Security Classification

DOCUMENT CONTROL DATA – R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
North Carolina State University Department of Electrical Engineering Raleigh, North Carolina 27607	UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

DATA COMPRESSION PANEL DISCUSSION

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Jay W. Schwartz Thomas S. Huang Jacob Ziv
Lee D. Davisson J. B. O'Neal, Jr.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
4 April 1972	5	29

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
F44620-69-C-0033 b. PROJECT NO. 7921 c. 61102F d. 681304	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR - TR - 72 - 0615

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research 1400 Wilson Boulevard (NM) Arlington, Virginia 22209

13. ABSTRACT

Data compression, in this paper, is used to denote the reducing of an input data set prior to transmission, as opposed to data reduction, the analytical processing of the set upon reception. Source encoding techniques are classified into sampling and analog to digital conversion, codebook techniques, predictive subtractive coding and delta modulation, along with aperature and partitioning techniques. Some of the most recent work discussed includes adaptive predictive coding, picture bandwidth compression-source coding, and a techniques for subjective performance measures. The paper concludes with an information theoretic approach to data compression.

AD 740615

DD FORM 1 NOV 65 1473

Security Classification

DATA COMPRESSION PANEL DISCUSSION

Jay W. Schwartz, Chairman

Institute for Defense Analyses
Arlington, Virginia 22202

Lee D. Davisson*
University of Southern California
Los Angeles, California 90007

J. B. O'Neal, Jr.
North Carolina State University
Raleigh, North Carolina

Thomas S. Huang
Massachusetts Institute of Technology
Cambridge, Massachusetts

Jacob Ziv**
Bell Telephone Laboratories
Murray Hill, New Jersey

I. INTRODUCTION

The term data compression is used to denote operations which are performed to reduce data prior to transmission. This technology has also been called by other names: "data compaction", "bandwidth compression", etc. The term data reduction usually refers to operations performed at a processing center to extract information in a convenient form for interpretation. Data compression results when data reduction is performed at the transmitting end.

It is convenient to differentiate between two forms of data compression: extraction and source encoding. Extraction is a term for data reduction performed at the transmitter rather than the receiver in order to reduce the amount of data transmitted. It refers to operations which "extract" desired information, casting aside unimportant or less important properties of the data. Examples of extraction operations performed in space programs are the calculation of moments, discarding samples with values below a threshold, and discarding samples which do not represent a local maximum. Pattern recognition performed prior to transmission is a form of extraction. In contrast to source encoding, extraction operations normally would be performed at the receiver as data reduction if they were not performed at the transmitter. Source encoding refers to operations which produce outputs (encoded data) from which the input data can be reconstructed, possibly with distortion. To obtain data compression, source encoding relies on knowledge of source statistics or on tolerance for distortion at the receiving end. Unlike extraction operations, source encoding operations usually would not be considered as a step in data reduction; indeed the first step in data reduction at the receiver probably is to undo them: i.e., to decode. Extraction in general has potential for much greater compression of data than does source encoding. However, extraction techniques often are outside the interests of communication theory. Therefore this discussion will be confined to source encoding.

Shannon[1] has considered the problem of coding an information source provided that there is a way to measure distortion of the final recovered information at the receiving point. He showed that a function R(D), called the rate distortion function, exists which is a measure of the information rate required to specify the output of a source with a distortion less than D. That is R(D), in bits per source symbol, is the minimum data rate which can be used to describe the information with a distortion less than D. This concept can form the basis of a quantitative approach to source encoding in the following manner. A coding scheme would be evaluated for use with a source by evaluating data rate as a function of the distortion measured after encoding and decoding. Plots for various coding techniques would be compared with each other, and with R(D) for that source. To select the "best" scheme for a particular application, one would specify a maximum distortion, D^*, and compare performance of the various coding schemes under the constraint that $D \leq D^*$. Further, one would compare this performance with $R(D^*)$ to determine how much better one might do with better schemes. The primary difficulty in realizing this systematic approach is in finding suitable distortion measures, but another difficulty is that to make the problem tractable, it must be possible to divide a message into independent portions. Neither difficulty is easily overcome. For example, a picture seldom can be divided into independent segments. Aside from a few special cases, investigations in conjunction with the rate distortion function assume a mean-square-error distortion measure on a sample-by-sample basis.[1,2,3,4]

This is not always an adequate measure of distortion. Lacking ways to measure distortion quantitatively as would be required for the approach outlined above, subjective measures are used by each researcher to determine which reconstructed pictures are "adequate" and which are "better" than others. Similarly, for voice messages "articulation" or "intelligibility" is often used to evaluate reconstructed sound

II. CLASSIFICATION OF SOURCE ENCODING TECHNIQUES

Sampling and A/D Conversion Sampling can provide data compression, but usually only when the desired information has a very low bandwidth compared to the available channel: for example, a temperature that can be adequately described by one sample an hour. When it is necessary to represent an analog function such as a TV scan or a voice message by PCM, it is usually found that sampling

* On leave of absence from Princeton University.

