MULTICHANNEL LINEAR PREDICTIVE CODING OF COLOR IMAGES*

Petros A. Maragos, Russell M. Mersereau, and Ronald W. Schafer

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

ABSTRACT

This paper reports on a preliminary study of applying single-channel (scalar) and multichannel (vector) 2-D linear prediction to color image modeling and coding. Also, the novel idea of a multi-input single-output 2-D ADPCM coder is introduced. The results of this study indicate that texture information in multispectral images can be represented by linear prediction coefficients or matrices, whereas the prediction error conveys edge-information. Moreover, by using a single-channel edge-information we obtained, from original color images of 24 bits/pixel, reconstructed images of good quality at information rates of 1 bit/pixel or less.

INTRODUCTION

Two-dimensional linear prediction was successfully applied to coding monochrome images at rates below 1 bit/pixel [1,2] and to clustering homogeneous image textures by using 2-D LPC distances [3]. Motivated by the above success of 2-D linear prediction, we tried to extend its use to multispectral images either by autoregressively modeling each channel separately or by using a vector 2-D linear predictor which exploits cross-correlation between channels. These two approaches ressemble the notions of component and composite encoding methods for color video signals [4]. A major contribution of this paper is the introduction of a multi-input single-output ADPCM coder whose output will be a single-channel edge-information signal; this reflects the idea that for most natural color images the edges occur at approximately the same location in every channel. Although our results refer only to 3-channel color images (red, green, blue), our theoretical formulation addresses the general case of an N-channel multispectral image.

MULTICHANNEL 2-D LINEAR PREDICTION

Let $\mathbf{x}(m,n) = [x_1(m,n),\ldots,x_N(m,n)]^T$ represent an N-channel 2-D image vector signal, where $[\cdot]^T$ denotes the transpose of a vector and $x_i(m,n)$ represents a single-channel scalar 2-D sequence of image intensity in a certain spectral band. By exploiting the autocorrelation of every channel and the cross-correlation between channels, we formulate the following 2-D vector autoregressive model for $\mathbf{x}(m,n)$:

$$\mathbf{x}(m,n) = \sum_k \sum_\ell \mathbf{A}(k,\ell)\mathbf{x}(m-k,n-\ell) + \mathbf{b} + \mathbf{e}(m,n) \quad (1)$$

where we predict the vector $\mathbf{x}(m,n)$ from its neighbor vector values weighted by "prediction matrices" $\mathbf{A}(k,\ell)$ of order $N\times N$. In (1), (k,ℓ) range over all integer pairs in a set Π, called the region of support of the prediction mask, and this set determines whether the mask is causal, quarter-plane, etc. The causality of the prediction mask is necessary for the recursive computability of (1). The bias vector $\mathbf{b} = [b_1,\ldots,b_N]^T$ accounts for the fact that the intensity image samples are explicitly biased by a dc-level vector $\mathbf{d} = [d_1,\ldots,d_N]^T$ since they are always nonnegative. The 2-D vector prediction error signal $\mathbf{e}(m,n)$ is the output of a $N\times N$ matrix prediction error filter

$$\mathbf{F}(z_1,z_2) = \mathbf{I} - \sum_k \sum_\ell \mathbf{A}(k,\ell) z_1^{-k} z_2^{-\ell} \quad (2)$$

when the input is $\mathbf{x}(m,n)$ and where $\mathbf{I}$ denotes the $N\times N$ identity matrix. The relation between $\mathbf{b}$ and $\mathbf{d}$ is

$$\mathbf{b} = \Big[\mathbf{I} - \sum_k \sum_\ell \mathbf{A}(k,\ell)\Big]\mathbf{d} \quad (3)$$

Consider the $N\times N$ average prediction error matrix

$$\mathbf{E} = \sum_m \sum_n \mathbf{e}(m,n)\mathbf{e}^T(m,n) \quad (4)$$

In (4), (m,n) range over all integer pairs corresponding to pixel locations inside some region of support of $\mathbf{x}(m,n)$ which we call the <u>analysis frame</u>. The i-th diagonal entry of the matrix $\mathbf{E}$ represents the mean-squared prediction error in the i-th channel. The criterion to find the optimal parameters $\{\mathbf{A}(k,\ell), \mathbf{b}\}$ of the model is to minimize the trace of $\mathbf{E}$. The inclusion of $\mathbf{b}$ in

*This work was supported by the Joint Services Electronics Program under Contract #DAAG29-81-K-0024.

