M2.16

LOW BIT-RATE IMAGE CODING USING 2-D LINEAR PREDICTION AND 2-D STOCHASTIC EXCITATION

K.K. Paliwal
Computer Systems and Communications Group
Tata Institute of Fundamental Research
Homi Bhaba Road, Bombay 400 005, India

Abstract- The stochastic excited linear predictive coding method has been recently proposed for low bit-rate coding of speech. In the present paper, we investigate the use of this coding method for image coding and study its different parameters. We show that it is not necessary to transmit local bias values of the image frames. We also show that the stochastic excitation is not adequate to represent the prediction residual signal. In order to get good performance from this coder, it is necessary to generate the codebook from the actual prediction residual signal.

1. Introduction

Recent trend in the image coding area seems to be to borrow concepts from the existing speech coding techniques and do their two-dimensional (2-D) extension to develop new image coders. Some examples of the coders developed in this fashion are the ADPCM coder [1], the sub-band coder [2], the vector quantizer [3,4], the multipulse excited (MPE) coder [5] and the regular pulse excited (RPE) coder [6].

In the speech coding area, the linear prediction (LP) analysis technique has been very popular over the last several years [7,8]. In the LP-based speech coders, the short-time speech spectrum is represented by a few LP coefficients and the residual signal obtained by passing the speech signal through the LP inverse filter is modelled and quantized. In the earlier LP-based speech coders [9], the modelling and quantization are done in such a manner that the difference between the unquantized and quantized residual signal is minimum. However, the present-day LP-based speech coders [10-13] model and quantize the residual signal in such a manner that the perceptually-weighted difference between the original and the reconstructed speech is minimized. This allows graceful degradation in the performance of the present-day coders with decrease in the bit-rate; as opposed to drastic degradation in the earlier coders. Examples of the present-day speech coders are the MPE coder [10], the RPE coder [11] and the stochastic excited (STE) coder.

In the multipulse excited (MPE) speech coder [10], the residual signal is modelled as a sequence of pulses. The model parameters (pulse positions and amplitudes) are determined from the speech signal by an analysis-by-synthesis procedure which estimates one pulse at a time and is sub-optimum in nature. These model parameters (pulse positions and amplitudes) are scaler-quantized. Though the model-parameter estimation procedure is sub-optimum, it tries to minimize the perceptual difference between the original and the reconstructed speech. Because of this, the MPE coder gives good quality speech at medium bit-rates (9.6-16 kbits/sec). But, it can not be used for low bit-rate (4.8 kbits/sec) coding of speech. The MPE coding has been successfully extended to image coding by Horn et al. [5], where it can produce good quality images at a bit-rate of 1 bit/pixel. However, it has not been possible to apply the MPE coding for images for lower bit-rates because the quality deteriorates drastically.

Recently, the stochastic excited coding of speech has been proposed for low bit-rate coding at 4.8 kbits/sec [13]. Here the prediction residual signal is modelled by a sequence of random Gaussian numbers. Both modelling and quantization of the residual signal is done in such a way that the perceptually meaningful difference between the original and the reconstructed speech is minimum. The concept of codebook coding (or, vector quantization) is used here for efficient quantization. In the present paper, our aim is to explore the STE coding method for image coding. Though the STE coding is computationally very expensive, recent reduction in cost and increase in speed of digital signal processing hardware have encouraged us to investigate the STE coding method for coding images at low bit-rates (<0.5 bit/pixel).

2. The STE image coder

A. Model

In the STE image coder, the image is synthesized at the receiver using the model shown in Fig. 1. Here, $x(m,n)$ denotes the 2-D sequence representing the image intensity samples. Since all the image intensity samples are positive, a bias term B is introduced in the model. For synthesizing the image signal, an optimum codevector is selected from the codebook and multiplied by a suitable gain factor G to derive the excitation signal $u(m,n)$ to the LP synthesis filter $H(z)=1/A(z,w)$ where

$$A(z,w)=1-\sum_{k}\sum_{l}_{(k,l)\in R} a(k,l)z^{-k}w^{-l}.$$

Here, $\{a(k,l)\}$ are the LP coefficients and R is the region of support of the predictor. The difference equation describing the synthesis of the image signal is given by

$$\hat{x}(m,n)=\hat{y}(m,n)+B$$
$$=\sum_{k}\sum_{l}_{(k,l)\in R} a(k,l)\hat{y}(m-k,n-l)+B+u(m,n).$$

In this model of image synthesis, the required parameters are bias coefficient B, LP coefficients $\{a(k,l)\}$, address of the optimum codevector and the gain factor G. These parameters have to be computed from the image signal $\{x(m,n)\}$, quantized and transmitted to the receiver. The procedures for computing these parameters and their subsequent quantization are described below.

B. Estimation of bias and LP coefficients

For computing the bias and the LP coefficients, the total image is divided into non-overlapping analysis frames of size (MxN). The bias coefficient B is computed for each of these analysis frames as a local mean of intensity samples and quantized uniformly using 7 bits.

