AN IMAGE CODING SCHEME USING BLOCK PREDICTION OF THE PYRAMID SUBBAND DECOMPOSITION

R. Rinaldo and G. Calvagno

Dipartimento di Elettronica e Informatica
Via Gradenigo 6/a, 35131 Padova, Italy
Tel: +39-49-828 7735, Fax: +39-49-828 7699

ABSTRACT

In this work, we propose a hybrid vector quantization scheme for general images that aims at exploiting similarities among detail signals in a multiresolution decomposition of the image. In a multiresolution decomposition, the image is decomposed into a set of subbands which are localized in scale, orientation and space. Our coding scheme consists of dividing each subimage of the multiresolution decomposition into square *range* blocks. The range blocks are matched against *domain* blocks chosen in the lower resolution subimage with the same orientation, and coded through a description of the map transforming the domain block into the range block. The pool of of domain blocks acts as a codebook for the range block, as in vector quantization, with the difference that the codebook is built from blocks inside the multiresolution decomposition. If the prediction procedure is not satisfactory with respect to a target quality, the block is coded using a geometric vector quantizer for laplacian random variables.

1. INTRODUCTION

Pyramid subband decomposition provides a multiresolution representation of images that has been used by several authors for image analysis and coding [1, 2]. The redundancy of the multiresolution representation has been fully recognized for fractal images and it has also been efficiently exploited for coding of general images [2, 3, 4].

In this work, we propose a coding procedure that uses a hybrid scheme of vector quantization for the coefficients of a multiresolution subband decomposition of the input image. The coding scheme consists of partitioning each subimage of the multiresolution decomposition into a set of nonoverlapping *range* blocks. The range blocks are matched against possibly overlapping *domain* blocks of the same dimension taken from the adjacent lower resolution subimage with the same orientation. The scheme efficiently exploits the redundancy of the multiresolution decomposition since range blocks are *predicted* from domain blocks. Thus, the pool of domain blocks acts as a *codebook* for the range block, as in vector quantization, with the important difference that the codebook is built from blocks *inside* the subband decomposition itself. To enlarge the codebook size, domain blocks are multiplied by a constant (i.e., scaled in amplitude) and possibly isometrically transformed before matching (i.e., reshuffling of the block coefficients is carried out) [5]. A range block of a subimage at a certain scale can be coded by giving the relative position of its matched domain block in the next lower resolution subimage with the same orientation, together with the resulting scaling factor and an identifier for the isometric transformation. To provide initial conditions to the coder, the four subbands with the lowest resolution are coded using a pyramid vector quantizer (PVQ) [6]. The block prediction procedure is then applied to the blocks of finer resolution subimages in a causal fashion, i.e., going from low to high resolution subbands. Whenever the block matching procedure is not satisfactory with respect to a target coding quality, the non-predicted range block is coded using PVQ. Thus, our coder uses a hybrid scheme of VQ, with the codebook built from blocks inside the subband decomposition, and geometric VQ.

The improvement over the technique described in [4] is significant. Specifically, the domain block search region for a range block at relative position (j, k) in one subimage, consists now of the blocks centered at position $(j/2, k/2)$ in the adjacent lower resolution subband with the same orientation. This reduces considerably the coding time. Moreover, the use of adaptive geometric PVQ for the non predicted blocks gives improved results in terms of achieved compression ratios.

2. PRINCIPLES OF SUBBAND CODING

Subband decomposition for images has first been introduced by Woods and O'Neil [7]. A popular scheme for two-dimensional subband decomposition is based on separable filters, as shown in Figure 1. Subband x^{ij}, $i, j = 0, 1$, is obtained by first filtering the columns of the input image with $H_i(z)$ and subsampling by a factor two, followed by filtering the rows with $H_j(z)$ and subsampling. In the scheme of Figure 1, $H_0(z)$ is a low-pass filter, while $H_1(z)$ has high-pass characteristics: thus, each subband x^{ij} has one fourth of the input image coefficients and is relative to details that pertain to different frequency regions of the input image spectrum. Assuming no coding error, the reconstruction is performed at the receiver by upsampling the rows of x^{ij}, and filtering with $G_i(z)$, followed by upsampling the columns and filtering with $G_j(z)$. The outputs of the four upsample/filter sections are summed together to give the reconstructed image $\tilde{x}$. It is possible to show that in the absence of coding errors, perfect reconstruction of the input image x can be obtained by appropriate design of the analysis filters $H_i(z)$ and synthesis filters $G_i(z)$ [8].

