PREMIER COLLOQUE IMAGE
Traitement, Synthèse, Technologie et Applications

BIARRITZ – Mai 1984 –

DESIGN OF CODEC FOR VIDEO CONFERENCING

CONCEPTION D'UN CODEC POUR VIDEO CONFERENCE

O. Ibikunle, N.M. Nasrabadi* and R.A. King.

Department of Electrical Engineering, Imperial College of Science and Technology, London SW7 2BT.

* 'Systems Designers', Systems House, 105 Fleet Road, Fleet, Hants, U.K.

RESUME

SUMMARY

Abstract

A hybrid image coding design, suitable for applications in limited motion imagery as in video-conferencing, is presented. The scheme intrinsically involves a transform domain approach to vector quantization. The fundamental design for non-temporal pictures is discussed, a corresponding effective adaptive scheme dependent on computed row intercorrelation spans within a transformed picture is proposed, and the inter-frame version of the codec is described. Simulation results have shown that good quality coding is obtained for a range of of data rates of 1.0-2.0 bits/pel for the intra-frame codec and 1.0 bits/pel for the interframe codec. The potential for further improvement is indicated with long vector-template quantization.

1 Introduction

Diverse image processing applications such as digital television, teleconferencing, remote sensing and biomedical monitoring and detecting procedures require suitable codecs for efficient data compression of pictorial information. Typical pictures (still or motion visual systems) utilised in such applications generally require large raw data rates in order to be faithfully retrieved after storage or transmission at their original fidelity (standard quality) and intelligibility (discernable detail). Excessive demands on memory storage and channel bandwidth are thus implicit and a codec must therefore serve to reduce the amount of data efficiently , to minimise the storage and transmission requirements. The efficiency of any image coding scheme is measured in terms of the resulting image degradation, its data compression potential and its implementation complexity.

Numerous codecs have been proposed for the image compression problem, however, the various techniques have mostly been classified into two essentially distinct categories [1] of prediction and transform coding respectively. A notable technique of predictive coding , which exploits the statistical correlation within the image data to permit differential signalling, is due to Zschunke [2]. His scheme involves a differential pulse code modulation (DPCM) technique with adaptive prediction dependent on contour directions within an image. The design reports good quality encoding of images with high definition contours, although low contrast compression did not differ considerably with fixed prediction coding. Transform coding involves implementing energy preserving transformations on image data to concentrate the maximum energy within a band of a minimum number of coefficients and to decorrelate the transformed data sufficiently. The resulting compaction and decorrelation permits the discarding of a number of low energy coefficients to achieve efficient coding. Chen and Smith [3] present an adaptive transform coder which allocates bits dependent on computed image activity indices and attractive fidelity coding is reported for monochrome and colour imagery.

Comparative studies [4] have generally inferred that predictive techniques are more sensitive to changes in image statistics and channel errors relative to transform coding. However, predictive coders have found wide use in real time coding of image processing applications because of their relative ease of design and speed of operation. Transform coders generally suffer relative equipment and implementation complexity particularly with respect to memory and computation requirements and real-time processing ability. Predictive techniques tend to have a better performance than transform codecs at higher bit rates and transform coders achieve superior compression at lower data rates [4,5].

A number of hybrid coding techniques have been proposed which combine the attractive features of transform and predictive methods and invariably offer good compression ability without many of the limitations of the distinct techniques. Habibi [5] has proposed two systems which basically involve application of orthogonal transformations in one spatial direction of the picture and effecting DPCM in the other direction. He reports good subjective results at low bit rates for moderately detailed scenery. Recent work by Nasrabadi and King [6] proposes fundamental designs of a hybrid codec that employs a vector quantization scheme in the transform domain in which row unitary transforms are implemented and vertical correlation is exploited. This paper presents an extension of the earlier work with a view to characterising this vector codec comprehensively and proposing a design for an adaptive version of the basic coding scheme of [6]. The extension to interframe coding for low-resolution video-conferencing is also discussed. It is noted that Gersho and Ramamurthi [7] have also applied a similar technique in the spatial domain and good results from encoding images of varied detail are reported, however, problems of local degradation were inherent at edge locations.