For computing the LP coefficients for each of the analysis frames, the region R of support of predictor is assumed here to be causal, though the spatial causality is not an inherent property of image formation. The region R is taken here to be a (QxQ) quarter-plane such that $R=\{(k,l);\ 0\leqslant k,l\leqslant Q-1 \text{ and } (k,l)\neq(0,0)\}$. The predictor order P is related to Q by $P=Q^2-1$. In the present study, we have considered only the third order linear predictor, because higher-order linear predictor does not increase the prediction gain appreciably [1].

The three LP coefficients are determined from the image data in the analysis frame by minimizing the total-squared value E of the 2-D prediction residual signal

$$E=\sum_{m=2}^{M}\sum_{n=2}^{N}\{e(m,n)\}^2,$$

where

$$e(m,n)=y(m,n)-a(1,0)y(m-1,n)-a(1,1)y(m-1,n-1)-a(0,1)y(m,n-1),$$

where

$$y(m,n)=x(m,n)-B.$$

This minimization leads to a set of linear equations which can be solved to determine the LP coefficients. This method of estimating LP coefficients is known as the covariance method [1]. There is another method of LP analysis, namely the autocorrelation method, which differs from the covariance method in terms of limits of summation in the above equation. Other details about this method can be found in [1].

For quantization of the LP coefficients, each LP coefficient is transformed into log area ratio (as done in speech coding [8]) to get uniform spectral sensitivity. Each log area ratio coefficient is then uniformly quantized using 6 bits.

C. Estimation of the optimum codevector

For computing the address and gain of the optimum stochastic codevector, the LP analysis frame of image is divided into non-overlapping search sub-frames of size (IxJ) such that M/I and N/J are positive integers. As mentioned earlier, in the STE coder, the residual image signal is represented by a sequence of Gaussian random numbers. So a codebook of L codevectors is constructed here from the Gaussian random numbers and stored both at the transmitter and the receiver.

In order to find the optimum codevector which represents the residual image signal, an exhaustive search procedure shown in Fig. 2 is used. Here, all the codevectors are processed sequentially one at a time. Each component of a codevector is scaled by a gain factor G which remains constant for the (IxJ) long stochastic search sub-frame and is reset to a new value for the next sub-frame. The components of scaled codevector are filtered through an LP synthesis filter and the bias coefficient is added to each of them. The synthesized image samples are compared with the corresponding original image samples to form a difference signal. The difference signal representing the objective error is further processed by another filter which makes this error perceptually more meaningful for a human observer. (The transfer function of this perceptual weighting filter can be derived from the existing knowledge of the human visual perception system. However, in the present study, this transfer function is taken to be unity.) The perceptually-weighted total-squared error is found over the sub-frame for each of the L stochastic codevectors in the codebook and the optimum codevector is selected as the one which results in least weighted-error. The gain factor G of the optimum codevector is also computed by minimizing the perceptually-weighted total-squared error. This is given by

$$G=\sum_{i=1}^{I}\sum_{j=1}^{J} y(i,j)\hat{y}(i,j) \Big/ \sum_{i=1}^{I}\sum_{j=1}^{J}\{\hat{y}(i,j)\}^2,$$

where $\{y(i,j)\}$ and $\{\hat{y}(i,j)\}$ are the bias-subtracted original and synthesized images. The logarithmic value of gain factor G is uniformly quantized using 6 bits. In order to facilitate the transmission of address of the codevector, the codebook size is taken to be a power of 2. For L=1024, the total bit-rate of the STE image coder is 0.274 bit/pixel.

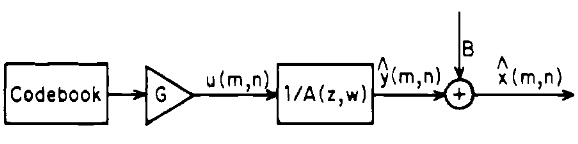

Fig. 1. Receiving end of the STE image coder.

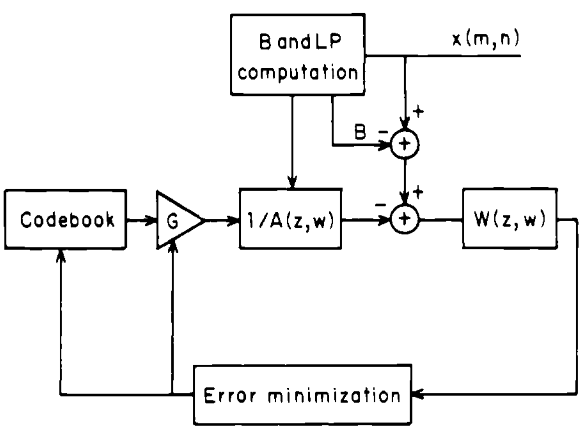

Fig. 2. Transmitting end of the STE image coder.