the unknown parameters guarantees that the prediction error $\mathbf{e}(m,n)$ will be a 2-D zero-mean vector sequence. The normal equations are:

$$\sum_k \sum_\ell \Phi(i,j:k,\ell)\mathbf{A}^T(k,\ell)+\mathbf{s}(i,j)\mathbf{b}^T=\Phi(i,j:0,0) \quad (5a)$$

$$\sum_k \sum_\ell \mathbf{s}^T(k,\ell)\mathbf{A}^T(k,\ell) + N_s \cdot \mathbf{b}^T = \mathbf{s}^T(0,0) \quad (5b)$$

where we observe the matrix correlation and vector shift lags respectively:

$$\Phi(k,\ell:i,j) = \sum_m \sum_n \mathbf{x}(m-k,n-\ell)\mathbf{x}^T(m-i,n-j) \quad (6a)$$

$$\mathbf{s}(k,\ell) = \sum_m \sum_n \mathbf{x}(m-k,n-\ell) \quad (6b)$$

In (5), (k,ℓ) and (i,j) range over the set Π. In (6), (m,n) range over the analysis frame, and N_s in (5b) denotes the number of samples inside the analysis frame.

An alternative way of modeling $\mathbf{x}(m,n)$ would be to autoregressively model each channel separately:

$$x_i(m,n) = \sum_k \sum_\ell a_i(k,\ell)x_i(m-k,n-\ell)+b_i+e_i(m,n) \quad (7)$$

for $i=1,2,\ldots,N$, where the optimal scalar linear prediction coefficients $a_i(k,\ell)$ and bias coefficient b_i are obtained by minimizing the mean-squared value of the scalar prediction error signal $e_i(m,n)$ over the analysis frame, as explained in [1,2]. Obviously the scalar models in (7) are a subcase of the vector model in (1) with the prediction matrices $\mathbf{A}(k,\ell)$ being diagonal.

One approach to compute the correlation and shift lags in (6) is to assume the vector image signal to be zero outside the analysis frame, which is similar to the autocorrelation method of 1-D linear prediction. Alternatively, samples on the borders of the frame could be supplied as needed in the computation of (6); this latter approach is called the covariance method. The covariance method gives better estimates of the predictor parameters and of the bias, and a smaller mean-squared prediction error than the autocorrelation method. However, neither method can guarantee stability of either the resulting scalar or matrix autoregressive models.

The stability of the matrix filter $1/\mathbf{F}(z_1,z_2)$ is necessary for the stable reconstruction of $\mathbf{x}(m,n)$ from the prediction matrices, the bias, and the prediction error signal $\mathbf{e}(m,n)$. This stability is equivalent to the scalar 2-D polynomial $\det[\mathbf{F}(z_1,z_2)]$ being minimum phase, where "det[•]" denotes determinant of a matrix. With the covariance method, the estimation of the bias interacts with the stability in the following way: From (2) and (3) we infer that $\mathbf{b} = \mathbf{F}(1,1)\mathbf{d}$. Therefore, if the image signal has a nonzero dc-level ($\mathbf{d}\neq\mathbf{0}$) and we arbitrarily require $\mathbf{b}=\mathbf{0}$ in (5), then we force the determinant of $\mathbf{F}(1,1)$ to become zero, which forces the model to be marginally unstable since $\det[\mathbf{F}(1,1)]=0$ corresponds to a pole on the unit-surface. Moreover, as we proved in [2], if the prediction mask has a quarter-plane region of support, then a necessary condition for stability is

$$\det[\mathbf{F}(1,1)] > 0 \quad (8)$$

Finally, if we use the autocorrelation method with a 2-D separable prediction mask, then the stability of the inverse prediction error filter is guaranteed in both the scalar and the vector cases.

MULTICHANNEL ADPCM CODING

We used the above theoretical formulation of 2-D linear prediction for the design of the predictors in the feedback loops of an ADPCM image coding scheme of the feed-forward type. Initially, each channel of the multichannel image was coded separately using a single-input single-output ADPCM, as described in [1,2], at an average information rate of 1 bit/pixel or less. This resulted in a bit rate of about N bits/pixel for an N-channel color image. However, since our interest was in much lower bit rates and because we wanted to exploit correlation between channels, we used the multi-input single-output ADPCM scheme shown in Fig. 1.