Paraunitary systems are perfect reconstruction systems

where the filter bank can be obtained by designing a single filter $H_0(z)$. In fact, one possible way to achieve perfect reconstruction is to choose the analysis/synthesis filter coefficients such that

$$\begin{aligned} g_0(n) &= h_0(-n) \\ g_1(n) &= h_1(-n) \\ h_1(n) &= (-1)^{1-n} h_0(1-n), \end{aligned} \tag{1}$$

where $h_0(n)$ is a filter whose z-transform $H_0(z)$ satisfies

$$H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) = 2. \tag{2}$$

Note that relation (2) implies that the filter $h_0(n)$ and its even translations form an orthonormal family, and that the two filters $h_0(n)$ and $h_1(n)$ are orthogonal [8].

In [9, 10] it is shown that orthonormal bases of wavelets correspond to a subband coding scheme with orthogonal filters satisfying equations (1) and (2). Considering the case of FIR filters, it is possible to show that conditions (1) and (2) can be met exactly only for even-length filters. Furthermore, only trivial solutions with linear phase are possible. Approximate solutions with odd length linear phase FIR filters are however possible. A design method for linear phase FIR analysis and synthesis filters can be found in [11]. A consequence of the linear phase constraint is that relation (2) can only be satisfied approximately. More general solutions for exact perfect reconstruction even-length, odd-length FIR linear phase filters are given by biorthogonal systems, where the orthogonality conditions (1) and (2) are relaxed.

The basic separable two-dimensional scheme of Figure 1 can be used to obtain finer decompositions of the input spectrum. In a multiresolution scheme, the decomposition of Figure 1 is iterated to obtain a multiresolution decomposition of $x(m,n)$. Denoting with x_0^{00} the input image $x(m,n)$, at each decomposition level l the image x_{l-1}^{00} is decomposed into the four subimages x_l^{00}, x_l^{01}, x_l^{10}, x_l^{11}, for $l = 1, \ldots, l_M$. The result of such a decomposition is a set of subimages which are localized in scale, orientation and space.

3. ADAPTIVE GEOMETRIC VECTOR QUANTIZATION

A fundamental result of rate-distortion theory is that it is possible to achieve better performance in terms of average distortion for a given output rate r in bits per sample by coding vectors of input data instead of scalars [12, 13]. Interestingly enough, this is true even in the case of a source that generates independent variables. This observation explains the interest of vector quantization for data compression, particularly in image and speech coding for communication or storage on digital links and media.

VQ is a non linear map $Q(\cdot)$ from a multidimensional input space $\mathbb{R}^L$ into a discrete output set of representative points. Thus, each L-dimensional vector $\mathbf{x}$ of input data is mapped into one of $K = 2^{rL}$ output vectors $\mathbf{y}_1, ..., \mathbf{y}_K$. Since r bit per dimension are sufficient to specify the output vector $\mathbf{y}_i = Q(\mathbf{x})$, compression is obtained at the expense of an average distortion

$$\mathrm{E}[d(\mathbf{x}, Q(\mathbf{x}))], \tag{3}$$

where $d(\mathbf{x}, \hat{\mathbf{x}})$ represents the cost of reproducing any vector $\mathbf{x}$ as the reproduction vector $\hat{\mathbf{x}}$ and E[·] denotes statistical expectation.

To specify a vector quantizer, one needs the set of output vectors $\mathbf{y}_1, ..., \mathbf{y}_K$, (or *codebook*) and the specification of the rule $Q(\cdot)$ to map the input vector to one of the output vectors. This rule, in turn, implies a partition $R_1, ..., R_K$, of $\mathbb{R}^L$, where

$$R_i = Q^{-1}(\mathbf{y}_i) = \{\mathbf{x} : Q(\mathbf{x}) = \mathbf{y}_i\}. \tag{4}$$

Determining the output vectors and the partition $R_1, ..., R_K$, in order to minimize the distortion (3) is a non-trivial task, and several approaches to the solution of this problem have been proposed in the literature [13, 14].

