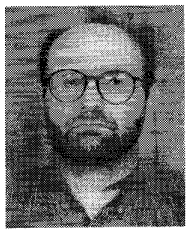
[23] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. 1980 IEEE Int. Conf. ASSP,* Apr. 1980, pp. 291–294.

[24] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform" *IEEE Trans. Image Processing,* vol. 1, no. 2, pp. 205–220, Apr. 1992.

[25] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing,* vol. 41, no. 12, pp. 3445–3479, Dec. 1993.

Steve S. Yu (S'85–M'90) was born in Shanghai, China, in 1956. He received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1982, and the M.S. degree in physics from Wayne State University, Detroit, MI, in 1984, and the M.S. and Ph.D. degrees in electrical and computer engineering from Illinois Institute of Technology, Chicago, in 1985 and 1994, respectively.

Since 1989, he has been working in the AT&T Bell Laboratories as a Member of Technical Staff. His research interests include image/video compression, signal processing, and wireless communication.



Nikolas P. Galatsanos (S'84–M'85–SM'95) was born in Athens, Greece, in 1958. He received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1982. He received the M.S. and the Ph.D. degrees both in electrical and computer engineering from the University of Wisconsin-Madison in 1984 and 1989, respectively.

Since August of 1989, he has been on the faculty of the Department of Electrical and Computer Engineering at the Illinois Institute of Technology, Chicago, IL, where he is currently an Associate Professor. His current research interests include recovery and compression problems for visual communications and medical imaging applications.

Dr. Galatsanos is a member the Technical Chamber of Greece. He also served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING.