A SURVEY OF
PICTORIAL DATA-COMPRESSION TECHNIQUES

Jack L. DiGiuseppe
A SURVEY OF PICTORIAL DATA-COMPRESSION TECHNIQUES

Jack DiGiuseppe

CONCOMP: Research in Conversational Use of Computers
ORA Project 07449
F.H. Westervelt, Director

supported by:
DEPARTMENT OF DEFENSE
ADVANCED RESEARCH PROJECTS AGENCY
WASHINGTON, D.C.

CONTRACT NO. DA-49-083 OSA-3050
ARPA ORDER No. 716

administered through:
OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

March 1969
ABSTRACT

The results of a survey of pictorial data compression techniques are summarized in this report. The survey was motivated by a study of half-time graphics communication over voice-grade lines. The principal compression techniques surveyed include the following: the optimization of the parameters of ordinary pulse coded modulation, pulse coded modulation using added pseudo-random noise, differential pulse coded modulation, predictive differential pulse coded modulation, run-length encoding, brightness contour detection and encoding, area encoding, picture compression in the Fourier domain.
# TABLE OF CONTENTS

**ABSTRACT** ........................................ iii

1. CONTEXT OF THE PROBLEM ................................ 1

2. PREVIOUS WORK ........................................ 6
   2.1 PURE PULSE CODED MODULATION ....................... 7
      2.1.1 Prefilter Characteristics ...................... 9
      2.1.2 Scanning Pattern ............................... 10
      2.1.3 Sampling Rate .................................. 11
      2.1.4 Brightness Quantization ....................... 13
      2.1.5 Transmission Code .............................. 16
      2.1.6 Post filter Characteristics ................... 18
      2.1.7 Summary ....................................... 18
   2.2 DIFFERENTIAL PULSE CODE MODULATION ............... 19
      2.2.1 Ordinary DPCM .................................. 19
      2.2.2 Predictive DPCM ................................ 20
      2.2.3 Advantages of DPCM ............................. 22
      2.2.4 Disadvantages of DPCM ......................... 23
      2.2.5 Summary ....................................... 23
   2.3 PCM USING FREQUENCY SEPARATION .................... 23
   2.4 PROPERTY DETECTION AND ENCODING ................... 25
      2.4.1 Run-Length or Differential-Coordinate Encoding ............ 25
      2.4.2 Restricted Run-Length Encoding ................. 28
      2.4.3 Dual Mode and Edge Detection Systems ............ 33
      2.4.4 Contour Detection and Encoding ................ 39
      2.4.5 Area Encoding .................................. 43
   2.5 COMPRESSION TECHNIQUES USING PICTURE TRANSFORMATION 44

**REFERENCES** ........................................ 47
LIST OF FIGURES

Figure 1. Some Results of Michel's Run-length Encoding (24) .............. 27
Figure 2. Waveforms of the Synthetic-Highs System (38) ..................... 35
Figure 3. A one-Dimensional Example of the Synthetic "Highs" System for Step Function Input (44) .............. 40
LIST OF TABLES

Table I. Contour Coding for Figures 16, 17, and 18(44) . . 41
1. CONTEXT OF THE PROBLEM

The general problem to be addressed is the efficient storage and transmission of graphical, textual, and pictorial data. The data are assumed to exist initially in the form of "hard copy," such as on paper, and are to be converted to a form which can be stored on devices accessed by a high-speed digital computer. It is intended that the data be stored in a central data processing facility which can be accessed rapidly by many remote stations, each containing a device suitable for receiving and displaying the data.

It is further assumed that the data will not be radically changed at the remote station; however it is hoped that the ability to make limited modifications will be included. Therefore, the scheme used to convert the data from external "hard copy" form to its internal stored form need not include the ability to incorporate efficiently modifications since these will usually be few in number. In addition, the conversion scheme need not be fast in terms of processing time, since the efficient conversion of the data from hard copy to computer storage will normally only occur once. This would not be the case if the data were continually subjected to modification in an interactive mode at the remote terminal. In such a case, after each modification, the data would have to be reconverted to an efficient internally stored form during the interactive session. This would call for a very rapid
conversion scheme, since it is assumed that after each modification of the data an updated version would have to be transmitted to the remote terminal. One possibility, if the modifications are minor during a given terminal session, would be to efficiently reconver the data at the end of a terminal session. This would avoid the degrading effect of cumulative minor modifications to a given datum.

The type of problem I am discussing arises in the storage and retrieval of data which have been accumulated from various sources. In general, these data have been assembled to provide rapid access of previously published material to a large number of individuals. These individuals do not wish to modify the data; they wish to search through the data for information about which they may know only certain key words—an author or a date, for example—which are associated with the data. Each item of information, such as a page of a journal article, available within such a system would be converted only once into an efficient internally stored form. Most of the information stored in such a system would consist of pure text, which could be keypunched for card input or read and converted to some character code, such as ASCII, by optical character-recognition devices. For text this is undoubtedly an efficient internally stored form. However, a great deal of the information to be stored would consist of a combination of pictures, drawings, tables, graphs, and text.
There is no obvious efficient way to store and transmit this mixed type of information.

This kind of information would also commonly be encountered in an industrial organization where a central store of managerial and engineering data could be accessed by a manager or an engineer who wishes to view previously prepared graphs, charts, or drawings without necessarily wishing to modify them.

In the two examples given above, an important factor is the time required to access and display the desired data. This is especially true in an information retrieval system where the user is led from one piece of data to the next, such as when one book references another, or a given keyword leads to a whole list of possible articles and each article must be examined individually. From a "human factors" standpoint the system must be capable of responding rapidly to the user.

I am not concerned here with the time required to locate a piece of data in the various mass storage devices of the system or with the way in which such files should be organized to minimize this access time. The problem I am interested in is the time required to transmit the accessed data and display it at the remote terminal. This transmission time is, of course, a function of the bandwidth of the transmission link, which brings in another factor.
the cost. In order for the kinds of systems that have
been mentioned to be practical they must be relatively
inexpensive on a per-terminal basis. One of the costs
associated with the terminal is that of the communications
link between the remote terminal and the central information
facility. It could range from voice-grade public telephone
lines to a private, leased, high-performance communications
link. The cost and bandwidth increase as one goes from the
former to the latter. We would like to keep the transmis-
sion time low and at the same time use an inexpensive com-
munications link. Whether or not this will be feasible de-
pends on the amount of data to be typically transmitted and
the response time desired. Other factors which will appear
later are the coded form of the data and the cost of a re-
 mote terminal capable of decoding and displaying the data.