Geometric quantizers exploit the geometric properties of memoryless sources of random variables. Let $f_X(a)$ be the one-dimensional probability density function of a random variable X. Then any L-dimensional vector $\mathbf{x}$ of independent and identically distributed (iid) variables drawn according to $f_X(a)$ has a joint probability density function

$$f_{\mathbf{x}}(\mathbf{a}) = \prod_{i=1}^{L} f_X(a_i). \tag{5}$$

It is possible to show [12], that for large L the vector $\mathbf{x}$ belongs with high probability to the *typical set* defined as

$$S_\epsilon = \{\mathbf{a} : 2^{-L(h(X)-\epsilon)} < f_{\mathbf{x}}(\mathbf{a}) < 2^{-L(h(X)+\epsilon)}\}, \tag{6}$$

where

$$h(X) = -\int f_X(a) \log_2 f_X(a) da$$

denotes the differential entropy of the scalar random variable X. Thus, one can conclude from (6) that, for large L, vectors of iid random variables drawn according to $f_X(a)$ concentrate with high probability in a subset of $\mathbb{R}^L$ where

$$f_{\mathbf{x}}(\mathbf{a}) \simeq 2^{-Lh(X)}. \tag{7}$$

Therefore, for random vectors of iid variables, the reproduction points of a vector quantizer should be uniformly located on the set S_ϵ, which can be determined once the probability density $f_X(a)$ is known.

Based on these results, Fischer [15] designed a *pyramid* vector quantizer for L-dimensional vectors of iid laplacian random variables. It turns out that the relevant volume for quantization is located around the L-dimensional hyperpyramid defined by

$$\sum_{i=1}^{L} |a_i| = \frac{L}{\lambda}. \tag{8}$$

In [6] a PVQ has been presented to code vectors of subband coefficients which have approximately a laplacian distribution [7]. The PVQ codebook vectors are determined by the intersection of a cubic lattice with a scaled L-dimensional pyramid. The pyramid is selected to be close to the nominal one, defined by (8), on the basis of the value

$$||\mathbf{x}||_1 = \sum_{i=1}^{L} |X_i| \tag{9}$$

computed from the actual components of each input vector $\mathbf{x}$. A product gain-shape code is used, in which $||\mathbf{x}||_1$ is coded using a Lloyd-Max gaussian quantizer [16] to index the actual pyramid, and the remaining Lr bits are used to identify the lattice point on the pyramid. It can be shown that the mean squared error distortion obtained with such PVQ is

$$D_{PVQ}(r) \simeq \frac{e^2}{3\lambda^2} 2^{-2r}. \tag{10}$$

Therefore, given a nominal variance $\sigma^2 = 2/\lambda^2$ of the subband coefficients and the rate rL in bits per vector, the PVQ computes the output vector among the $N \leq 2^{rL}$ points in the intersection of the cubic lattice with an L-dimensional hyper-pyramid determined on the basis of the actual value of $||\mathbf{x}||_1$. At the decoder, we need the nominal variance, a code for $||\mathbf{x}||_1$ and rL bits per vector of input coefficients.

An adaptive version of the PVQ is obtained by imposing a target distortion and determining the rate r_k for each input vector $\mathbf{x}_k$ using relation (10) and the sample variance

$$\sigma_k^2 = \sum_l \mathbf{x}_k^2(l). \tag{11}$$

The rate r_k is then uniformly quantized. The vector of block coefficients is fed into the PVQ using a nominal value for the variance that is recalculated with (10) from the quantized r_k. Thus, besides the $N_r + r_k L$ bits for each input vector, at the decoder we need side information about the quantized values r_k.

4. DESCRIPTION OF THE CODER

In this section we describe in some detail the proposed coder. The description is relative to 512×512 images, but it can be easily adapted to different sizes of the input image. The actual coding is performed on the coefficients of a 5 level multiresolution subband decomposition that is computed using the 24-tap linear phase nearly orthogonal and quasi-perfect reconstruction filter bank of [11]. The image is symmetrically extended before filtering to avoid border effects.