** On leave from the Scientific Department Israel Ministry of Defense.

necessitates greater communication capacity than the original analog message. Certain implementations of "aperture" techniques, discussed presently, may be considered as employing asynchronous sampling to achieve data compression.

Source coding techniques for data compression usually operate on digital data: some because they must, others because they can be instrumented more conveniently for digital than for analog data. Therefore, when evaluating a source encoding technique, it is sometimes necessary to penalize its performance with the cost, in terms of bandwidth and/or signal power, of analog to digital conversion. Moreover, sampling rate, quantization, and reconstruction procedure often can affect the relative performance of source encoding techniques. In some cases investigators are not free to experiment with these parameters as freely as they should: for example, TIROS pictures which have been used in evaluating techniques often have been sampled and reconstructed before being made available.

Code-Book Techniques Certain source encoding techniques can be considered to represent each source word with a code word. Some familiar examples of the techniques which fall in this class rely on the statistics of the source symbols: the Shannon-Fano[5], Huffman[6], and Morse code. Gilbert and Moore[7] discuss a number of these codes. Other examples rely on an understanding of tolerable distortions rather than source statistics: for example, sophisticated quantization schemes and companding.[8,9]

A few of these code-book techniques have found wide practical usage: the Morse Code and logarithmic companding are probably the best examples. To take advantage of correlation from one source word to another conveniently, code book techniques may be used in conjunction with other forms of source encoding.

Predictive Subtractive Coding and Delta Modulation Another general approach to source encoding is to predict a sample value, based on previous samples, and then to transmit the difference between the true and predicted sample values.[10] Of course, some form of code-book technique must be employed on the difference values in order to obtain data compression. The most practical implementation of predictive-subtractive coding is delta-PCM in which the difference values are merely quantized.[11,12] Delta-modulation is the special case of delta-PCM with one-bit quantization.

When compared with most other source encoding techniques, delta-PCM has the important advantage of producing encoded data at a fixed synchronous rate. It has the undesirable property of predictive-subtractive coding that transmission error propagates through succeeding decoded values. The potential compression available with delta-PCM is limited to at best a factor equal to the number of bits per sample in an adequate PCM representation, unless one resorts to supplementary coding which would sacrifice the fixed synchronous data rate.

Aperture Techniques Aperture techniques are similar to predictive-subtractive coding in that use is made of the difference between the true value and an estimated value. With aperture techniques, however, either the true or estimated value is transmitted rather than the difference between them.[13,14] Data compression is obtained by transmitting nothing when the true and estimated values differ by less than a specified amount, i.e., when the true value falls within a certain "aperture" centered on the estimated value.

Aperture techniques can avoid the problem of error propagation that results when differences are transmitted. Furthermore, they can permit the use of interpolation instead of prediction in forming estimates. Encoded data is produced at an asynchronous rate which depends on the success in estimating sample values; because of this, each transmitted value must be tagged to identify the sample which it represents.

Partitioning Techniques Another approach to source encoding is to consider the source data as a combination of distinct parts, and to treat each part separately. The best known partitioning technique is vocoding in which voice messages are divided into component frequencies which are described by the transmitted data. The original voice message is reconstructed, with a loss in fidelity, from the component frequencies. Somewhat similar philosophy gives rise to techniques proposed for encoding television data by treating high and low frequency components separately, or by treating different bits of the sample values separately.[15,16]

For example, if the intensity at each picture element is read to 6-bit accuracy, each sample value can be treated as two values--say, the higher-order 3-bits forming one value and the lower-order 3-bits forming a second value. The rationale behind this approach is that the higher-order (most significant) bits change more slowly, corresponding to lower frequency components, than do the lower-order bits. Carrying this approach further, bit-plane encoding[17] treats each bit of a sample separately, treating a source producing n-bit samples as n binary sources.