It is assumed at the outset that the data will be
stored in digital form. This requires that the data be
somehow sampled and coded in terms of sequences of bits.
The first thing that must be decided is the spatial rate
at which a piece of printed matter must be sampled in order
to preserve the information present. I shall start out by
assuming that the material to be digitized is black and white
with no shades of gray. Theoretically if the material is
sampled at twice the highest spatial frequency, or greater,
then it is possible to reconstruct perfectly the material.
Some redundancy is needed to compensate for noise introduced in the sampling, transmission, and reconstruction processes. A limited sample of line drawings and fine print such as pica-typewriter characters yields line-width values of 0.01 inches or larger, with line separation being two or more line-widths. This implies a spatial frequency of 1 cycle/0.03 inches or 33.3 cycles/inch as an upper limit. Therefore, the sampling rate should be 66.7 samples/inch minimum. For convenience, consider a sampling rate of 100 samples/inch. If this rate is used to digitize a standard 8-1/2-inch by 11-inch page, the number of bits that would result would be \((8.5 \times 100) \times (11 \times 100) = 935,000\). It is assumed that the page is entirely black and white and that a 0 bit represents white and 1 bit represents black. If one were to take a very naive approach and attempt to transmit the picture by transmitting all the sampled bits, how long would it take? Using voice-grade telephone lines at the teletype rate of 110 baud or 110 bits per second, it would take 8,500 seconds. At 2,000 baud, it would require 467.5 seconds. Actually a transmission rate of about 500,000 baud would be needed to make the transmission time reasonable for interactive remote terminal usage. But the cost of such a high data rate is incompatible with the goal of an inexpensive terminal system.
Another consideration is the cost of storing 935,000 bits for each digitized page of data. If the 935,000 bits are compacted into bytes, we require 116,875 bytes of storage, or 28.5 pages of storage assuming 4096 bytes per page. This is obviously a prohibitive amount of storage for a single 8-1/2 x 11 page of data and dramatically points out the need for some form of coding and compression of the data.

2. PREVIOUS WORK

A great deal of work has already been done in this area, but the results have been disappointing. Most of the work has been directed toward television bandwidth compression. One measure of success for any data compression or coding scheme is a suitably defined compression ratio. A commonly used definition is the following: let B be the number of bits produced by scanning a given piece of graphic material at a spatial frequency equal to twice the highest spatial frequency encountered in any data to be scanned (this is known as the Nyquist rate); let N be the number of bits that result when the original B bits are compressed and encoded according to the scheme being evaluated; then the compression ratio is $R = B/N$. Most previously reported schemes achieve maximum values of $R$ that range between 30 and 40, with values less than 10 being most common. The previous analysis implies the need for a compression ratio on the order of 200 or 300. This would bring the transmission
time for an 8-1/2 x 11 inch page into the range of 1 or 2 seconds using a 2,000-baud transmission rate. The next few sections will describe some compression techniques which have been investigated.

2.1 PURE PULSE CODED MODULATION

Pulse coded modulation, commonly abbreviated PCM, was invented by Reeves (1) in the 1930's; however, it did not find wide application until very recently. This was primarily due to the unavailability of circuit components capable of high-speed switching and pulse regeneration. PCM has several advantages (2,7) relative to analog transmission techniques:

1. It uses time-division multiplexing in contrast to frequency-division multiplexing commonly used with analog transmission techniques;

2. PCM signals can be transmitted over long distances using pulse regeneration circuits, or repeaters, spaced at regular intervals without cumulative deterioration in the signal-to-noise ratio;

3. PCM techniques are compatible with the transmission of digital data to, from, and between digital computers;

4. PCM signals are easily switched.

The disadvantages of PCM are the cost of terminal equipment and, compared to the original analog signal, the increased bandwidth.
PCM transmits a picture in the following way. The picture is scanned according to some pattern, usually but not exclusively with a television-type raster scan, and samples of picture brightness are taken at regular intervals along the scan pattern. The interval between samples is determined by the resolution requirements of the system. If it is assumed that all picture details are to be exactly reproduced, then the sampling theorem requires that the sampling interval be equal to the period of twice the highest spatial frequency in the picture.

Each sample point generates a sequence of one or more bits. In the case of monochrome pictures, there are two levels, black and white, and each sample point corresponds to one bit, e.g., binary 1 represents black and binary 0 represents white. When pictures containing a continuum of grayness levels are sampled, each sample generates n bits where \( g, (0<2^n-1<g<2^n) \), is the number of grayness levels required to reproduce the picture according to some exactness criterion. Therefore, an essentially infinite number of grayness levels will be represented by a finite number of discrete levels, and it will be impossible to reproduce exactly the picture. In addition, in a picture having abrupt transitions from white to black the spatial frequencies will be quite high, and it will therefore be practically impossible to sample at a rate equal to twice
that of the highest spatial frequency.

Since PCM transmission of pictures yields only an approximation of the original picture, studies have been made to find the optimum parameters for a PCM transmission system.\(^2\,4\,5\,6\,8\) The parameters are evaluated both subjectively by human observers and objectively using various optimization criteria.

The following is a representative list of PCM parameters to be optimized:

1. prefilter characteristics,
2. scanning pattern,
3. resolution or sampling rate,
4. number of brightness quantization levels,
5. transmission code,
6. and postfilter characteristics.

2.1.1 **Prefilter Characteristics**

The purpose of a prefilter \(^8\,9\,10\) is to reshape the power spectral density of the signal in order that the signal may be quantized with greater accuracy. Since the signal itself consists of a continuum of picture brightness amplitudes, it will not be possible to represent precisely the amplitude at each sample point using a finite number of discrete levels. The difference between the signal amplitude and its quantized representation is known as quantization noise. It has been found that prefiltering
in error were pseudorandomly chosen the errors will be pseudorandomly distributed in the reproduced picture and, therefore, highly uncorrelated.

In statistical coding, one scan pattern may have advantages over another. The unconditional probability distribution of brightness levels does not vary for different scan patterns or for a certain picture sample space. However, this is not true for the various orders of conditional probability. For a particular sample space it should be possible to find a set of scan patterns for which the entropy, and therefore the number of bits required for transmission, is a minimum if the optimum statistical code is used. It is questionable whether the savings in transmission would be sufficient to warrant the added complexity. Also it is doubtful that the picture sample space normally encountered would exhibit sufficient stationarity from one picture transmission to the next to allow the use of a single statistical coding scheme.

2.1.3 Sampling Rate

The sample spacing for black and white pictures without grayness levels is determined by the width of the thinnest black or white region which will be encountered. Moreover the human eye must be able to resolve the region under normal viewing conditions. Straight lines of any slope and thickness should be reproduced as straight lines and should
not have a "staircase"-like appearance. In pictures which contain grayness levels, the picture brightness can be described as a function of the two picture planar coordinates $f(x,y)$. If the $x$ components of spatial frequency for all pictures to be sampled are less than some maximum frequency $f_{xm}$, then, according to the Sampling Theorem, a horizontal sampling rate $R_x > 2f_{xm}$ is sufficient to reconstruct those components. If a similar number $f_{ym}$ exists for the $y$ components of spatial frequency, then a vertical sampling rate $R_y > 2f_{ym}$ is sufficient. $R_x$ and $R_y$ expressed in samples per unit length determine the matrix of sample points to be used for PCM. Since the final picture will be viewed by a human observer, it may be possible to limit the sampling rate according to the bandwidth of the human eye with respect to brightness frequency response. This involves research into the psychophysics of human vision, which is beyond the scope of this preliminary paper. The literature on PCM that I have seen does not specify scan rates in sample points per unit length. Usually the size of the sampling raster is given in terms of sample points alone, and the actual dimensions of the raster in inches are not stated. Some common raster sizes that have been used in experimental systems are 128 x 128, 256 x 256, and 512 x 512, the last being appropriate to the PCM transmission of television pictures. A cursory examination of
black and white material, such as an 8-1/2 x 11-inch pica typewritten page or a line drawing, indicates the need for vertical and horizontal sampling rates on the order of 100 sample points per inch.

