Conference on

Machine Processing of

Remotely Sensed Data

October 16 - 18, 1973

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

INFORMATION PRESERVING CODING FOR MULTISPECTRAL DATA [1]

by

J. R. Duan and P. A. Wintz

Laboratory for Applications of Remote Sensing
and
Department of Electrical Engineering
Purdue University, West Lafayette, Indiana 47907

I. ABSTRACT

A general formulation of the data compression system is presented. A method of instantaneous expansion of quantization levels by reserving two codewords in the codebook to perform a folding over in quantization is implemented for error free coding of data with incomplete knowledge of the probability density function. Results for simple DPCM with folding and an adaptive transform coding technique followed by a DPCM technique are compared using ERTS-1 data.

II. INTRODUCTION

In view of the massive amount of multispectral scanner data (LARS, 1973) that will be accumulated with the use of aircraft and satellites such as ERTS-1, it is desirable if the total number of bits recorded can be reduced and the original version of measurements can be reconstructed without error from the coded version. The ground station storage-space problem can thus be reduced and quantity of storage tapes for distribution of these data can also be reduced.

Much effort in the past, however, has been directed toward non-information preserving techniques (at least from the numerical error point of view) based on the model for the human visual system (Budrikis, 1972). Extensive bibliographies on topics related to data compression have been compiled [11]-[14]. In this study a class of information-preserving techniques for minimizing the amount of information (Shannon, 1949) that has to be retained, coded and stored to represent the multispectral data is discussed. As the application for scientific measurements such as the multispectral data is not intended for display purposes alone but rather to extract and derive information, it is anticipated that new information can be obtained when new processing techniques are developed. Therefore, it is difficult to predict the value of the coded data if error is introduced. The information-preserving techniques become very attractive. Deviations from the error-free reconstruction are also discussed. Absolute error bounds are used in the error control section to ensure a uniform quality throughout the data. A probabilistic error criterion is also proposed. A general formulation of the bit reduction problem will be presented first which shows the necessary steps involved and points out the various aspects of the data compression system.

[1] The work reported in this paper was sponsored by the National Aeronautics and Space Administration (NASA) under Grant Number NGR 15-005-152 and NGL 15-005-112.

III. MATHEMATICAL FORMULATION OF THE PROBLEM

It is convenient to think of the bit reduction basically as a stochastic minimization problem. Consider a set of discrete data D which consists of countable subsets I_i, i=1,2,.... For any i, I_i is a vector subset, i.e., $I_i=\{X_1,X_2,\ldots,X_n\}$ and the X_i, i=1,2,...,n, are m x 1 random vectors collected from a data source. Let b denote the total bits required for any given piece of information I, where $I \subseteq D$. Now b is a function of the transformation, the associated quantization schemes and any controlled redundancy introduced. Through an appropriate choice of these elements, the minimum can be found. We consider the general case where the number of bits for different blocks of data may be different; therefore, the minimization is over the expected value for a given subset of data $I \subseteq D$.

$$b_{min}(Q_o,X,T)=\underset{\substack{T_\ell \varepsilon F \\ Q_o}}{\text{Min}} \; \underset{X\varepsilon I}{E} \; \{S\left(Q_o[T_\ell(X)]_k\right)+C(T_\ell,k)\}$$

subject to the constraint that

$$\{e[\hat{X}(T_\ell,Q_o,Q_1,k),X]\}<\varepsilon$$

where

$$\hat{X}=Q_1[T_\ell^{\,t}\{Q_o[T_\ell(X)]_k\}]$$

$\hat{X}$: is the reconstructed data vector, m x 1.

F : is the set of all available affine transformation.

I : is any member in D.

T_ℓ: is any m x m matrix in F.

Q_o: is the nonlinear quantization operator with $[L_{01},L_{02},\ldots,L_{0m}]$ quantization levels for its corresponding arguments.

Q_1: is the nonlinear quantization operator with $[L_{11},L_{12},\ldots,L_{1m}]$ quantization levels for its corresponding arguments.

S : is the function giving the number of bits required to code k transformed variables.

k : is the number of transformed variables kept, being determined through the constraint.