III. SOME RECENT WORK IN SOURCE ENCODING

Adaptive Predictive Coding* (L.D. Davisson)

As noted, compression can be attained by encoding the differences between the true and predicted sample values where the prediction is based upon the past as seen at the receiver. The coding is done relative to some distortion measure--usually it is required that the reconstructed sample values disagree from the true values by less than a preset amount.

Ideally, such a predictor would minimize the message entropy. As it is not known how to do this, one must resort to simpler quantities, usually mean square error minimization[18] but also possibly conditional probability mode maximization.[19] The prediction can in principle be nonlinear as well as linear, employing polynomials in the sample values[18,19] or conditional histograms.[20] In many cases linear methods are sufficient[20] whereas in some cases profitable use can be made of nonlinear methods.[21]

The optimization of the predictor can be done in advance if the data distributions are

*This work was supported by Joint Services Electronics Program administered by the Air Force Office of Scientific Research, under Grant AF-AFOSR-67-1622-A. F44620-69-C-0033

known or if appropriate quantities can be estimated from "representative" samples under the assumption of stationarity. In most applications such information will be only roughly available necessitating the use of adaptive methods if nearly optimum performance is to be attained.

One simple adaptive predictor[20,81] of the form suggested by Widrow[88] and Lucky[83] for different applications depends on local gradient following and provides good performance while being quite practical to implement. Suppose that the data are a time series $\{s_n\}$ with received values (in the absence of channel errors) $\{y_n\}$. The n^{th} predicted value is a linear weighting of the immediate past M values:

$$\hat{s}_n = \sum_{j=1}^{M} \alpha_j^n y_{n-j}.$$

The n^{th} prediction error is then

$$e_n = s_n - \hat{s}_n = s_n - \sum_{j=1}^{M} \alpha_j^n y_{n-j}$$

The weight vector $\{\alpha_j^n\}$ is adjusted after each prediction by the following algorithm

$$\alpha_j^{n+1} = \alpha_j^n + \gamma e_n y_{n-j} \qquad j = 1,2,\cdots,M$$

where γ is a small positive constant. Still simpler versions result by replacing e_n, y_{n-j} or both by their signs. It can be shown that the mean square prediction error on a stationary time series is

$$\sigma^2[M;\gamma] \leq \sigma^2[M](1+M\gamma\lambda) + o(\gamma)$$

where $\sigma^2[M]$ is the minimum variance using the optimum predictor and $\lambda = \overline{y_n^2}$. It is seen that nearly optimum performance is attained by the adaptive predictor while retaining simplicity and ability to follow data nonstationarities.

Picture Bandwidth Compression-Source Coding
(T.S. Huang)

A symposium on picture bandwidth compression was held recently at the Massachusetts Institute of Technology. It was the first symposium devoted entirely to the field, and was intended to present the current status of television bandwidth compression and of related areas. The large and enthusiastic response to this meeting indicates the growing interest in the area.

Many picture bandwidth compression schemes have been proposed and tried out. Most of these operate on intra-line redundancy and achieve bit rates of around 3 bits/sample. (Direct PCM transmission of pictures requires 7 or 8 bits per sample.) One can attain lower bit rates by operating on two-dimensional redundancy. The most recent work in this direction has been done by mapping portions of a picture with a linear transformation, and by quantizing the transformed picture through some optimum strategy.[14,18] A procedure that Fourier transforms blocks of a picture and codes the transformed data with an adaptive procedure promises to produce good pictures with about one bit per sample.[26]

In transmitting motion pictures, one can exploit the frame-to-frame redundancy. Because of the formidable experimental problems, however, very little work has been done in this area until recently.[37,38] Similar reasons account for the scarcity in studies on color picture coding. One such study[39] suggests that a color picture can be sent at about the same bit rate as an equivalent-quality monochrome picture.

Looking into the future, we expect to see experimental studies on motion and color picture coding play a growing role. On the theoretical side, it is hard for us to see any significant progress. Although Shannon's rate distortion theory seems to provide an ideal mathematical background for studying picture bandwidth compression, the possibility of quantifying picture quality so that we may have a suitable fidelity criterion is as remote as ever. Let us hope that studies in visual pattern recognition may throw some light on this knotty problem.

Bounds on Subjective Performance Measures
(J.B. O'Neal, Jr.)