2.1.4 **Brightness Quantization**

Given a brightness sample one must decide which discrete brightness symbol to use to represent it. For black and white, only two symbols are used. If the brightness sample is above a certain threshold value, one symbol is transmitted, otherwise the other symbol is sent. If grayness levels are to be sent, then it must be decided how many different grayness levels are going to be used. If too few levels are used, the reproduced picture will have false contours and appear to be broken-up into distinct regions of constant brightness. If too many levels are used, then excessive or redundant information will be transmitted and the transmission time will be unnecessarily long. Experiments performed to evaluate various choices for the number of brightness levels indicate that anywhere from $2^4$ to $2^7$ levels can give satisfactory pictures. Powers of 2 are used because normally the brightness symbols are represented by a binary code. To partition the full range of brightness from white to black into intervals of grayness, such that each grayness interval is assigned one brightness symbol, two basic methods are used: uniform and logarithmic
partitioning. Logarithmic partitioning is based on experimental evidence that the human eye's response to light is proportional to the logarithm of the light intensity. Experimental evidence indicates that for a given picture quality logarithmic quantization requires one less bit per brightness symbol than uniform quantization. That is, half as many brightness symbols are sufficient. PCM using 6 bits per brightness symbol and uniform quantization is often used as a standard against which other data-transmission and data-compression schemes are compared.

A method described by Roberts(14) makes it possible to improve the quality of pictures which might otherwise contain quantization noise (e.g., false contours) without increasing the bits per brightness symbol required. The method involves adding pseudorandom noise to the picture before quantization and subtracting the same noise at the receiver. He has found that 3 bits per sample would be sufficient for transmitting most television pictures.

Roberts defined an error criterion which could be used to judge quantitatively the quality of reproduced pictures. He assumed that the errors \( E \) in a reproduced picture consist of two components: the apparent noise in the output, \( V \), and the tonal differences between the input and the output, \( D \). Therefore \( E = V + D \). The total error is defined as the mean square-error.
where $x$ is the input signal and $y$ is the output. A flat distribution is assumed for the input, making $p(x)$ constant.

In this case $E$ becomes,

$$E = \mathcal{A} \int_0^1 \int_0^1 p(y|x)(x-y)^2 \, dy \, dx,$$

where $\mathcal{A}$ is defined so that the minimum mean square error is unity. The apparent noise is defined as the variance of the output relative to the mean output with the input held constant,

$$V = \mathcal{A} \int_0^1 \int_0^1 p(y|x)(y-\tilde{y}_x)^2 \, dy \, dx,$$

where $\tilde{y}_x = \int_0^1 p(y|x) \, dy$.

The tonal error is a measure of the deviation of the mean output from the input in a region of constant input, and is defined so that $E = V + D$,

$$D = \mathcal{A} \int_0^1 (x-\tilde{y}_x)^2 \, dx.$$

Now for straight PCM over a noiseless channel $E=1$, $D=1$, $V=0$ so the entire error is due to tonal error. If random noise with a flat distribution and an amplitude equal to one grayness level, $2^{-n}$, is added to $x$, then

$$E = 1 + (1-2^{-n}), \quad D=2^{-n}, \quad V=2(1-2^{-n}).$$
In this case the total error has been reduced at the expense of increased apparent noise. For \( n \geq 3 \) the reproduced pictures were free of false contours and \( D = 0 \) from a practical standpoint; also \( V = 2 \). If pseudorandom noise is used, the same noise which is added at the transmitter can be subtracted at the receiver. This has the effect of reducing \( V \) without increasing \( D \). Now \( E = 1 + 2^{-n} \), \( D = 2^{-n} \), \( V = 1 \).

2.1.5 Transmission Code

The most common code used in PCM transmission is a straight binary code. If a non-statistical code is to be used there is very little reason for using a code other than the straight binary code, unless noise on the channel is a serious problem. In the case of a noisy channel with specified error characteristics and specified picture brightness statistics, it is possible to choose from a set of codes a particular code which is optimum, in the sense that the average noise power in the reproduced picture is a minimum. (2)

Another possibility would be to use variable length statistical coding. To do this it is necessary to compile statistics concerning the distribution of brightness symbols for all pictures to be transmitted. Studies of television picture statistics indicate that the unconditional picture brightness distribution is uniform but that adjacent samples are highly correlated. (15) Schreiber (12,13) has determined
the conditional probability distribution up to the second order for television pictures and has reported that the conditional entropy, $H_y(x)$, ranged from 1.85 for the simplest picture to 3.36 for the most complex, with an average at 2.62 bits per sample. This is to be compared with standard 6-bits-per-sample PCM and implies that statistical coding could decrease the average number of bits per picture by a factor of 2.2. This is not a very significant decrease considering the complexity of the encoding and decoding equipment. He gives a figure for the second-order conditional entropy, $H_{yz}(x)$, of 1.49 bits for a simple picture, which is 0.36 bits lower than $H_y(x)$ for the same picture. This would indicate that the added advantage of second-order statistics is small when compared with the increased complexity of the equipment necessary.

In another paper, Schreiber (11) evaluates a lower bound for the average entropy per picture element. He postulates a portion of a picture of uniform brightness where the neighboring picture elements are so highly correlated that any element-to-element variation in brightness is due primarily to Gaussian noise, whose r.m.s. value is equal to one brightness quantization level. The entropy for this case is computed to be 1.12 bits per brightness symbol. Therefore, for this limiting situation, the maximum possible compression ratio is $6/1.12 \approx 5.3$ to 1 when compared
with 6-bit PCM. This implies the use of variable-length statistical codes of high order involving rather long coding and decoding tables. Such complexity to achieve a compression ratio of less than 5.3 to 1 is unlikely, especially since a compression ratio at least ten times this is really needed.

2.1.6 Postfilter Characteristics

The postfilter is generally the inverse of the pre-filter. It can also be used to eliminate "noise" whose spatial frequencies lie outside the range of the picture spatial frequencies.