The philosophy of each feedback loop in Fig. 1 is that for the i-th channel the $\mathbf{P}_i$ predictor forms an estimate from past samples of the reconstructed image signal $\hat{x}_i(m,n)$. This estimate is subtracted from the incoming image signal $x_i(m,n)$ to form the difference signal $d_i(m,n)$ which is quantized and encoded into the 2-D signal $c(m,n)$ for transmission. At the receiver, the quantized difference signal $\hat{d}_i(m,n)$ excites the i-th inverse prediction error filter to produce the reconstructed image signal $\hat{x}_i(m,n)$ for the i-th channel.

The design of the multi-input single-output quantizer Q in Fig. 1 is governed by the intuition that for most natural color images the edges occur at approximately the same location in every channel. The edge-information in the i-th channel is conveyed mainly by the prediction error signal $e_i(m,n)$. However, assuming small quantization errors, the difference signal $d_i(m,n)$ approximates $e_i(m,n)$. Therefore an encoded quantized difference signal would contain mainly information about the edge-location. This is depicted in Fig. 2 where the binary images (a), (b), (c) show the encoded quantized (2-levels/pixel) difference signals of the red, green, and blue channel separately for a head and shoulders image with well defined edges. The binary image of Fig. 2(d), however, shows the 2-levels/pixel common encoded quantized difference signal which is the output of the multi-input single-output quantizer of Fig. 1. By comparing the images of Fig. 2, we realize that by using a single-channel for information about edge-location we do not loose many edges. The encoded signal $c(m,n)$ was formed by first finding a single-channel difference signal:

$$d(m,n) = \sum_{i=1}^{N} w_i \cdot d_i(m,n) \qquad (9)$$

where the w_i's are weighting coefficients, and then quantizing and encoding $d(m,n)$ as follows:

$$c(m,n) = \begin{cases} 1 & , \quad d(m,n) \geq \Theta \\ 0 & , \quad -\Theta < d(m,n) < \Theta \\ -1 & , \quad d(m,n) \leq -\Theta \end{cases} \qquad (10)$$

The encoded signal $c(m,n)$ represents the sequence of codewords. The quantized difference signals are determined as follows:

$$\hat{d}_i(m,n) = c(m,n) \cdot \Delta_i \quad , i=1,2,\ldots,N \qquad (11)$$

The threshhold Θ in (10) and the step sizes Δ_i in (11) are adapted over each M×M analysis frame of the image according to the rule:

$$\Theta = K \cdot \sigma_e \quad , \Delta_i = D \cdot \sigma_{e_i} \qquad (12)$$

where σ_{e_i} is the rms value of the i-th prediction error signal $e_i(m,n)$ in the analysis frame, and σ_e is the rms value of a single-channel prediction error signal formed by a linear combination of all the $e_i(m,n)$ using the same weighting coefficients as in (9). The constants K and D are determined empirically [1,2]. The 3-level quantization logic of (10) allows us to set $\Theta=0$ and thus quantize the difference signal with 1-bit fixed length codewords. Alternatively, if $\Theta \neq 0$, by adjusting K we can produce at the output of the quantizer a large percentage of zero levels which will reduce significantly the entropy of the quantized difference signal and enable us to use Huffman codewords of variable length in order to achieve an average bit rate of much less than 1 bit/pixel.

In addition to the encoded quantized difference signal, we must transmit to the receiver "side-information" about the predictor parameters, the bias and the step size. The predictors $\mathbf{P}_i$ in Fig. 1 are designed either as scalar predictors (with prediction coefficients operating on the i-th channel) or as vector prediction (with predictor matrices operating on all the channels simultaneously). Unfortunately, the issue of stability and the limited available mathematical tools for 2-D polynomials limit our choices among various approaches. For scalar predictors the autocorrelation method with a 2-D separable prediction mask guarantees stability and it allows us to quantize the prediction coefficients in the domain of the log-area-ratios, exactly as done with LPC coding of speech. Alternatively, we can use the "stabilized" covariance method with a non-separable 2-D mask, as explained in [1,2], and use a logarithmic quantizer to quantize the coefficients inside a fixed range. For vector predictors, we can use the autocorrelation method with a 2-D separable mask for guaranteed stability. The quantization of the entries of the resulting prediction matrices is still under investigation. The components of the bias vector $\mathbf{d}$ and the step sizes Δ_i are quantized by using log-quantizers.