Quantizing noise is present whenever analog information is encoded into digital form suitable as an input to any digital system such as a computer or digital transmission line. Bounding the impairments caused by this noise is one of the ubiquitous activities of those involved in information theory. The performance metric to be bounded is of primary importance. To be useful it must measure the quality of the system in the eyes of its users. Bounds on the ratio of signal power to quantizing noise power have been derived. For speech and television signals, however, it has long been known that noise power alone is not a definitive measure of signal quality. For these signals, noise at low frequencies is more damaging than noise at higher frequencies. Frequency weighting networks have been designed from subjective tests which can be used to accurately measure the detrimental effect of noise. The attenuation of these networks falls off with frequency in such a way that, when noise is passed through the network, the output power is an accurate measure of the subjective effect of the noise. The ratio of signal power to frequency weighted noise power is a much better performance metric than the usual signal to noise ratio. Bounds have been derived for S/N_Y, the ratio of signal power to frequency weighted noise power for speech and television signals. Values of S/N_Y greater than the bound are not possible for any encoding system.

When expressed in dB the bounds on signal power to frequency weighted noise power are the sum of three terms: $S/N_Y \leq T_B + T_P + T_S$. T_B is a term which increases linearly with bit

rate. T_F is determined by the statistical nature of the signal. It is a measure of the predictability (or redundancy) of the signal. T_S is a quantity determined by the appropriate subjective frequency weighting function. The terms T_B, T_P and T_S and the effects measured by each one are relatively independent of the others, i.e., T_B is dependent only on the bit rate, T_P is determined only by the predictability of the signal and T_S is determined only by subjective considerations (as embodied in the frequency weighting function).

In order to obtain numerical results for bounds of this type, values of T_P must be estimated for practical signals like speech, television and still pictures. This is done at some risk for T_P depends on the statistics of the class of signals for which the systems are to be used. The best we can do is to estimate T_P for, what we believe at this time to be, "typical" signals and calculate bounds based on these "typical" signals. The calculation of T_S can be made for speech using the C message weighting characteristic and for television by using the standard frequency weighting functions of Barstow and Christopher.

An Information Theoretic Approach to Data Compression
(J. Ziv)

Suppose that we have a data source which emits a sequence of symbols $x_1, x_2, x_3 \in \chi$ (an arbitrary set) at a rate of ρ_S per second. This sequence is fed into an "encoder" which assigns to each successive block of n source symbols, say $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$, a channel input of duration of $n/\rho_S = T$ seconds. At the receiving end of the channel, the T-second output is transformed by a "decoder" into an n-sequence, say $\underset{\sim}{\hat{x}} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$, which is delivered to the destination. The "distortion" between the source output sequence $\underset{\sim}{x}$ and the received sequence $\underset{\sim}{\hat{x}}$ is defined as

$$d^{(n)}(\underset{\sim}{x}, \underset{\sim}{\hat{x}}) = n^{-1} \sum_{k=1}^{n} d(x_k, \hat{x}_k),$$

where $d(x,\hat{x}) \geq 0$ is an arbitrary function.

The following distortion functions are considered:

(a) $d(x,\hat{x}) = |x-\hat{x}|^s \quad (s > 0)$
when χ is a subset of the reals.

(b) $d(x,\hat{x}) = \begin{cases} 0, & x = \hat{x} \\ 1, & x \neq \hat{x} \end{cases}$
when χ is a discrete (i.e., countable) set.

(c) $d(x,\hat{x}) = \begin{cases} 0; & |x-\hat{x}| \leq \sigma \\ 1; & |x-\hat{x}| > \sigma \end{cases}$
when χ is a bounded set typically the interval $\left[-\frac{A}{2}, \frac{A}{2}\right]$.

The classical problem is that of a "memoryless" source, where successive source outputs are statistically independent with identical probability distribution. In this case it is meaningful to let the system performance criterion (fidelity criterion) be the statistical expectation of the distortion $d^{(n)}(\underset{\sim}{x}, \underset{\sim}{x})$. A quantity of interest is $\bar{d}^*(T)$, the smallest attainable value of the fidelity criterion when the coding delay is T seconds.

The following topics are of interest:

A) The possibility of achieving $\bar{d}^*(T)$ by an A/D converter followed by an optimal digital encoder.
B) The possibility of using a uniform quantizer as the A/D converter.