2. Fundamental design of vector codec (I). The operational block diagram of the vector codec is as shown in Fig.1. A bank of orthogonal (one dimensional ordered Hadamard) transformations are performed along each row of an N1 x N2 picture. This transformation is described thus

$$T(n_1, m_2) = \sum_{n_2 = 0}^{N_2 - 1} x(n_1,n_2)\ A_H(n_2,m_2) \qquad (1)$$

for $m_2 = 0,1,2,3,\ldots,N_2 - 1$; $A_H(n_2,m_2)$ is the one dimensional Hadamard transform (HT) kernel and $x(n_1,n_2,)$ and $T(n_1,m_2)$ are the appropriately indexed spatial and transformed pictorial data respectively. A number of high frequency (sequency for HT) components are discarded, after normalization by computed standard deviations (with respect to the columns). The standard deviations are given by

$$\sigma(m_2) = \left\{\frac{1}{N_1} \sum_{n_1=0}^{N_1-1} [T(n_1, m_2)]^2 - [\mu(m_2)]^2\right\}^{\frac{1}{2}} \qquad (2)$$

for $m_2 = 0,1,2,\ldots\ldots,N_2 - 1$ where $\mu(m_2)$ is the expected mean per column data. Discarding of normalised coefficients is on a zonal sampling basis and is also dependent on the desired bit rate and code-book dimensions. With high correlation exhibited amongst the transformed rows after coefficient normalization, a vector-quantizer with a vector mask of size (k x 1), partitions the picture into vertical (k x 1) vectors. Each picture vector therefore spans k-column-wise transform coefficients all localized in k neighbouring rows. The quantizer, exploiting the vertical correlation, maps these picture templates into standard templates by addressing a previously constructed code-book of such vectors and matching each transformed picture template to a corresponding nearest distance (in the euclidean sense) vector within the code-book. Each selected codebook vector is accordingly identified with a binary word for onward transmission over the channel. An identical codebook at the receiver containing all representative templates tabbed by corresponding binary words, is addressed by the incoming word for selection of the appropriate vector template for decoding. After complete decoding, denormalization of the coefficients is effected and the inverse HTs as given by (3) are implemented to retrieve the picture.

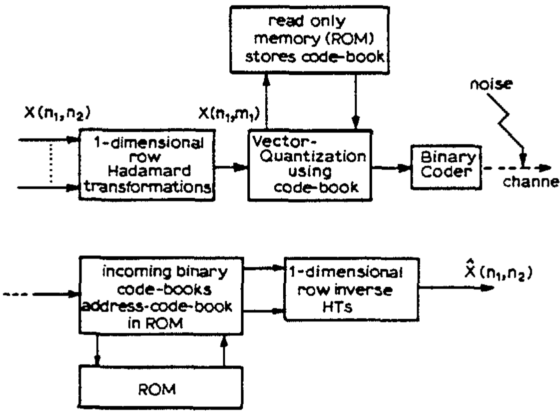

Fig.1 Codec I - Fundamental design of vector-codec

$$x(n_1,n_2) = \sum_{m_2=0}^{N_2-1} T(n_1,m_2)\ B_H(n_2,m_2) \qquad (3)$$

where $B_H(n_2,m_2)$ is the one dimensional Hadamard inverse kernel.

The resulting bit rate, R, for a code book of size (M,k) and a discarding ratio $c' = (c/N_2)$ is given as:

$$R = \frac{\mathrm{Log}_2\ M}{kc'} \qquad (4)$$

A Read Only Memory (ROM) may be practically employed for storage of the codebooks templates. The standard templates may be locally constructed from the image and be optimized over several images by applying an algorithm due to Linde et al [8] for codebook standardization.

3. Adaptive Vector-Quantization and coding scheme design (II)

The adaptivity (Fig. 2) introduced into the design may be applied for variable quantization at a fixed bit rate or variable rate. It is accomplished by varying the vector-mapping window (k_v x 1) in accordance with the statistical variation in the vertical direction of the row-transformed picture. The statistical variation is described by the euclidean distance between transformed rows. Thus the adaptive mapping is such that if there is sufficient intercorrelation between $k_v = b$ neighbouring rows then the rows are encoded with a codebook of vector lengths k = b. Previously developed codebooks for different template lengths, all jointly satisfying fixed or varied data rates (see (4)) depending on the desired adaptivity, are switched to and addressed in accordance with pre-computed thresholds of row inter-separation (euclidean sense).

More formally considered, if Dij is an array of euclidean distances between every pair of transformed rows i and j, then for positional separations between rows or row adjacencies, rs = j - i , an array of thresholds of euclidean distance may be computed for each row adjacency (the maximum rs being determined by the maximum desirable vector length k_m of the design). The threshold array used in this codec may be written thus

$$T(rs) = \frac{1}{L(rs)} \sum_{m=1}^{L(rs)} D_{ij}(rs) \qquad (5)$$

for $i = 1,2,\ldots\ldots,N-1$, $j = 2,3,\ldots\ldots\ldots,N$; $rs = 1,2,\ldots\ldots\ldots, k_m-1$ where L(rs) is the array of the number of occurrences of each rs in the picture of length N. The thresholds may be computed over a large set of pictures to broadly optimize the values.

The scanning and switching algorithm is designed to reflect elaborate protocol to ensure that the resulting switching operation is sufficiently fail-safe such that the inter-correlation within a span of the k_m rows would be fully explored before justifying the selection of a particular codebook. Sequential tracking and non-overlapping matching are other important factors involved in the protocol. The scanning algorithm is implemented by proceeding through all i,j combinations such that the following protocol is satisfied:

$$D_{ij}(rs) \leq T(rs) \qquad (rs = 1) \qquad (6a)$$
$$D_{ij}(rs) \leq T(rs) \qquad (rs \neq 1) \qquad (6b)$$
$$D_{ij'}(rs) \leq T(rs) \qquad (rs \neq 1) \qquad (6c)$$

for all $i = 1,2,\ldots\ldots\ldots, k_m - 1$, $j = 2,3,\ldots k_m$, and $j' = j + 1$ within each set of k_m

rows. Condition (6a) is unconditionally assumed as it is required, to start off the scanning scheme, since the minimum k_v for vector coding is $k_v = 2$ (i.e. rs = 1). Conditions (6b) and (6c) are conditionally satisfied, however, when (6b) holds (6c) must also hold, this ensures sequential scanning. The last j for which (6b) and (6c) are simultaneously valid, within a scan of k_m rows, represents the size of the vector windows to be encoded and thus a codebook of size (M, k = j) is switched to accordingly. Non-overlapping coding is ensured by noting the last j and resuming the intercorrelation examination scans for the next set of k_m rows from the j + 1th row.

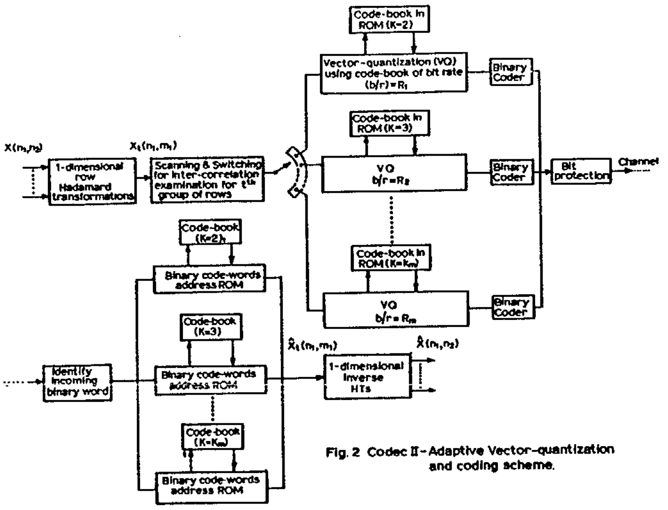

Fig. 2 Codec II - Adaptive Vector-quantization and coding scheme.

4. The interframe version (III)

The extension to interframe picture coding, of our design, essentially involves applying the vector-quantization procedure in the temporal direction where correlation between frames is exploited. However, in this case two dimensional transforms are implemented on sub-blocks of each frame and the vector-quantizer maps sufficiently intercorrelated pixels within corresponding sub-blocks of a frame sequence. Rather than apply the adaptive scheme discussed in section 3, a procedure which takes into consideration the temporal motion between frames, is proposed [9]. Two codebooks of vector lengths k = L and k = L/2 are used where computed image activity indexes, which are functions of transformed a.c. coefficients, identify corresponding sub-blocks within frame sequences as being areas of stationary-motion or rapid motion. Frame sub-block sequences of rapid movement are coded with the codebook of k = L/2, indicative of smaller correlation spans amongst temporal pixels and stationary-motion sub-block sequences are encoded with the codebook containing longer vectors to reflect larger correlation spans.