$T_\ell^{\,t}$: transpose of T_ℓ

C : is the function which gives the number of bits required for the controlled redundancy consisting of the specifications inserted in the bit string for fast retrieval and the bookkeeping information.

The above formulation includes a class of data compression systems,if appropriate identifications are made for the transformation matrix, the function S and the parameters involved. Two specific examples of the deterministic type will be mentioned.

Example 1: Differential Pulse Code Modulation (DPCM) coding:

Let

$$T_\ell = \begin{pmatrix} 1 & & & & \\ -1 & 1 & & 0 & \\ 0 & -1 & 1 & & \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

$C(T_\ell,k) \equiv 0$

$S(y)=b_1(y_1)+b_2(y_2)+\ldots+b_n(y_n)$

where $b_i(y_i)$ equals the number of bits used for data sample y_i.

and delete the minimization and expectation operators in equation A, the result will be the familiar case of DPCM.

Example 2: Block Quantization Coding:

Let $T_\ell \varepsilon F$ and $k=C_1$, a fixed constant, $C(T_\ell,k)\equiv 0$,

$Q_0(y)=[Q_{01}(y_1),Q_{02}(y_2),Q_{03}(y_3)+\ldots.]$, $S(x)$ same in Exp. 1,

the result is the block quantization coding scheme.

A global solution of $b_{min}(Q_0,X,T_\ell)$ is not readily available, partly due to the computational effort for the large amount of data volume such as the multispectral scanner data and partly due to the nonstationarity of the data, so that any strategy proven to be good for a subset of data such as I_i, for any i, may not be good for another subset. However, the absolute minimum of $b(Q_0,X,T_\ell)$ for a given piece of data $I \subseteq D$, subject to all available variations of the parameters and transforms in the set F, is not of great importance anyway, because usually a suboptimum algorithm with simpler manipulations on the data may be found with performance which may differ only slightly from the optimum result (Tasto, 1972).

Experimental results for two specific cases are compared and presented here. The first case is regular DPCM and the second case involves a transformation followed by a subsequent DPCM and is referred to as differential transform coding. It consists of two sections (1) adaptive error control section and (2) the residue difference correction section.

IV. DIFFERENTIAL PULSE CODE MODULATION (DPCM) AND QUANTIZATION WITH FOLDING TECHNIQUE

For DPCM coding the data to be coded is the difference between two correlated data samples. The quantization requirements will change when the data range is changed. Even through the variance of the difference data may be smaller than the original data, the range spanned by the data may now be even larger. In fact, 9 bits are needed to encode the data exactly. Otherwise, one will be faced with either coarse quantization or with the problem of quantizer saturation, or rather, dynamic overload. The percentage of these overload occurrences can be made as low as desired by designing the quantizer to match the data, once the probability density function of the data (or probability mass function for discrete data) is completely known. However, usually only the sample probability density function of a selected subset is assumed known. It is not efficient to have a quantizer with many bands having an extremely low probability of occurrence, and it is also unlikely that the quantizer can be effectively changed from time to time. Therefore, in the effort of trying to code the data without error while not having to use a large number of codes,a folding technique is proposed. By reserving two codewords in the codebook to do folding when the dynamic range of the data undergoes some sudden changes out of the regular operating bounds, the saturation error can be completely eliminated while the number of codes can remain to be small. Only a rough idea about the probability density function is needed to design the codes. This is especially effective for variable length codes where shorter codes can be used for the normal operating range of the data. The number of codes needed for coding a 9-bit DPCM data can be reduced from 512 to 16 codes. The performance of DPCM with folding for ERTS-1 for 4 channels using 16 predesigned variable length codes is contained in Table 1. Also contained in Table 1 are the optimum results if the probability mass function of the data is found through examining all of the data and the variable length Huffman codes thus generated are used to code the data. If 4-bit fixed length codes are used with the folding technique, 4.5 bits are needed. This is not a fixed value due to the fact that it varies with the number of times folding of data occurs. If no folding occurs for certain areas of the data, four bits per data per channel are needed.