EXPERIMENTAL RESULTS

We successfully applied the multichannel adaptive prediction ADPCM coding to color aerial photographs and head and shoulders images. These color images had only 3 channels (red, green and blue) with a total resolution of 24 bits/pixel. The analysis frames consisted of 16×16 or 32×32 pixels. The prediction masks had a quarter-plane region of support with 2×2 or 3×3 samples in extent. By coding each channel separately at 1 bit/pixel or less, color reconstructed images of high quality resulted at a rate of ≈3 bits/pixel or less. By using a multi-input single-output ADPCM with adaptive scalar prediction and 3-level quantization color reconstructed images of good quality resulted at a total rate of ≈1 bit/pixel or less (down to ≈0.8 bit/pixel). These rates correspond to compression factors of about 24:1 or more. The mixing of the different channels in Eq. (9) was done by using as weighting coefficients 0.3, 0.6 and 0.1 for the red, green and blue channel respectively, since the green color is the most important and the blue is the least important for edge-content [4].

By using multichannel ADPCM with adaptive matrix (instead of scalar) predictors we obtained coded images whose quality was similar to the quality of the images coded by using scalar predictors. Since matrix linear prediction gives a smaller prediction error residual than scalar linear prediction, we are continuing to investigate ways of achieving higher image quality using matrix predictors.

REFERENCES

[1] P. A. Maragos, R. M. Mersereau, and R. W. Schafer," Some Experiments in ADPCM Coding of Images," Proc. ICASSP-82, Paris, France, May 1982, pp. 1227-1230.

[2] ________, "Two-dimensional Linear Prediction and its Application to Adaptive Predictive Coding of Images," submitted for publication to IEEE Trans. on ASSP, Aug. 1983.

[3] ________, "Two-Dimensional Linear Predictive Analysis of Arbitrarily-Shaped Regions," Proc. ICASSP-83, Boston, Mass., April 1983, pp. 104-107.

[4] J. O. Limb, C. B. Rubinstein, and J. E. Thompson, "Digital Coding of Color Video Signals-A Review," IEEE Trans. on Commun., Vol. COM-25, No. 11, Nov. 1977, pp. 1349-1385.

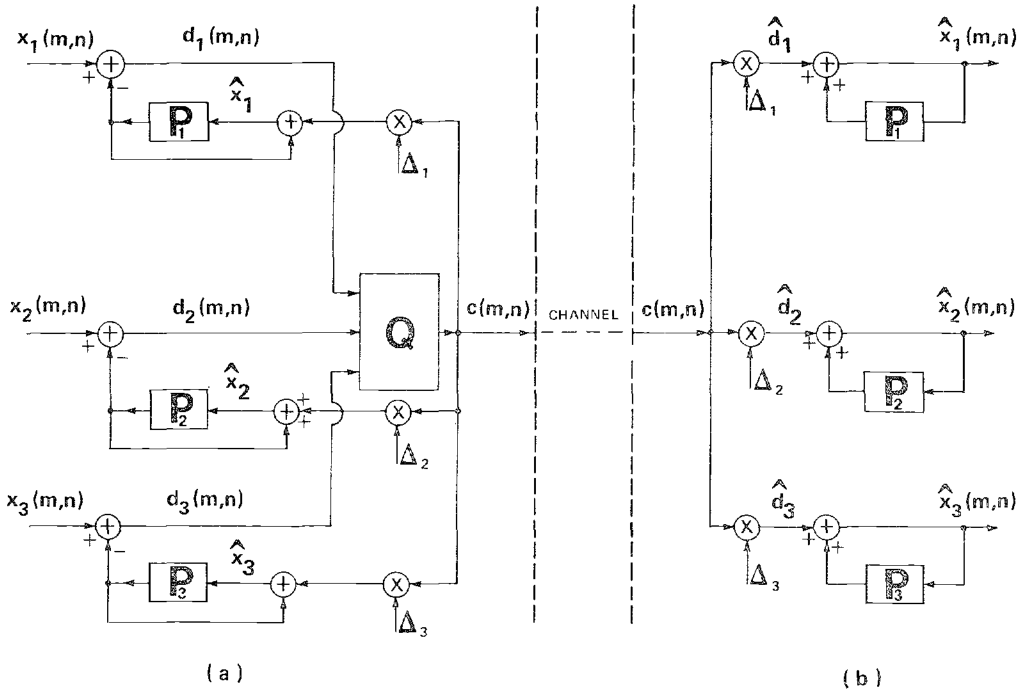

Figure 1 – Multi-input single-output ADPCM (a) Coder, (b) Decoder

Figure 2 – Binary encoded quantized difference signals, (a) Red channel, (b) Green, (c) Blue, (d) Combined