2.1.7 Summary

The PCM techniques which have been described seem to be capable of achieving compression ratios as high as 5 to 1 when compared with standard 6-bit PCM. Roberts' method using the addition of pseudorandom noise appears to be a simple way to achieve a 2-to-1 compression ratio. Other schemes which achieve higher compression ratios generally involve statistical variable-length encoding. The complexity of statistical encoding and decoding and the nonuniformity of picture statistics would generally make such schemes unattractive in view of the low compression ratios yielded.
2.2 DIFFERENTIAL PULSE CODE MODULATION

Another widely studied transmission technique is differential pulse code modulation, abbreviated DPCM. The method consists of sending a quantized version of the difference between a function of the previous samples and the current sample. Let the i-th difference signal be \( d_i \); let \( F_{i-1} \) be the function of the previous samples; and let \( s_i \) be the i-th sample. Then DPCM takes on the form

\[
d_i = s_i - F_{i-1}
\]

where \( i \) is indexed over all sample points of a picture. It is assumed here that the picture has been raster-scanned and converted into a one-dimensional function of time.

2.2.1 Ordinary DPCM

For ordinary DPCM, \( F_{i-1} \) takes on the form

\[
F_{i-1} = \sum_{j=1}^{i-1} [d_j],
\]

where the brackets mean that \( d_j \) has been quantized. The DPCM system has the following diagram:
2.2.2 **Predictive DPCM**

Another type of DPCM is called predictive DPCM. In this case, \( F_{i-1} \) is a function which predicts the next sample's value on the basis of the value of previous samples. The quantized difference between the predicted and actual value is transmitted. The diagram follows:

![Diagram of Predictive DPCM](image)

Various types of prediction functions can be used. They can be classed as linear or polynomial predictors. A linear prediction is a weighted sum of previous values and can be a "previous value," "slope," "planar," or "circular" predictor. The simplest linear prediction is one that predicts the current sample \( s_i \) to be equal to the immediately preceding sample \( s_{i-1} \) or to the sample directly above it on the previous scan line; however, the latter would involve storing an entire scan line. Slope prediction is a bit more complicated, involving two previous values, while planar and circular prediction are even more complicated involving 3 and 7 previous values, respectively, from more than one scan line. The equations for the various
forms of linear prediction and the raster sample points used.
are shown below.\(^{(16)}\)

\[
\begin{align*}
  s_p &= s_{1,0} \text{ (previous value, same line)} \\
  s_p &= s_{0,1} \text{ (previous value, line above)} \\
  s_p &= 2s_{1,0} - s_{2,0} \text{ (slope)} \\
  s_p &= s_{1,0} + s_{0,1} - s_{1,1} \text{ (planar)} \\
  s_p &= s_{1,0} + s_{2,1} - s_{2,1} + s_{3,0} - s_{3,1} + s_{1,2} - s_{1,1} \text{ (circular)}
\end{align*}
\]

(x marks the current sample value to be predicted)

Each type of prediction can predict certain picture entities. For example, planar and circular prediction will predict horizontal and vertical lines; lines at angles of \(\pm 52^\circ\) are predicted by circular prediction. Harrison\(^{(16)}\) experimented with various types of linear prediction and found that the predictive DPCM signal power is lower than the ordinary PCM signal power for the same picture. In addition, the DPCM error signals still contained considerable redundancy because, when the error signal alone was displayed on a TV monitor, the original pictures were clearly distinguishable.

Linear prediction is a special case of polynomial prediction. In general, polynomial prediction involves previous values raised to some power and, therefore, is considered too complex for our purposes. In fact, results indicate that "previous value" prediction eliminates most of the
redundancy and little is gained by higher-order prediction. (22)

2.2.3 Advantages of DPCM

The real advantage of DPCM is that the unconditional entropy of the error signal (based on the quantized error signal monogram statistics) is approximately equal to the first-order conditional entropy of the actual signal. (12)

This means that the advantages of statistical coding can be obtained more easily by coding and transmitting the error signal. O'Neal (17) developed a theory of predictive DPCM which he used in conjunction with signal statistics to derive optimal prediction function coefficients. He found that a 3:2 savings in bit rate over straight PCM could be achieved. Graham (18, 21) incorporated two properties of human vision into his predictive DPCM system and was able to achieve a 2:1 reduction in bit rate using 8-level (3 bits) error signal quantization. The first property of human vision which he used is the inability to detect amplitude distortion in chaotic, complex regions of a picture. The second property is the sensitivity of the human eye to distortion in simple, low-contrast, predictable, structured regions. Thus where the error signal is large (unpredictable, chaotic region) a few coarse quantum levels are used whereas for small errors, fine accurate quantization is used.

Tests on four types of pictures gave entropies ranging from...
2.1 to 2.64 bits per picture element. Graham also mentions alternate-mode predictive DPCM, where one of two prediction functions is used at each step depending on which gives the smaller error. For an assumed joint distribution function of the form \( p(i,j) = K\alpha^{|i-j|} \) Oliver\(^{(19)} \) found that the entropy of a statistically coded "previous-value"-predicted error signal approached 2.918 bits per picture element.

2.2.4 **Disadvantage of DPCM**

The disadvantage of DPCM is that it is much more sensitive to noise than ordinary PCM. Since a reconstructed picture element depends on all past picture elements, a single error appears in all remaining picture elements.\(^{(17,20)} \) Because of the large amount of data transmitted for any picture, transmission errors could be a serious problem. Error recovery procedures in the transmission link might solve this problem.

2.2.5 **Summary**

In summary, it can be said that DPCM, like PCM, yields only very modest compression ratios; ratio of 2:1 appears to be the best that can be achieved. In addition, the picture degradation due to errors appears to be a serious drawback.

2.3 **PCM Using Frequency Separation**

Before going on to techniques which involve picture property detection, one other PCM technique should be mentioned. This consists of breaking down the scanned picture
signal into two or more component signals, each containing a particular band of frequencies. Then each component signal is sampled and quantized using a sampling rate and a number of quantization levels suitable to that band of frequencies. The sampling rate is generally equal to twice the highest frequency in the band. The criteria for determining the number of quantization levels for a given frequency band is based on the two properties of human vision mentioned previously. Bands containing the higher frequencies represent high-contrast, very detailed portions of the picture; therefore brightness amplitude can be coarsely quantized due to the eyes' insensitivity to brightness at such frequencies. Low-contrast regions of the picture are represented by the lower-frequency bands, to which the eyes' brightness sensitivity is high; consequently many quantization levels are needed. Kretzmer took a 4MC television signal and passed it through a low-pass filter to get the "lows" component signal (0-0.5MC) which he sampled and quantized with 7 bits. The "lows" signal was subtracted from the original video to get the "highs" signal (0.5-4MC) which was quantized with 3 bits. The result was approximately a 2:1 compression ratio compared with straight 7-bit PCM. He also proposed another scheme using 7 bits for 0-0.5 MC, 4 bits for 0.5-1 MC, 3 bits for 1-2 MC and 2 bits for 2-4 MC. Results are not reported but the gain should have been less.
2.4 PROPERTY DETECTION AND ENCODING

The next few transmission techniques to be described are based primarily on the definition, detection, and encoding of certain picture details or properties.

2.4.1 Run-Length or Differential-Coordinate Encoding

The first of these is run-length encoding (also called differential-coordinate encoding). In this method, the picture property which is detected and encoded is a string of picture brightness samples which can all be assigned the same quantization level. The length of such a string or "run" is what is actually coded and transmitted. Some information about the quantization level must also be sent.