Given a target value P for the reconstruction mean squared error, we subtract from the 16×16 low-pass subband x_5^{00} its mean value, and subdivide x_5^{00} into four 8×8 blocks. The mean value is coded separately using an 8 bit uniform quantizer. Each 8×8 block is transformed using the Discrete Cosine Transform (DCT). As known, the transform coefficients have approximately a laplacian distribution, although their variance depends on the coefficient position inside the block [16]. In order to build vectors of DCT coefficients with the same variance, we split the 8×8 blocks into four 4×4 subblocks, and group the coefficients of the subblocks located in corresponding positions inside the 8×8 blocks, thus originating four vectors of L=64 coefficients. These vectors are coded using adaptive PVQ as explained in section 3. Using 64 coefficient vectors should guarantee performance close to the asymptotic values. Similarly, subbands $x_5^{01}, x_5^{10}, x_5^{11}$ are divided into four 8×8 block vectors that are PVQ coded directly. The imposed distortion is $0.5P$ and the number of bits for $||\mathbf{x}||_1$ is $N_r = 6$ in both cases.

Quantized subbands at level 5 provide the codebook for the prediction of subbands at level 4. Specifically, subimage x_4^{01} (or x_4^{10}, x_4^{11}) is divided into non overlapping 4×4 range blocks b_k. A domain block d_h of the same dimension is searched in x_5^{01} (or x_5^{10}, x_5^{11}) to minimize the distance

$$D(b_k, d_h) = \sum_l \sum_m (b_k(l,m) - \tau_k[d_h](l,m))^2 \tag{12}$$

between the range block and an appropriately transformed domain block $\tau_k[d_h]$. Here, l, m, indicate the coefficient position inside the block. The transformation τ_k has the form

$$\tau_k[d_h] = \alpha_k(\mathcal{I}_k[d_h]), \tag{13}$$

where α_k is a multiplicative scaling factor, and $\mathcal{I}_k$ is one of the possible four rotations of the domain block coefficients, chosen to minimize (12). The transformations are similar to those proposed for fractal block coders [5].

For each of the four rotations, the optimal scaling parameter is readily obtained by setting to zero the derivative of $D(b_k, \tau_k[d_h])$, namely

$$\alpha_k^o = \frac{\sum_{l,m} b_k(l,m) d_h(l,m)}{\sum_{l,m} d_h^2(l,m)}. \tag{14}$$

The scaling parameters are quantized using a uniform quantizer followed by entropy coding [16], in which the quantization step is proportional to the square root of the target mse P.

The 4×4 blocks in x_5^{01} considered for matching of a range block at relative position (j,k) in x_4^{01}, are located in a subregion centered at the corresponding position $(j/2, k/2)$ in x_5^{01}. Three bits per direction are used to specify the address of the domain block relative to this position.

If the prediction is not satisfactory and gives an mse greater than P, the 4×4 block is coded by using adaptive PVQ. Four consecutive non predicted 4×4 blocks are grouped together to form a vector of L=64 coefficients. Also in this case, we use $N_r = 6$ bits for $||\mathbf{x}||_1$. As explained in section 3, the rate r_k for a block is quantized using 0.25 bit steps. The resulting output symbols are entropy coded and sent as side-information to the decoder.

The procedure is repeated for subbands at level 3. In this case, we consider two possible range block dimensions, namely 4×4 and 8×8. Again, domain blocks are searched around the corresponding position in x_4^{01} (or x_4^{10}, x_4^{11}) and six bits in total are used to specify the domain block position. Prediction is considered first for 8×8 blocks, which are split into four 4×4 subblocks whenever the minimized distortion gives an mse greater than P. If the prediction of a 4×4 block is still not satisfactory, it is coded using adaptive PVQ.

Subbands at level 2 and 1 are coded similarly, with the only difference given by the largest allowed dimension for range blocks. In Table 1 we report the range block dimensions considered for the subbands at the various levels. Note that PVQ is performed in all cases only for those 4×4 blocks that could not be predicted.

