MULTICHANNEL LINEAR PREDICTIVE CODING OF COLOR IMAGES

GEORGIA INST OF TECH ATLANTA SCHOOL OF ELECTRICAL ENGINEERING  PA MARAGOS ET AL. 1984 AFOSR-TR-84-0355

UNCLASSIFIED DAAG29-81-K-0024

END

7-75

NL
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS TR 14
This paper reports on a preliminary study of applying single-channel (scalar) and multi-channel (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 the investigators obtained, from original color images of 24 bits/pixel, reconstructed images of good quality at information rates of 1 bit/pixel or less.
MULTICHANNEL LINEAR PREDICTIVE CODING OF COLOR IMAGES

Petros A. Naregos, Russell N. Horsdome, and Ronald W. Schafer

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 successes of 2-D linear prediction, we tried to extend its use to multispectral images either by autoregressive modeling each channel separately or by using a vector 2-D linear predictor which exploits the cross-correlation between channels. These two approaches resemble 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 \( x(m,n) = [x_1(m,n),\ldots,x_N(m,n)]^T \) represent an N-channel 2-D image vector signal, where \([\cdot]^T\) denotes the transpose of a vector and \(x_k(m,n)\) 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 \(x(m,n)\):

\[
x(m,n) = \sum_{k=1}^{L} \sum_{i=1}^{M} A(k,i) x(m-k,n-i) + b(m,n)
\]

where we predict the vector \(x(m,n)\) from its neighbor vector values weighted by "prediction matrices" \(A(k,i)\) of order \(N\times N\). In (1), \((k,i)\) range over all integer pairs in a set \(E\) 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 \(b = [b_1,\ldots,b_N]^T\) accounts for the fact that the intensity image samples are explicitly biased by a dc-level vector \(d = [d,\ldots,d]\) since they are always nonnegative. The 2-D vector prediction error signal \(e(m,n)\) is the output of a \(N\times N\) matrix prediction error filter

\[
P(z_1,z_2) = 1 - \sum_{k=1}^{L} \sum_{i=1}^{M} A(k,i) z_1^{-k} z_2^{-i}
\]

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

\[
b = [I - \sum_{k=1}^{L} A(k,i)] d
\]

Consider the \(N\times N\) average prediction error matrix

\[
E = \sum_{m,n} e(m,n) e^T(m,n)
\]

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

This work was supported by the Joint Services Electronics Program under Contract 60AG29-81-E-0024.
the unknown parameters guarantees that the prediction error \( s(m,n) \) will be a 2-D zero-mean vector sequence. The normal equations are:

\[
\sum \sum \theta(k,j)k^m \theta(k,j) = \theta(0,0)
\]

\[
\sum \sum k^m \theta(k,j) \theta(k,j) = \theta(0,0)
\]

where \( \theta(k,j) \) is the correlation between the vector model in the unknown channel. The vector model in the unknown channel, where \( k \) and \( j \) denote the number of samples inside the analysis frame.

An alternative way of modeling \( s(m,n) \) would be to autocorrelatively model each channel separately:

\[
x_1(m,n) = \sum \sum a(k,l)k^m \theta(k,j) + b_l + e_1(m,n)
\]

for \( i=1,2,\ldots,N \), where the optimal scalar linear prediction coefficients \( a(k,l) \) and bias coefficients \( b_l \) are obtained by minimizing the mean-squared value of the scalar prediction error signal \( e_1(m,n) \) over the analysis frame, as explained in [1,2]. Obviously the scalar models in (5) are a subcase of the vector model in (1) with the prediction matrices \( \theta(k,j) \) 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 \( 1/F(z) \) is necessary for the stable reconstruction of \( s(m,n) \) from the prediction matrices. The bias, and the prediction error signal \( e(m,n) \). This stability is equivalent to the scalar 2-D polynomial \( \text{det}[F(z)] \) being minimum phase, where \( \text{det}[.] \) denotes determinant of a matrix. With the covariance method, the estimation of the bias involves the stability in the following way: From (2) and (3) we infer that \( b = F(1) \). Therefore, if the image signal has a non-zero dc-level (d=0) and we arbitrarily require \( b=0 \) in (5), then we force the determinant of \( F(1) \) to become zero, which forces the model to be marginally unstable since \( \text{det}[F(1)] = 0 \) corresponds to a pole on the unit circle. Moreover, as we proved in [2], if the prediction mask has a quarter-plane region of support, then a necessary condition for stability is

\[
\text{det}[F(1)] > 0
\]

Finally, if we use the autocorrelation method with a 2-D separable prediction mask, then the stability of the linear 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 predictor 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 \( W \) bits/pixel for an \( N \)-channel 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 1-th channel the \( P_l \) predictor forms an estimate from past samples of the reconstructed image signal \( x_i(m,n) \) This estimate is subtracted from the incoming image signal \( x(i,m,n) \) to form the difference signal \( d_l(m,n) \) which is quantised and encoded into the 2-D signal \( c(m,n) \) for transmission. At the receiver, the quantised difference signal \( d_l(m,n) \) excites the 1-th inverse prediction error filter to produce the reconstructed image signal \( x_i(m,n) \) for the 1-th channel.

The design of the multi-input single-output quantiser \( Q \) 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 1-th channel is conveyed mainly by the prediction error signal \( e_1(m,n) \). However, assuming small quantisation errors, the difference signal \( d_1(m,n) \) approximates \( e_1(m,n) \). Therefore an encoded quantised 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 quantised (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 quantised 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 lose many edges. The encoded signal \( c(m,n) \) was formed by first finding a single-channel difference signal:

\[
d(m,n) = \sum_{i=1}^{W} w_i d_i(m,n)
\]

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

\[
c(m,n) = \begin{cases} 1 & d(m,n) > \theta \\ 0 & -\theta < d(m,n) < \theta \\ -1 & d(m,n) < -\theta
\end{cases}
\]

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

\[
d_i(m,n) = c(m,n) - A_i, \quad i=1,2,\ldots,N
\]

The thresholds \( \theta \) in (10) and the step sizes \( A_i \) in (11) are adapted over each WM analysis frame of the image according to the rule:

\[
\theta^* = K \sigma_i, \quad A_i = D \sigma_i
\]

where \( \sigma_i \) is the rms value of the \( i \)-th prediction error signal \( e_i(m,n) \) in the analysis frame, and \( \sigma \) 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 such less than 1 bit/pixel.

In addition to the encoded quantised difference signal, we must transmit to the receiver "side-information" about the predictor parameters, the bias and the step size. The predictors \( P_i \) in Fig. 1 are designed either as scalar predictors (with prediction coefficients operating on the 1-channel) or as vector predictors (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-error-ratio, exactly as done with LMC coding of speech. Alternatively, we can use the "stabilised" covariance method with a non-separable 2-D mask, as explained in [1,2], and use a logarithmic quantiser 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 \( \theta \) and the step sizes \( A_i \) 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 bit/pixel. The analysis frames consisted of 16x16 or 32x32 pixels. The prediction masks had a quarter-plane region of support with 2x2 or 3x3 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 8 bits/pixel or less. By using a multi-input single-output ADPCM with adaptive scalar prediction and 3-level quantizer color reconstructed images of good quality resulted at a total rate of 6 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 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

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

(a) Binary encoded quantized difference channel
(b) Green
(c) Blue
(d) Combined