Michel (24,25) has reported some interesting results on run-length encoded black and white material (no intermediate grayness levels). The method involves transmitting code words which represent contiguous sequences of all black or all white sample points. Since the black or white run-lengths could be very large values (as in the case of an empty page) it is necessary to restrict the run-length values for which there are codes. This limits the length (in bits) and number of code words. Run lengths which exceed the maximum length for which there is a code word must be expressed in terms of two or more code words in sequence. Michel computed statistics concerning run lengths contained in 300-word pica typewritten copy. He used enlargements of the pica
letters to measure run lengths and then applied known results for the monogram, diagram, and trigram frequencies for English letters\(^{(25)}\) to compute the run-length probabilities. Using Shannon-Fano encoding as modified by Huffman\(^{(27)}\), Michel derived code words based on the derived probability distribution and found that the average code-word length (in bits) was within 1% of the entropy (in bits) of the run lengths. Correlation between run lengths was not taken into account. Michel's code had a prefix which preceded a run-length code word if it was a black run, otherwise all run lengths were assumed to be white, and each white run was assumed to terminate with one black dot. This pica-tailored code was used to encode a variety of data. Figure 1 gives some of his results.

The compression ratio for the 300-word pica letter is \(R = \frac{10^{6}}{10^{5}} = 10\). This value seems low since the code is tailored to be optimum for this case. For the empty page \(R=180\). However, intuitively, 5,500 bits seems like an excessive amount of information to be associated with an empty page. The reason for this is that Michel felt that the average transmission which he would encounter would involve pica-style text and should therefore be tailored for that case, at the expense of efficiency in other cases. It is difficult to compare these compression ratios against a norm because it is difficult to measure the amount of information contained in a picture. In fact, as stated by Pearson\(^{(28)}\),
Table I. Portion of Pico-Tailored Code

<table>
<thead>
<tr>
<th>Item</th>
<th>Code Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>1</td>
</tr>
<tr>
<td>2 dots</td>
<td>0010</td>
</tr>
<tr>
<td>3 dots</td>
<td>011</td>
</tr>
<tr>
<td>4 dots</td>
<td>110</td>
</tr>
<tr>
<td>5 dots</td>
<td>111</td>
</tr>
<tr>
<td>Margin (also empty line)</td>
<td>1100</td>
</tr>
<tr>
<td>7 dots</td>
<td>1000</td>
</tr>
<tr>
<td>8 dots</td>
<td>1110</td>
</tr>
</tbody>
</table>

Table II. Binary Digits per 8'/ by 11-inch Page

<table>
<thead>
<tr>
<th>Type of Copy</th>
<th>Binary Digits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty page</td>
<td>2,500</td>
</tr>
<tr>
<td>Simple drawing, Fig. 4101</td>
<td>88,900</td>
</tr>
<tr>
<td>Grid with 2-inch spacing</td>
<td>40,900</td>
</tr>
<tr>
<td>300-mesh graph paper</td>
<td>300,900</td>
</tr>
<tr>
<td>Circuit schematic, Fig. 2/11</td>
<td>129,400</td>
</tr>
<tr>
<td>Page filled with text type</td>
<td>1040,400</td>
</tr>
<tr>
<td>Grid with 1/4-inch spacing</td>
<td>920,000</td>
</tr>
<tr>
<td>Page filled with 8-point type</td>
<td>105,100</td>
</tr>
<tr>
<td>Two forth and one white dot alternating horizontally</td>
<td>3,880,990</td>
</tr>
</tbody>
</table>

Figure 1. Some Results of Michel's Run-Length Encoding (34)
the actual information present for transmission is a function of the process used to digitize the original picture. The higher the resolution used in the scanning process, the larger the amount of information present, up to a point. At even higher resolution the redundancy increases and the amount of information becomes constant (see Schreiber(11)). The results for the 300-word pica letter could be evaluated on the basis of Shannon's(29) results concerning the redundancy of English. The zeroth-order measure of average information per symbol (letters and spaces) in English text gives 4.76 bits per symbol. Assuming that the average English word has five letters and a space, then 300 words would contain an average amount of information equal to 8,568 bits. Some bits must be added to this to designate the format of the page such as margins, line spacing, paragraph indentation, etc. Let the total be 10,000 bits, then \( R' = \frac{10^6}{10^5} = 100 \), in contrast to Michel's value of \( R = \frac{10^6}{10^5} = 10 \) for the 300-word pica letter. If higher-order measures are used for English, such as 4.2 and 3.6 bits per symbol for the first- and second-order or the eight-order value of 2.4 bits per symbol, the difference between \( R \) and \( R' \) approaches 200 versus 10.

2.4.2 Restricted Run-Length Encoding

Another approach to run-length encoding was taken by Cherry, et al. (30, 31) They analyzed the run-length statistics for some class of documents and attempted to find
an optimum set of standard run-lengths. Then any run-length was restricted to be one of the standard run-lengths. Long run-lengths were, in effect, broken down into a set of standard run-lengths. They were considering a situation in which the scanning, encoding, and transmission were all proceeding simultaneously. In this case, extremely long run-lengths could cause gaps in the transmission while the end of the run was being awaited; also transmission buffer overflow became possible when many very short runs occurred in succession. They performed a queueing theoretic analysis for the purpose of choosing standard run lengths, transmission rate, and buffer size so that the probabilities of buffer underflow and overflow were below a certain level.

The problem of transmission buffer underflow or overflow does not arise when the pictorial data are compressed and encoded off-line.

One of the real advantages of using standard run-lengths is that variable-length statistical coding becomes unnecessary. If the run-lengths are chosen properly they are uniformly distributed and can therefore be efficiently represented by a fixed-length code. The longest standard run places an upper limit on the compression ratio. This seems to be a disadvantage to using restricted run-length encoding. Robinson and Cherry(31) used standard run-lengths of 1,2,4,10 sample points and 7-bit brightness quantization.
Using various test pictures, they achieved compression ratios in the range of 3 to 6. Each run was specified by 9 bits: 7 bits for brightness quantization and 2 bits for run-length. If the entire picture consisted of runs equal to the maximum run-length of 10, then the compression ratio would be
\[
\frac{nx^7}{(m/10)x^9} = 7.8,
\]
the maximum compression ratio possible using 1, 2, 4, 10 as the standard run-lengths. The authors found that slight variations in the first three values did not significantly affect the results.