The scaling parameter α_k is set to zero whenever the variance of the range block is less than $0.5P$. The blocks corresponding to the same spatial location at finer scales

with the same orientation are also tested to determine if their variance is negligible. In such a case, these blocks are not coded at all. The first range block with $\alpha_k = 0$ is encoded with a special symbol indicating the insignificance of all the corresponding blocks in finer scales. The technique is similar to that described in [3].

Additional bits per block are needed to specify the block type in order to correctly decode the stream of bits relative to each block. For the subbands at level 4, one bit per block is sufficient to distinguish between the predicted and PVQ coded blocks. Each 8×8 block at level three can be split into four 4×4 blocks or predicted as it is. One bit is sufficient to distinguish between these two possibilities. Only in the case the block is split, one bit for each of the resulting 4×4 blocks is needed to distinguish between the prediction and PVQ alternatives. Thus, additional bits are used only when necessary.

The same strategy is applied to classify blocks of subimages at levels two and one of the multiresolution decomposition. Starting from the largest allowed block dimension, one bit specifies if the block is predicted as it is or split. In case it is split, the four originated blocks are recursively classified with one bit indicating prediction or further splitting, for the intermediate block sizes, and prediction or PVQ, for the smallest block size.

5. RESULTS

In this section, we present some experimental results to evaluate the performance of the proposed hybrid vector quantization coding scheme. The used test images are 512×512 gray-level, coded with 8 bits per pixel (bpp). The objective image reconstruction quality was determined by evaluating the peak-signal-to-noise ratio (PSNR).

Figure 2 shows the PSNR versus the bit rate relative to the coding of the test image "Lenna". In the same plot, the results of the proposed coding method are compared with those obtained by using the coder presented in [4] (PPC) and with the popular JPEG coding system. The plot shows that the hybrid VQ coder performs better over the entire range of bit rates, with an improvement in PSNR that is almost independent of the bit rate. The curve labeled PVQ in the plot is relative to the application of adaptive PVQ alone to all the subband coefficients. Specifically, the subbands are divided into 8×8 blocks that are coded using PVQ with 6 bits for $||\mathbf{x}||_1$. Note that the performance of our scheme is similar to that of PVQ at a rate of about 0.4 bpp. As a matter of fact, when the prediction procedure gives poor results, our scheme reduces to adaptive PVQ of subband coefficients.

The visual quality obtainable with our coder can be appreciated by comparing the original image "Lenna" shown in Figure 3 with the image coded at .25 bpp shown in Figure 4. As it may be seen, no blocking effect can be noticed in figure 4, even though some artifacts and smearing can be detected. Ringing effects, typical of subband coded images, are also negligible.

The computation time required to code "Lenna" at .25 bpp was about 2 minutes on a Sun SPARCclassic (this time does not include the computation of the pyramid subband decomposition). Coding time figures that were obtained for different bit rates and other images are similar. Notice that the coding time may vary depending on the image reconstruction fidelity: when the target quality required is high, (i.e., for small values of the target mse P) the block prediction procedure resorts more often to block splitting and a greater number of blocks are coded with PVQ. We remark here that the value of the parameter P is not directly related to the actual mse that will be obtained after coding.

6. REFERENCES

[1] P.J. Burt, E.H. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Trans. on Communications*, pp. 532–540, April 1983.

[2] A. Pentland, B. Horowitz, "A Practical Approach to Fractal-Based Image Compression," in *Proc. Data Compression Conference*, pp. 176–185, 1991.

[3] J.M. Shapiro, "An Embedded Wavelet Hierarchical Image Coder," in *Proc. IEEE ICASSP*, Vol. IV, pp. 657–660, March 1992.

[4] R. Rinaldo and G. Calvagno, "Image Coding by Block Prediction of Multiresolution Subimages," submitted to the *IEEE Trans. on Image Processing*.

[5] A. E. Jacquin, "Fractal image coding: a review," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1451–1465, Oct. 1993.

[6] M.E. Blain and T.R. Fischer, "A Comparison of Vector Quantization Techniques In Transform an Subband Coding of Imagery," *Signal Processing: Image Communication*, pp. 91–105, Vol. 3, 1991.

[7] J.W. Woods and S.D. O'Neil, "Subband coding of images," *IEEE Trans. on Acoustics Speech and Signal Processing*, vol. 34, no. 5, Oct. 1986.