5. Simulation Results

The codecs discussed in the foregoing sections were simulated on a digital computer. Codecs (I) and (II) were used to encode the picture shown in Fig. 3a which is composed of 128 x 128 pixels at a resolution of 8 bits/pel. Figs. 3b and 3c are some results of fixed-length coded pictures at 2 bits/pel and 1.5 bits/pel respectively with normalized mean square errors (NMSE - an objective criterion of fidelity) with respect to the original picture of 0.5×10^{-2} and 2.294×10^{-2} respectively. Detailed simulations [10] have generally indicated that improved coding performance is accomplished for a desired bit rate, with longer vector windows of quantization (i.e. large k for a given R). Also for large k, very low-bit rates are obtainable, at appreciable fidelity, for a reasonable codebook size.

Figs. 4a and 4b are some results obtained from the adaptive codec (II) simulation. Fig. 4a is obtained from adaptive quantization at a fixed bit rate of 2.0 bits/pel by switching between four codebooks (k = 2,3,4,5). The effect of codec II is more pronounced when used as a variable rate codec as seen from Fig. 4b where an average bit rate of 1.1 bit/pel is obtained by switching between codebook rates of R = 1.0 (k = 6) bits/pel and 1.5 bits/pel (k = 5) respectively. This impresses the potential of the adaptive variable rate codec being very useful for good compression at desirably low data rates. This is further conceivable as it is possible to exploit the designed adaptive scheme by employing longer templates within the codebook pool at lower bit rates and sequentially smaller templates for higher bit rates. Thus, keeping the highest bit rate of the variable rate codec within a sufficiently low level, attractive compression is guaranteed with assured minimal overall codebook size. The result of fig. 4c is due to noise simulation, to view the performance of the basic codec in the presence of severe noise levels. Employing a high error rate of 10^{-2}, it is seen that appreciable detail is retained in the degraded picture. Hence, for more typical error levels, better performance of the codec is guaranteed. This also justifies, to a high degree, the transform domain approach to vector coding since localised errors incurred in the transform domain would be more uniformly distributed on inverse transformation to the spatial domain.

Fig. 3a. Original picture (128 x 128 pel) quantized to 8 bits/pel.

Fig. 3b. Fixed Vector-length coding at 2.0 bits/pel (k = 5), NMSE = 0.5×10^{-2}

Fig. 3c Fixed length coding at 1.5 bits/pel (k = 6), NMSE = 2.3×10^{-2}

Fig. 4a Adaptive quantization (k = 2,3,4,5) at fixed rate, R = 2.0 bits/pel, NMSE = 0.6×10^{-2}

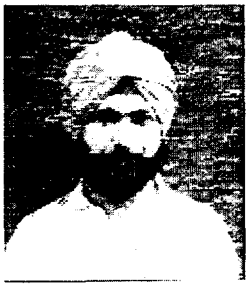

Fig. 4b Adaptive coding at 1.1 bits/pel, NMSE = 2.23×10^{-2}

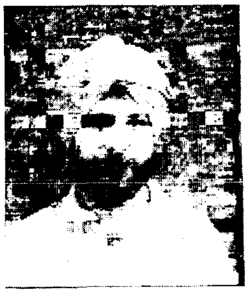

Fig. 4c Application of high level noise (error rate = 10^{-2}) at 1.5 bits/pel (k = 6)

Fig. 5a shows the original sequence (136×200 frame size) encoded by the interframe version of our codec. Fig. 5b shows a result of encoding at 0.8 bits/pel/frame. The performance, at such bit rates, shows that the interframe codec can achieve considerable savings in transmission rates in video conferencing, particularly where cameras are stationary and scenes consist mainly of small areas moving in front of relatively large stationary backgrounds.

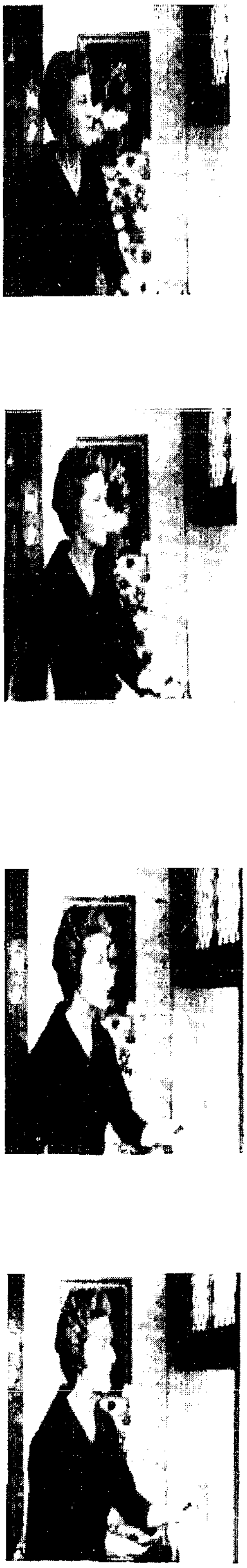

Fig.5a Original sequence of limited motion imagery (136 x 200 pel/frame)