V. DIFFERENTIAL TRANSFORM CODING

A. ADAPTIVE ERROR CONTROL SECTION

The multispectral data can be considered as a three-dimensional random process $x(n,m,\ell)$ defined over $1\leq n\leq J$, $1\leq m\leq K, 1\leq \ell\leq L$ and $0\leq x\leq N$. The first step is to partition the data into a convenient format. In order to take advantages of both the spectral domain structure as well as the two-dimensional spatial structure, a nxmxl array of data can be reindexed to form a one-dimensional vector data (Wintz, 1972).

After segmentation the multispectral data will be arranged in a vector form with each entry of the vector representing one picture element. Each pixel can take on any one of the intensity levels, M. M is limited by the number of bits r assigned to each sample, $M=2^r$.

Let Y be the vector of the data in its transformed domain and X be the data vector. Then,

$$Y=[T_\ell]X$$

where T_ℓ is the selected orthonormal transformation matrix.

From a quantized and truncated version of Y, e.g., $\hat{Y}=(Y_1,Y_2,\ldots,Y_k,0,0,\ldots,0)$, nonzero up to k terms, X is reconstructed. Let $\hat{X}$ be the reconstructed data, we have $\hat{X}=[T_\ell]^t\hat{Y}$.

The number of terms kept in Y, e.g. k is chosen according to how it is desired to control the difference function $e(X,\hat{X})$. Therefore, the number of bits assigned to each block is dependent on the convergence rate of the block data under the prescribed difference criterion.

B. Choice of Transformation Matrix

There is no restriction as to the type of transforms in set F to be used, except, perhaps, a limitation on the matrix size and the tolerable computational effort involved in obtaining the transformation results. The common choices are Fourier, Hadamard, Karhunen-Loeve, Haar and Slant transforms. Fast transform algorithms exist in the implementation of some of these transforms (Pratt, 1969). Only results for optimum transform is contained in this investigation due to the fact that other transforms will not result in better performance (Habibi, 1971). The optimum transform for minimizing the number of transform samples needed for best reconstruction in mean square error sense is the Karhunen-Loeve (K-L) transform. Some other optimum properties of this transformation can be found in (Okamoto, 1968). The K-L transformation matrix is often defined as being composed of eigenvectors of the covariance matrix of the data. However, if the mean of the data is not zero, the correlation matrix and covariance matrix are different and their corresponding normalized matrices are also different. This fact leads to four different transformation matrices and the principal components are not invariant if the data are manipulated by an affine or scale transformation. The performance of applying the four sets of basic functions thus generated can be compared. It is found that scaling of the data results in a more stable probability density function for the transformed data, i.e., the change of the dynamic ranges of the coefficients from one area of data to another does not jump violently. Also, since the correlation matrix rather than the covariance matrix is used, the effort in first obtaining and then subtracting the sample mean from the data can be skipped in the transformation of the data. Hence computational complexity can thus be reduced.

C. RESIDUE ERROR CORRECTION SECTION

In this section the replica of the original data is reconstructed and the difference is found and coded, which consists of errors introduced through quantization, saturation of quantization levels and truncation of the transformation coefficients. In the transform method, even if all the transform terms are kept, a certain amount of quantization errors will still occur. Mean square error between the replica and the original that can be tolerated is usually used as a criterion to determine the number of bits needed for each coefficient and the number of transform terms kept which is fixed as in the case of block quantization coding (Huang and Schultheiss, 1963). The error, however, cannot be effectively controlled in this fashion and may be concentrated on certain areas and, therefore, vital information may be lost even though the overall mean square error may be small. The bounds set for the maximum deviation from the original data for the reconstructed replica can be used in addition to the mean square error criterion. (1) Set an absolute bound for uniform convergence of every data point, i.e., $|X-\hat{X}|\leq\varepsilon$ where ε may be several gray levels. (2) Set an absolute bound for only α% of the data, that is, the bounds can be set in a probability sense such that $Prob(|X-\hat{X}|\leq\varepsilon)=\alpha$, where α can be arbitrarily set according to needs. The convergent property of each block processed under this criterion is different; the number of terms needed vary in the reconstruction of the replica. The distribution of the number of coefficients kept for absolute bounds 2 and 8 are plotted in Figure 1, where B=8 actually means $-7\leq X-\hat{X}\leq 8$ and B=2 means $-1\leq X-\hat{X}\leq 2$. The results for an ERTS-1 subframe (channel 2) is shown in Figure 2 for illustration purposes only. The advantage of a probabilistic description of the absolute error bound is that less number of the transform terms need be kept, see Figure 1, the curve in the middle, and thus results in more compression, see Figure 2 (d), where only 2.6 bits are needed rather than 3.4 bits in Figure 2 (c). Since the result shown is for channel 2 only, which usually has a higher entropy than the other channels, the number of bits required is a little higher than that required if 4 channels are processed. Using variable length codes, the picture data in Figure 2 (b) (c) (d) require 3.3, 1.6 and 2.2 bits per picture element more respectively for error free reconstruction. From the sum of two stages, it can be seen that if more bits are used in the transform domain the less efficient is the data compression. If only a few terms of the transform coefficients are used then the more efficient would be the compression system. In fact, for a 400x400x4 block of ERTS-1 subframe using 0.397 bits/pel/channel for the transform terms and 2.508 bits/pel/channel for the subsequent DPCM (Duan and Wintz, 1973).