Run-lengths are found to have an exponential distribution function \((30,32)\). The probability that a run is of length \(n\) is \(P(n)=Ar^{-n}\), and since \(\sum_{n=0}^{\infty} P(n)=1\), then \(A=r-1\) for \(r>1\). Therefore, \(P(n)=(r-1)r^{-n}\); and the average value of \(n, \bar{n} = \sum_{n=0}^{\infty} nP(n) = r/r-1\). Now if a sample is sent only at the beginning of each run, neglecting run-length information, then the rate of sample generation will be decreased by the factor \(r/r-1\), which is an upper bound on the compression ratio. The added bits to specify run-length make this bound impossible to achieve. Seyler \((32)\) found that a run-length probability density function of the form \(p(T)=ae^{-\alpha T}\) was accurate for statistically stationary television signals having an exponential element auto-correlation function. His analysis led to the conclusion that a worst-case compression ratio of 5 was possible. He also looked at the
problem of choosing the maximum run-length value such that the bit rate would be a minimum, for various values of \( \alpha \). Wyle et al. measured run-length statistics and devised a coding scheme which makes it possible to begin reproducing a run before the entire run-length has been received. They report an average compression ratio of 4 for black and white pictures and 2.9 for continuous-tone pictures. The maximum ratio was 8.5. Golomb looked at the problem of finding an optimum variable-length statistical code when the run-length distribution function is of a particular form. He found a general form for the code corresponding to the general geometric distribution \( p^n q \). His code can be encoded and decoded using a brief algorithm rather than an infinite code book. Kubba and Sekey have studied the problem of optimum detail detection for the determination of run lengths in noisy pictures. They use the theory of statistical inferential estimation to find \( \log (H_n / H_{n+1}) \) which indicates the preference for \( H_n \) versus \( H_{n+1} \) where, \( H_n \) is the hypothesis: \( x_n \) and \( x_{n+1} \) are independent so \( x_{n+1} \) starts a new run; and \( H_{n+1} \) is the hypothesis: \( x_{n+1} \) is part of a run already in progress. \( x_n \) and \( x_{n+1} \) are neighboring picture brightness samples.

In summary, run-length encoding seems to be capable of providing compression ratios between 5 and 10, with some very simple pictures giving ratios as high as 20 (see Michel's "Nature Morte," Fig.1). Variable-length statistical coding
seems necessary to achieve ratios as high as 10, but the standard runs method results in ratios not too much less and is more easily implemented.

Another method used to detect changes in picture brightness, and the position of those changes, is described by Gouriet (37). His method involves converting the analog picture signal to a staircase-like quantized signal which is then differentiated, giving a signal consisting of positive and negative pulses corresponding to positions in the quantized signal, where there are positive or negative steps, respectively. These "position pulses" are then rectified and used to drive a ramp function generator. Each pulse resets the generator to zero output, so that the composite output is a "sawtooth" where each "tooth" begins and ends at times corresponding to "position pulses," and the amplitude of each "tooth" is proportional to the distance between the pulses. Now the "sawtooth" signal is differentiated, yielding pulses whose position has no significance but whose amplitude is a measure of the distance between successive changes in brightness. This stream of "distance pulses" could then be transmitted at a regular rate over an analog channel. Another signal must also be sent consisting of brightness samples corresponding to points in the video signal where the changes in brightness occur. The "position pulse" stream could be used to derive the
brightness samples. Gouriet found that the combined entropy of the distance and brightness signals was smaller, by a factor of seven, than the picture entropy in the case where picture elements in a scan line are assumed to be independent.

Newell and Geddes\(^{(38)}\), in testing Gouriet's technique, found that for a raster having 180 lines and 500 elements per scan line, the average frame had 7000 brightness changes with the maximum being 12,000; most scan lines had more than 50. These results were based on 6-bit brightness quantization. When 13 nonuniformly spaced grayness levels were used, the average was 12,000 brightness changes. In both cases 12 pictures were tested. An unrestricted distance signal was found to require an extremely high signal-to-noise ratio, and it was necessary to restrict the distance signal to a maximum of 20 picture elements between brightness changes. They concluded that a practical bandwidth reduction factor of 3.1 seemed possible.

2.4.3 Dual Mode and Edge Detection Systems

Schreiber and Knapp\(^{(39)}\) experimented with a dual mode system which transmitted a low frequency or "lows" signal using ordinary PCM. The "highs" signal was run-length encoded using 3 bits for brightness quantization and 5 bits for run length. The compression ratio achieved was estimated to be 4:1. Newell and Geddes\(^{(38)}\) tested a similar system
using Gouriet's analog technique for run-length transmission and found that a bandwidth reduction of 3:1 was practical. They also tested the "synthetic highs" technique of Schreiber et al. This method is based on the detection of brightness "edges" in a scan line. The "lows" signal is separated using a low-pass filter and is transmitted separately using PCM or reduced bandwidth analog transmission. The highs signal is derived by differentiating the original signal and, by means of very few quantization levels, replacing the derived signal by pulses whose height and duration are determined by the portion of the signal within a given quantization level. The minimum quantization level is chosen so that the low-frequency contribution to the derivative is ignored. These pulses are then transmitted in analog fashion or are run-length encoded by transmitting the distance between pulses and the discrete pulse height. At the receiving end the pulses are passed through a specially designed transversal filter which synthesizes a "highs" signal which is recombined with the "lows" to give the reproduced picture (see Fig.2). The "synthetic highs" technique gave qualitatively better pictures than the quantized highs method with generally the same noise and bandwidth requirements.

Julesz has investigated a technique which combines a type of run-length encoding at the transmitter with linear interpolation at the receiver. The beginning of a run is
Figure 2. Waveforms of the Synthetic-Highs System (38)

(a) Video input.
(b) 'Lows'.
(c) Derivative of video input.
(d) Quantized derivative.
(e) Brightness pulses for 'highs' after decoding at receiver.
(f) 'Synthetic highs'.
(g) Final waveform, (b) + (f).
based on the detection of an edge or abrupt change in brightness on a scan line. These edges are detected by comparing the differences in brightness of adjacent pairs of unquantized brightness samples using the differences \( u_{i-1} - u_{i-2} \) and \( u_i - u_{i-1} \), where \( u_i \) is a brightness sample at the \( i \)-th sample point and \( u_{i-1}, u_{i-2} \) are the two previous samples on one side of \( u_i \). An \( \epsilon \) is used to assign an index to difference functions in the following way: if \( u_{i-1} - u_{i-2} > \epsilon_1 \), the index equals \(+1\); if \( |u_{i-1} - u_{i-2}| < \epsilon_1 \) the index equals \(0\); if \( u_{i-1} - u_{i-2} < -\epsilon_1 \), the index equals \(-1\). This leads to the following types of brightness changes and associated index pairs: \( (0,1), (1,0), (1,1), (-1,0), (-1,-1), (1,-1), (1,1) \). If any of these changes should occur, except for the \((1,1)\) or \((-1,-1)\) case, a new run is begun and the old run is terminated, quantized, and sent along with a quantized version of \( u_i \) at that point. For practical considerations, 16 Nyquist intervals was the maximum allowed run-length, so 4 bits were required. Ten bits were used for brightness quantization (6 bits would have been sufficient) except when the change was within \( \pm 2 \) Nyquist intervals of the last change, then coarser 4-bit quantization was used. This takes advantage of the eye's insensitivity to brightness accuracy in a region of rapid change, a property of human vision mentioned earlier at the receiver, linear interpolation is used
to reconstruct the signal across a run-length between two edges.