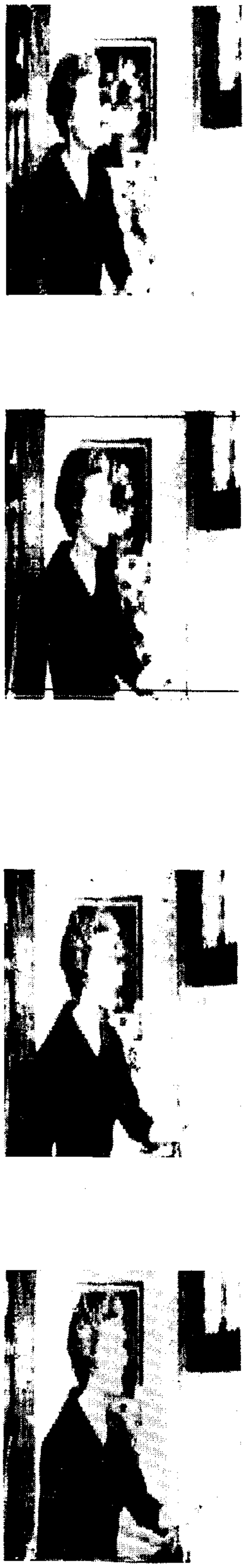

Fig. 5b Encoded sequence at 0.8 bits/pel/frame.

Problems of memory may seem imminent in our codec, however, for desirably low data rates and long-vector template coding it can be substantially minimized. Both adaptive schemes (II and III), would in practice require overhead information to be transmitted or stored identifying which codebook was switched to and addressed at the coder. Real-time processing may be further enabled by incorporating efficient parallel processing into any hardware implementation. Such processing is realizable for codec II, by computing the D_{ijs} over sections of the transformed rows rather than entire lengths. An increase in resolution of original pictures would further ensure high quality coding.

5. Conclusion

In this paper we have presented a transform domain vector codec design. We believe our design reflects a relatively simple approach to image data coding problems. Simulations have indicated attractive compression potential. The interframe version of this hybrid-codec is applicable in low resolution limited motion tele-conferencing.

References

1. A.K. Jain. "Image Data Compression; A review", Proc. IEEE, Vol. 69, No. 3, 1981, pp. 349 - 389

2. W. Zschunke. "DPCM Picture Coding with Adaptive Prediction", IEEE Trans. on Comm., Com-25, No. 11, 1977, pp. 1295 - 1302

3. W.H. Chen and C. H. Smith. "Adaptive Coding of Monochrome and Colour Images ", IEEE Trans. on Comm. Tech., Vol. Com-25, 1977, pp. 1285 - 1292

4. A. Habibi "Comparison of the nth-order DPCM encoder with Linear Transformation and Block-Quantization Techniques", IEEE Trans. Comm. Tech., Vol. Com-19, 1971, pp. 948 - 956

5. A. Habibi. "Hybrid Coding of Pictorial Data", IEEE Trans on Comm., Com-22, No. 5, 1974, pp. 614 - 624

6. N.M. Nasrabadi and R.A. King. "Image Coding using Vector-Quantization in the transform Domain", Pattern Recognition Leters I (1983), pp. 323 - 329

7. A. Gersho and B. Ramamurthi. "Image Coding using Vector-quantization",Proc. ICASSP, 1982, pp. 428 - 431

8. Y. Linde, A. Buzo and R.M. Gray. "An Algorithm for Vector Quantizer Design", IEEE Trans. Comm., Com-28, 1980, pp. 84 - 95

9. N.M. Nasrabadi and R.A. King. "Transform Coding using Vector Quantization", 1983 Conference of Information Sciences and Systems, Baltimore

10. J.O. Ibikunle. "A Study of a Vector Quantization Technique for Image Transform Coding", M.Sc. Thesis, Imperial College of Science and Technology 1983.