3. Results

The STE coder is studied here on a number of 512 x 512 images. Different parameters used in this coder are listed in Table 1. The performance of the STE coder is evaluated in terms of signal-to-noise ratio [14]. For the values of parameters listed in Table 1, the STE coder results in an SNR of 19.5 dB which is not very encouraging.

We have tried to study the effect of various parameters on the SNR performance of the STE coder. In order to see whether the introduction of bias term is necessary in the STE model shown in Fig. 1, we study the performance of the STE coder with and without bias term. The SNR results are found to be 19.51 dB with bias term and 20.36 dB without bias term. Thus, the use of bias term in the STE coder deteriorates its image quality. In addition, it requires 6 bits per frame for its transmission. So, it will be better not to transmit the frame-bias values.

Next, we study the effect of different methods (covariance and autocorrelation methods) of LP analysis on the performance of the STE coder. The STE coder results in an SNR of 20.36 dB for the covariance method and 20.24 dB for the autocorrelation method. In addition, it might be noted that both the covariance and autocorrelation methods for 2-D LP analysis do not guarantee the stability of the LP synthesis filter. So from performance point-of-view, the covariance method is preferable for 2-D LP analysis.

We have also studied the effect of larger codebook sizes on the performance

Table 1. Parameters of STE image coder

Image dimension	512x512
Image gray levels	256
LP analysis frame (MxN)	32x32
LP Order (P)	3
LP analysis method	Covariance
Search sub-frame (IxJ)	8x8
Weighting filter (W)	Unity
Codebook size (L)	1024
Total bit-rate	0.274 bit/pixel

of the STE coder. The SNR results are 20.36 dB for L=1024, 20.41 dB for L=2048 and 21.29 dB for L=4096. Though the SNR performance improves with the codebook size, it is still not very satisfactory even for L=4096. From this, we conclude that the stochastic model is not good to represent the prediction residual signal.

In order to improve the performance of the coder, we have tried to use codebook generated by using the Linde-Buzo-Gray algorithm [15] on the training data consisting of the actual residual signal. For L=256, I=4 and J=4, it resulted in SNR improvement of 4.8 dB.

4. Conclusion

In the present paper, the STE coding method is studied for image coding at low bit rates (0.274 bit/pixel). It is shown that it is not necessary to include bias term in the STE model. It is also shown that the stochastic model is not adequate for representing the prediction residual signal. For getting better performance, it is necessary to generate a codebook from the training data consisting of the residual signal.

References

[1] P.A. Maragos et al., "Two-dimensional linear prediction and its application to adaptive predictive coding of images", IEEE Trans. ASSP-32, pp. 1213-1239, Dec. 1984.

[2] J.W. Woods and S.D. O'Neil, "Subband coding of images", IEEE Trans. ASSP-34, pp. 1278-1288, Oct. 1986.

[3] A. Gersho and B. Ramamurthi, "Image coding using vector quantization", Proc. ICASSP, pp. 428-431, 1982.

[4] T. Murakami et al., "Vector quantization of video signals", Electron. Lett., pp. 1005-1006, Nov. 1982.

[5] C. Horn et al., "Image coding based on 2-D linear prediction and 2-D multipulse excitation", Proc. EUSIPCO, pp. 673-676, 1986.

[6] C. Horn and E.F. Deprettere, "Multi-pulse and regular-pulse LP coding of images", Proc. ICASSP, pp. 1083-1086, 1987.

[7] J. Makhoul, "Linear prediction: A tutorial review", Proc. IEEE, pp. 561-580, 1975.

[8] J.D. Markel and A.H. Gray, Linear Prediction of Speech, Springer-Verlag, Berlin, 1976.

[9] C.K. Un and D.T. Magil, "The residual-excited linear prediction vocoder with transmission rate below 9.6 kbit/s", IEEE Trans. Com-23, pp. 1466-1474, Dec. 1975.

[10] B.S. Atal and J.R. Remde, "A new model of LPC excitation for producing natural sounding speech at low bit rates", Proc. ICASSP, pp. 614-617, 1982.

[11] P. Kroon et al., "Regular-pulse excitation - A novel approach to effective and efficient multipulse coding of speech", IEEE Trans. ASSP-34, pp. 1054-1063, 1986.

[12] M.R. Schroeder and B.S. Atal, "Speech coding using efficient block codes", Proc. ICASSP, pp. 1668-1671, 1982.

[13] M.R. Schroeder and B.S. Atal, "Code-excited linear prediction (CELP): High-quality at low bit rates", Proc. ICASSP, pp. 937-940, 1985.

[14] A.K. Jain, "Image data compression: A review", Proc. IEEE, pp. 349-389, Mar. 1981.

[15] Y. Linde et al., "An algorithm for vector quantizer design", IEEE Trans. Com-28, pp. 84-95, Jan. 1982.