There were some problems with this scheme. If $\epsilon_1$ was too small, misalignment of edges from one scan line to the next resulted. If $\epsilon_1$ was too large, detail was lost. Therefore an $\epsilon_1$ (small) and an $\epsilon_2$ (large) were chosen, and the union of the two sets of edge points was used. Another problem was that for gradual changes in brightness an error would build-up over long intervals, because only adjacent differences are compared. This was solved by monitoring the difference between the original and coded picture and taking extra samples if the difference exceeded a tolerance $\epsilon_3$ (and the distance to the next edge was greater than 2 Nyquist intervals). Two sets of criteria tested were $\epsilon_1=3.6\%, \epsilon_2=10\%, \epsilon_3=5\%$ and $\epsilon_1=5\%, \epsilon_2=10\%, \epsilon_3=7.2\%$. Four basically different test scenes were used. The minimum average ratio of edge points to total raster points was 31.5%, with 42.4% being the maximum and 25.3% being the minimum for one scene. Fifty-six per cent, 72%, and 44% are the average, maximum, and minimum ratios, respectively, of coarsely quantized samples to selected samples. Statistics were also computed for run lengths between edges. For one of the scenes, less than 10% of the run lengths exceeded 16 Nyquist intervals and, therefore, required additional samples. These statistics were used to compute the entropy in bits per sample in order to estimate the results of
statistical encoding. The average, maximum, and minimum entropy was 2.35, 2.80, and 2.01, respectively, giving an average compression ratio of 2.9 compared to 7-bit PCM. The statistical encoding in this case takes into account only the run-length statistics and not the brightness-level statistics. A more practical fixed-length non-statistical code would give an average compression ratio of 2.4.

Julesz found that if any of the scenes were coded with a statistical code tailored for any other scene the results were only slightly different from the results when a scene was coded with its own code. This is important because in a practical system the same code would have to be used for all scenes. It should be noted that Julesz used $1/25$ of the normal resolution, and therefore much higher compression ratios could be expected with his scheme using finer resolution. However, the restriction on his run lengths to 16 Nyquist intervals limits his maximum compression ratio to 9.6 for non-statistical encoding.

Gabor and Hill(42) have described a method which involves detecting edges of brightness on alternate scan lines and, using linear interpolation, derives the edges for the scan line in between. This could result in sending half as many edge samples but seems to call for a much more complex receiver.
2.4.4 Contour Detection and Encoding

Consider the following assumptions: (1) the human eye responds most to edges of brightness and, in fact, emphasizes them; (2) sharp edges of brightness are relatively few in an average picture and they lie along connected contours; (3) less accurate reproduction of the low-frequency, low-contrast portions of a picture will not be noticeable. These assumptions were used by Pan (43) and Graham (44) as the basis for a dual mode picture transmission technique which is a two-dimensional extension of the edge-detection method of Schreiber, Knapp, and Kay (40). A two-dimensional low-pass filter is used to form an image resembling an out-of-focus version of the original picture which is then transmitted using ordinary 6-bit PCM and a much coarser sampling rate. In order to detect the edges of brightness the original picture is operated upon using a function which will emphasize abrupt changes in two dimensions. Either the gradient or the Laplacian operator can be used for this purpose. A nonlinear thresholding operation is performed on the resulting image to detect the presence of contours. Once a potential contour point is detected, neighboring points are examined in order to trace out the entire contour. Extremely short contours are rejected and are assumed to be the result of noise in the original picture. Each contour found is encoded and
transmitted to the receiver. To reconstruct the original picture at the receiver, a two-dimensional transversal filter is applied to the image containing the thin line contours, synthesizing a "highs" signal, which, when combined with the "lows" signal, gives the desired result. The characteristics of this transversal filter are a function of the low-pass filter and the edge-detection operator. This becomes apparent in the one-dimensional case where the edge is the step function shown in Fig. 3; the low-pass filter gives 3(b) and the gradient gives 3(c). Therefore, the transversal filter's response to 3(c) is constrained to be something like 3(d) in order that 3(b) and 3(d) will sum to 3(a).

![Figure 3. A one-dimensional example of the synthetic "highs" system for step function input](image)

Graham performed a computer simulation of this technique using a 256x256-element picture representation and 8-bit brightness quantization. A two-dimensional low-pass filter in difference equation form was used to determine the "lows" image, which was transmitted using 6-bit PCM and a 32x32 sampling matrix giving 3000-6000 bits. Both the gradient
and Laplacian methods were used for edge detection, and
transversal filters were derived for both cases. For one
of the test pictures the gradient yielded 11,886 edge points
while the Laplacian gave 26,119; the same threshold was
used in both cases. Results for three other test pictures
using the gradient are summarized in Table 1. The beginning
of a contour was coded as follows:

Table I

Contour Coding Data for Figures 16, 17, and 18

<table>
<thead>
<tr>
<th></th>
<th>Fig. 16</th>
<th>Fig. 17</th>
<th>Fig. 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of edge points</td>
<td>8900</td>
<td>13,980</td>
<td>2301</td>
</tr>
<tr>
<td>Number of contours</td>
<td>263</td>
<td>485</td>
<td>55</td>
</tr>
<tr>
<td>Number of bits for:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VDC</td>
<td>20,300</td>
<td>38,400</td>
<td>5,900</td>
</tr>
<tr>
<td>GDC</td>
<td>17,400</td>
<td>30,600</td>
<td>4,260</td>
</tr>
<tr>
<td>GMC</td>
<td>11,200</td>
<td>18,900</td>
<td>2,880</td>
</tr>
<tr>
<td>start point</td>
<td>6,312</td>
<td>11,640</td>
<td>1,320</td>
</tr>
<tr>
<td>lows</td>
<td>3,300</td>
<td>5,000</td>
<td>3,000</td>
</tr>
<tr>
<td>Total number of bits</td>
<td>58,212</td>
<td>102,540</td>
<td>17,360</td>
</tr>
<tr>
<td>Reduction factor</td>
<td>6.75</td>
<td>3.8</td>
<td>22.7</td>
</tr>
</tbody>
</table>

16 bits were used to designate a contour starting location,
3 bits designate the direction of the gradient, 2 bits are
for the gradient magnitude, and 3 bits designate the contour direction, making 24 bits total. A variable-length Huffman code is used to designate the following elements of contour continuation: contour vector direction change (VDC) which is integral multiples of 45°, -3° ≤ VDC ≤ 3°; gradient direction change (GDC) which has the same values as VDC; and gradient magnitude change (GMC) which has values 0 ≤ GMC ≤ 4. The end-of-contour symbol is an eighth value for VDC. The four test pictures required, on the average, 5.8 bits per contour-continuation point. A fixed-length non-statistical code would have required 9 bits.