REFERENCES

1. C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," Information and Decision Processes, R. E. Machol, Ed., McGraw-Hill, 1960.
2. A. M. Gerrish and P. M. Schultheiss, "Information Rates on Non-Gaussian Processes," IEEE Trans. on Information Theory, Vol. IT-10, No. 4, pp. 265-271, October 1964.
3. T. J. Goblick, Jr., "Theoretical Limitations on the Transmission of Data from Analog Sources," IEEE Trans. on Information Theory, Vol. IT-11, No. 4, pp. 558-567, October 1965.
4. R. A. McDonald and P. M. Schultheiss, "Information Rates of Gaussian Signals Under Criteria Constraining the Error Spectrum," Proc. IEEE, Vol. 52, pp. 415-416, April 1964.
5. C. E. Shannon, "A Mathematical Theory of Communication," Bell Systems Technical Journal, Vol. 27, July-October 1948.
6. D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," Communications Theory, W. Jackson, Ed. London: Butterworth, 1953, pp. 102-110.
7. E. N. Gilbert and E. F. Moore, "Variable-Length Binary Encodings," Bell Systems Technical Journal, Vol. 28, pp. 933-967, July 1959.
8. B. Smith "Instantaneous Companding of Quantized Signals," Bell Systems Technical Journal, May 1957.
9. D. H. Schaefer, "Logarithmic Compression of Binary Numbers," Proc. IRE, Vol. 49, p. 1219 1219, July 1961.
10. E. M. Oliver, "Efficient Coding," Bell Systems Technical Journal, Vol. 31, pp. 724-750, July 1952.
11. J. B. O'Neal, Jr., "Delta Modulation Quantizing Noise Analytical and Computer Simulation Resules for Gaussian and Television Inputs," Bell Systems Technical Journal, Vol. 45, pp. 117-141, January 1966.

[15] J. B. O'Neal, Jr., "Predictive Quantizing Systems (Differential Pulse Code Modulation) for the Transmission of Television Signals," Bell Systems Technical Journal, Vol. 45, No. 5, pp. 689-721, May-June 1966.

[16] B. Julesz, "A Method of Coding Television Signals Based on Edge Detection," Bell Systems Technical Journal, Vol. 28, pp. 1001-1028, July 1959.

[14] L. W. Gardenhire, "Redundancy Reduction the Key to Adaptive Telemetry," National Telemetering Conference, June 1964.

[13] E. R. Kretzmer, "Reduced-Alphabet Representation of Television Signals," IRE International Conv. Rec., Pt. 4, pp. 140-147, 1956.

[12] W. F. Schreiber and C. F. Krapp, "TV Bandwidth Reduction by Digital Coding," IRE International Conv. Rec., Pt. 4, pp. 88-99, 1958.

[17] J. W. Schwartz and R. C. Barker, "Bit-Plane Encoding: A Technique for Source Encoding," IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-2, No. 4, pp. 385-392, July 1966.

[18] A. V. Balakrishnan, "An adaptive Nonlinear Data Predictor," Proc. NTC, 1962.

[19] L. D. Davisson, "The Theoretical Analysis of Data Compression Systems, "Proc. IEEE, February 1968.

[20] L. D. Davisson, "Theoretical Considerations in Data Compression," MIT Picture Bandwidth Compression Symposium, MIT Press (to be published).

[21] L. D. Davisson, "An Application of Data Compression Concepts," Proc. Purdue Symposium on Information Processing, to be published.

[22] B. Widrow, "Adaptive Filters I: Fundamentals, "Stanford Elect. Lab. Rept. SEL-60-126, December 1966.

[23] R. W. Lucky, "Techniques for Adaptive Equalization of Digital Communications Systems," BSTJ, February 1966.

[24] W. K. Pratt, and H. C. Andrews, Application of Fourier-Hadamard Transformation to Bandwidth Compression. Proceedings of Picture Bandwidth Compression Symposium, April 2-4, 1969, Cambridge, Mass., to be published by M.I.T. Press.

[25] J. Woods, and T. S. Huang, Picture Bandwidth Compression by Linear Transformation and Block Quantization, Proc. of Pict. Band. Comp. Symp., op. cit.

[26] G. B. Anderson, and T. S. Huang, Picture Bandwidth Compression by Piecewise Fourier Transformation, Proc. of Purdue Univ. Centennial Symposium on System Sciences, Lafayette, Indiana, April 1969.

[27] J. E. Cunningham, Frame-Correction Coding, Proc. of Pict. Band. Comp. Symp., op. cit.

[28] F. W. Mounts, Frame-to-Frame Processing of TV Pictures to Remove Redundancy, Proc. of Pict. Band. Comp. Symp., op. cit.

[29] U. F. Grovemann, Coding Color Pictures, Proc. of Pict. Band. Comp. Symp., op. cit.