[8] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, A.V. Oppenheim Ed., 1993.

[9] S. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, pp. 2091–2110, Dec. 1989.

[10] A. Cohen, I. Daubechies, J.C. Feauveau, "Biorthogonal Bases of Compactly Support Wavelets," *Communications on Pure and Applied Mathematics*, pp. 485–560, June 1992.

[11] J.D. Johnston, "A Filter Family Designed for Use in Quadrature Mirror Filter Banks," in *Proc. IEEE ICASSP*, pp. 291–294, April 1980.

[12] T.M. Cover and J.A. Thomas, *Information Theory*, New york: Wiley, 1991.

[13] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Boston, MA: Kluwer, 1992.

[14] Y. Linde, A. Buzo and R.M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Communications*, pp. 84–95, Jan. 1980.

[15] T.R. Fischer, "A Pyramid Vector Quantizer," *IEEE Trans. on Information Theory*, pp. 568–583, July 1986.

[16] N.S. Jayant, P. Noll, *Digital Coding of Waveforms*, Englewood Cliffs, NJ: Prentice Hall, 1984.

Subband	4×4 PVQ	4×4 Pred.	8×8	16×16	32×32
$x_4^{01}, x_4^{11}, x_4^{10}$	$\surd$	$\surd$			
$x_3^{01}, x_3^{11}, x_3^{10}$	$\surd$	$\surd$	$\surd$		
$x_2^{01}, x_2^{11}, x_2^{10}$	$\surd$	$\surd$	$\surd$	$\surd$	
$x_1^{01}, x_1^{11}, x_1^{10}$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$

Table 1: Summary of the range block dimensions.

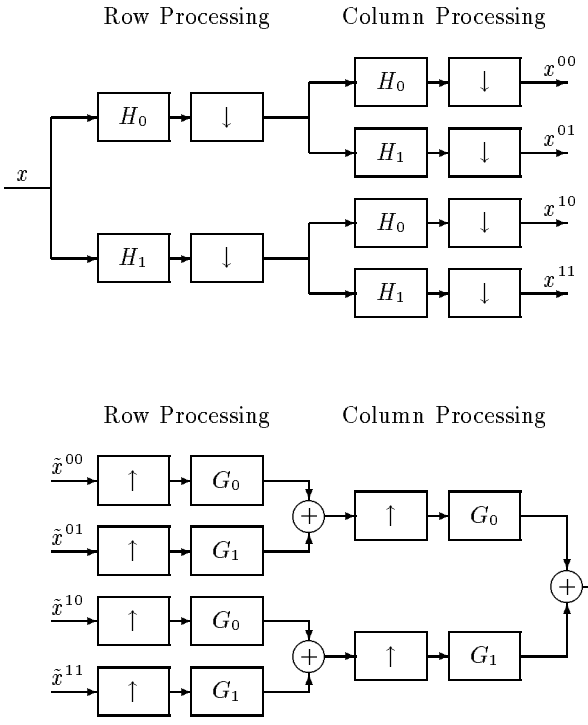

Figure 1: Separable 2D scheme for subband decomposition.

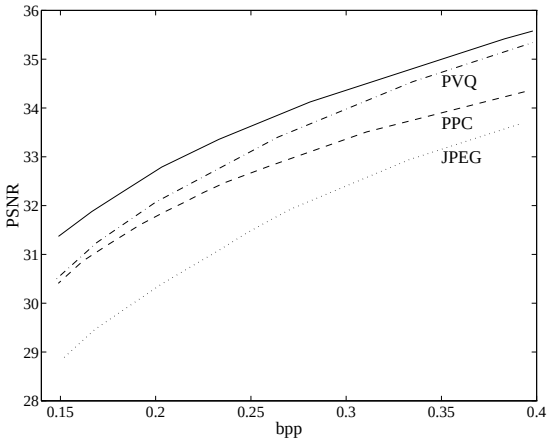

Figure 2: PSNR versus bpp for the test image "Lenna": comparison of the hybrid VQ vith JPEG, PPC and PVQ.

Figure 3: Original test image "Lenna".

Figure 4: Reconstructed image "Lenna" at .25 bpp.