The compression ratios achieved by contour encoding (see Table I) were generally higher than those which would be expected if previously described methods were applied to the same pictures. In addition, this method is not inherently limited in its maximum compression ratio as is restricted run-length encoding. Therefore, increased resolution should give a larger proportional increase in compression ratio. If the 256x256 sampling matrix were replaced by a 512x512 matrix, the number of samples would increase by a factor of 4, but the number of edge points would be expected only to double; therefore the compression ratios would be expected to at least double. There seem to be two disadvantages associated with this scheme. One is the amount of processing time required for each picture. Simulation processing times for a single picture
were on the order of 5 minutes, using an IBM 7094. The
digitized pictures were recorded on digital magnetic tape.
Optical filtering might help but would still leave the pro-
cessing time required for contour detection, tracing, and
encoding, and perhaps contour reconstruction. The other
disadvantage is the sensitivity to transmission errors.
Since a contour continuation message contains information
only about changes in contour properties, an error in one
contour continuation symbol would cause the remaining por-
tion of that contour to be reconstructed improperly.

Cheydleur (46) has proposed a scheme for the coding
of contours and lines of varying width. It is based on
the use of compactly coded syllables which specify such
things as contour location and contour alteration. A con-
tour location syllable specifies the beginning of a new con-
tour. It is followed by sub-syllables which specify its
horizontal displacement relative to another contour, the
slope of the left contour boundary, the rate of change of
the slope, the width of the contour, and the rate of change
of the width. There is also a sub-syllable which specifies
the vertical displacement to the position where there is
a change in any of the previously mentioned sub-syllables.
The values of the changes are specified by the sub-syllables
of a contour alteration syllable. Pictures with grayness
levels are coded by replacing the width sub-syllable by one
that specifies grayness value to the left of the contour.
As yet, no experiments have been performed to evaluate this technique.

2.4.5 Area Encoding

A method described by Cunningham\(^{(47)}\) involves replacing every \(n \times n\) area of the picture by the average brightness in that area. The intermediate values are determined at the receiver by means of interpolation. In addition, at the transmitter, the interpolated values are compared to the true values, and a pair of criteria are used to determine whether special correction symbols need to be sent. Averaging over a 3x3 area gave good quality pictures and a compression ratio of 6:1; 5x5 averaging gave poor quality pictures and a compression ratio of around 9:1. A modified correction scheme applied to 5x5 averaging brought the quality up and the ratio down to that of 3x3 averaging.

An encoding technique based on area-dependent statistics was investigated by Wholey.\(^{(48)}\) He selected the 12-element pattern shown in Fig. 4 and analyzed several pictures of a given class (weather maps) for the statistical relationship between the value of \(P\) and the various combinations of values for the 12 x's.

\[
\begin{align*}
&x_1 \ x_2 \ x_3 \ x_4 \ x_5 \\
&x_6 \ x_7 \ x_8 \ x_9 \ x_{10} \\
&x_{11} \ x_{12}^P
\end{align*}
\]

Fig. 4 12-element statistical encoding pattern.
Only black and white pictures were considered. These statistics were used to predict the most probable value for \( P \), given value assignments for the 12 x's. A particular picture is processed by scanning the pattern across it and making a prediction for \( P \) at every sample point. The picture is assumed to have a white border two elements wide. For each sample point at which a wrong prediction is made, a 1 is placed in the corresponding position in an error matrix. The resulting error matrix is a complete representation of the picture, which is then transmitted using statistical run-length encoding. For ten 7000-element weather maps the average error matrix had 5.5% 1's and the average compression ratio was 2.6:1. The entropy of the average error matrix was

\[
H = -[0.055 \log_2 0.055 + 0.945 \log_2 0.945] = 0.31,
\]

which gives a compression ratio of 3.2:1 ideally. When various picture processing techniques were applied to the pictures, which gave a straight-line approximation to the weather maps, the compression ratios increased to about 7.7:1.

In order to achieve large compression ratios with area encoding it is necessary that the areas themselves be large and also that the number of possible combinations of brightness within the areas be small or highly non-uniform in their probability distribution. These properties seem
self-contradictory and do not agree with statistics which have been compiled.(49)

2.5 COMPRESSION TECHNIQUES USING PICTURE TRANSFORMATION

Previous methods which have been described generally involve operations which attempt to extract and encode the essential information by operating on the original picture. It would be desirable to find a suitable transformation which, when applied to the original picture, yields an image containing the essential information in a more convenient form. A suitable transformation would preserve the information content in the original picture and would be reversible. Andrews and Pratt(50) have performed experiments in television bandwidth reduction using the finite two-dimensional Fourier transformation. This transform is a linear invertible one-to-one mapping. In addition, Andrews(51) proved that the Fourier transform is an information-preserving mapping. In general, the entropy of a function and its linear inverse are identical.

Andrews and Pratt used an original image containing 256x256 elements quantized to 64 levels of grayness. A digital computer was used to calculate the two-dimensional Fourier transform of the original image, using a highly efficient version of the Cooley-Tukey algorithm.(52) The Fourier image is then processed and transmitted. At the receiver, a second two-dimensional Fourier transform is performed to
obtain the original image. They have found that the double Fourier transform of a picture does not significantly degrade the quality of the reproduced picture. In addition, the Fourier image has the following interesting properties: most of the "information" lies along the coordinate axes and near the origin at low spatial frequencies; and the samples exhibit a greater degree of statistical regularity than do image domain samples. The authors used two methods to decrease the amount of information transmitted for the Fourier image. One method used a binary mask to blank out certain low-amplitude spatial frequencies. A second, more sophisticated method, is based on the fact that the energy contained in an image is the same in the spatial and Fourier domains. Therefore, the lowest spatial frequencies are generated and sent first, and as more spatial frequencies are sent the cumulative energy is computed. When the energy reaches the total energy, within some tolerance, transmission is terminated. The resulting number of Fourier image samples transmitted is a small fraction of the total. Gaussian 64-level quantization was used because linear quantization was found to be inadequate.

A reproduced picture with a compression ratio of 4:1 was very much like the original. Another picture compressed by a factor of 36 was quite blurred and unsatisfactory. Future work with this technique will be directed toward the
application of spatial domain data-compression techniques, like predictive and interpolative coding, to the Fourier domain.
REFERENCES


20. O'Neal, J.B., "Deltamodulation Quantizing Noise Analytical and Computer Simulation Results for Gaussian and TV Input


November 1963, pp. 1507-1517.


A SURVEY OF PICTORIAL DATA-COMPRESSION TECHNIQUES

Technical Report 16

J. DIGIUSEPPE

March, 1969

DA-49-083 OSA 3050

Technical Report 16

Qualified requesters may obtain copies of this report from DDC

Advanced Research Projects Agency

The results of a survey of pictorial data-compression techniques are summarized in this report. The survey was motivated by a study of half-time graphics communication over voice grade lines. The principal compression techniques surveyed include the following: the optimization of the parameters of ordinary pulse coded modulation, pulse coded modulation using added pseudo-random noise, differential pulse coded modulation, predictive differential pulse coded modulation, run-length encoding, brightness edge detection and the use of "synthetic highs," brightness contour detection and encoding, area encoding, picture compression in the Fourier domain.
picture, compression, compression ratio, encoding, statistical, transmission, quantization, entropy, run-length, contour, Fourier,