VI. CONCLUSIONS

(1) Given an incomplete knowledge of the true probability density function of the data, the concept of folding can greatly simplify the design of the quantizer with only a slight complication for the decoder. The idea of folding takes care of instantaneous increase of the dynamic range encountered.

(2) It is demonstrated that an overall average of around 3.5 bits per data sample per channel are needed for an error free reconstruction of ERTS-1 data with predesigned variable length codes and using folding technique.

(3) It is found that one can take advantage of both the transform techniques and DPCM in the error free coding of the data. This is done by first coding the data through a few terms of the transform coefficients so that the data can be reconstructed with controlled error (or rather controlled difference between the original and reconstructed versions) and then coding the element difference. This is especially true when the number of bits used for the transform terms is small.

(4) Further compression can be done through transform techniques and with some assurance of the absolute deviation of the reconstruction data such as

through the probabilistic description. Varying degree of compression can be achieved.

(5) At present it seems that the transform technique is so much more complicated compared to DPCM that it may be hardly implemented. However, since the multispectral data is eventually going to be processed by machine through transformation, it may prove to be useful to code the transform terms if specific transformation proven to be good for, say, a certain classification purpose, can be defined.

REFERENCES

(1) Duan, J.R. and Wintz, P.A. 1973. "Error Free Coding", Information Note 022073 The Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

(2) A. Habibi and P.A. Wintz, 1971. "Image Coding by Linear Transformations and Block Quantization", IEEE Trans. Commun. Technol., Vol. COM-19, pp. 50-60, Feb. 1971.

(3) J.J.Y. Huang and P.M. Schaltheiss, 1963. "Block Quantization of Correlated Gaussian Random Variables", IEEE Transactions on Communication Systems, Vol. CS-11, Sept. 1963.

(4) Laboratory for Applications of Remote Sensing (LARS), 1973. The Multispectral Scanner. Focus, No. 1. Purdue University, Lafayette, Indiana.

(5) Okamoto, M. 1968. Optimality of Principal Components. In Multivariate Analysis II, ed. P.R. Krishnaiah.

(6) Pratt, W.K. and Andrews, H.C., 1969. "Application of Fourier-Hadamard Transformation to Bandwidth Compression", presented at MIT Symp. on Picture Bandwidth, April 1969.

(7) P.J. Ready and P.A. Wintz, "Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery", IEEE Transaction of Communications Technology. (to appear)

(8) Shannon, C.E. 1949. A Mathematical Theory of Communication. Urbana: University of Illinois Press.

(9) Tasto, M., and Wintz, P.A. 1971. Image Coding by Adaptive Block Quantization. IEEE Transactions of Communications Technology, Com-19:957-71.

(10) Wintz, P.A. 1972 Transform Picture Coding. Proceedings of the IEEE. 60-7:809-820.

(11) University of Southern California, 1972. Bibliography on Digital Image Processing and Related Topics. Electronic Sciences Laboratory, USCEE Report 410.

(12) Wilkins, L.C. and Wintz, P.A. 1971. Bibliography on Data Compression, Picture Properties, and Picture Coding. IEEE Transactions on Information Theory, IT-17:180-97.

(13) Rosenfeld, A. 1968. Bandwidth Reduction Bibliography. IEEE Transactions on Information Theory, IT-14:601-02.

(14) Pratt, W.K. 1967. A Bibliography on Television Bandwidth Reduction Studies. IEEE Transactions on Information Theory, IT-13:114-15.

LARS run number and Flightline ID	Data Volume	Standard codes	Optimum codes
		bits/sample/channel	
72063500 102715233	1172x784x4	3.77	3.54
72059000 110617504-4	601x580x4	3.48	3.22
72051000 103716244	1873x652x4	3.33	2.93

Table 1 Comparison of Performance of DPCM Compression System for data taken at a different time and location.

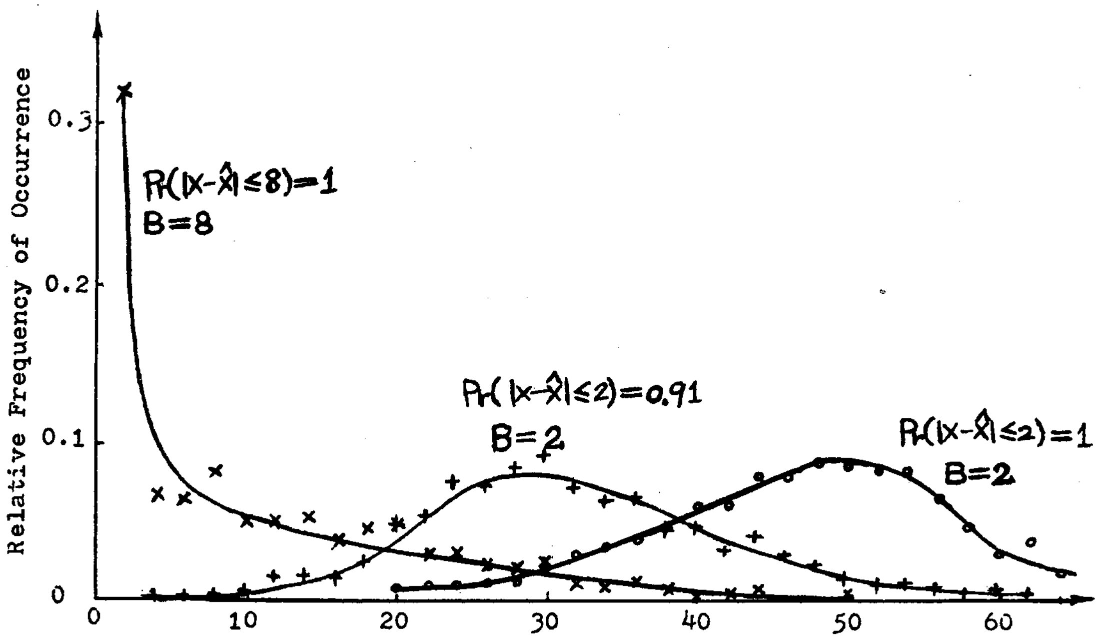
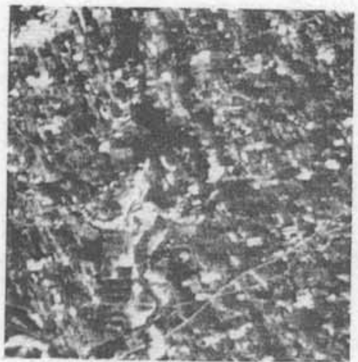


Figure 1 Distribution of the Number of Coefficients needed to ensure uniform convergence of every data point.



(a) Original

(b) 1.1 bit/pel
$E(X-\hat{X})^2=6.06$ with
$-7 \leq (X-\hat{X}) \leq 8$

(c) 3.4 bits/pel
$E(X-\hat{X})^2=0.43$ with
$-1 \leq (X-\hat{X}) \leq 2$

(d) 2.6 bits/pel
$E(X-\hat{X})^2=1.20$ with
$\mathrm{Prob}(-1 \leq (X-\hat{X}) \leq 2)=0.91$

Figure 2. The Original and Reconstructed Pictures for Channel 2 of ERTS-1 Multispectral Scanner Data