NASA TECHNICAL NOTE

NASA TN D-3475

NASA TN D-3475

COMPRESSION OF VIDEO DATA BY ADAPTIVE NONLINEAR PREDICTION

by Joseph A. Sciulli

Goddard Space Flight Center

Greenbelt, Md.

COMPRESSION OF VIDEO DATA BY ADAPTIVE

NONLINEAR PREDICTION

By Joseph A. Sciulli

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

ABSTRACT

Both a survey of the theory of adaptive data prediction
and a description of the computer simulation of the data
compression mechanism are presented. Results of simula-
tions of the conditional expectation predictor allow compar-
isons with other techniques. Also included are comments
on the problem of coding for a data compression system,
the characteristics of the Tiros TV cloud cover pictures as
an information source, and possible applications for data
compression systems.

CONTENTS

COMPRESSION OF VIDEO DATA BY ADAPTIVE
NONLINEAR PREDICTION

by
Joseph A. Sciulli
Goddard Space Flight Center

INTRODUCTION

In recent years studies of data compression have warranted the attention of many investigators. Since demands for large amounts of scientific data are increasing, methods for more efficient data transmission must be developed. Usually a communications system is designed so that the information source is sampled at a constant rate determined by the most active data periods. During a large percentage of time the data are relatively quiescent, and so redundant samples are transmitted. Data compression by prediction is a promising method of redundancy removal and is therefore the subject of many recent studies. A survey of the literature shows that two philosophies are being proffered for the solution to this problem. The first approach might be called the "state of the art" point of view where efforts have been focused on studying well-known, easily implemented techniques such as the zero-order and first-order predictors. References 1 and 2 are typical examples of this point of view. The philosophy of this approach is that the simpler schemes are within "state of the art" spacecraft instrumentation capability and are certainly easier to analyze and simulate. Those who have chosen this route generally feel that more sophisticated approaches are too complex to have any application value.

The second school of thought has chosen a more sound theoretical foundation in offering a solution to the data compression problem; the work of Balakrishnan (Reference 3) represents this latter philosophy. It is true that this approach at the present time appears to be difficult to instrument for spacecraft use; nonetheless, it is highly desirable to concentrate on the more sophisticated methods, especially since a high degree of onboard data processing capability (e.g., random access memory and arithmetic capability) will be available in the future.

This report is intended to develop part of the work reported in Reference 3 as well as to complement the work reported in Reference 4. It deals with the description, simulation, and analysis of the results of the application of the conditional expectation predictor to the compression of video data and presents observations on alternate methods and possible applications of the findings. Suggestions for future study are given in the concluding section.

THEORY OF ADAPTIVE PREDICTION SYSTEM

Before describing the prediction mechanism it would be worthwhile to formulate a working definition of an adaptive system. The word "adaptive" implies modification to meet new conditions. A truly adaptive system is characterized by its ability to (1) monitor its own performance with respect to some performance criteria, (2) learn of new conditions, and (3) adjust its structure to fit the new conditions. In a real communications system no *a priori* knowledge of the statistical structure of the information source is usually available. The data compression technique to be described in this report satisfies the definition of adaptivity and also requires no *a priori* statistical knowledge of the data.

Consider a sequence of descrete samples of the form shown in Figure 1. Assume that each sample may be any one of Q discrete values. Suppose a random variable X is defined such that

$$X = \left(x_{i-M}, x_{i-M+1}, \cdots x_{i-1}\right) . \tag{1}$$

The sample space size of the random vector X depends on the choice of the memory size M. Since each sample may assume any one of exactly Q discrete values, the sample space size of X for a memory size M is simply

$$S = Q^M . \tag{2}$$

Suppose, in addition, a second random variable χ is defined such that

$$\chi = x_j \qquad \text{for} \qquad 1 \leq j \leq Q \tag{3}$$

and corresponds to the data sample immediately succeeding X. Assume that we have been observing and recording the immediate successors to the random variable X over a number of samples denoted by L, the learning period, and that our operation must determine the optimal prediction for the ith sample. The optimal prediction $\hat{x}_i$ is given by

$$\hat{x}_i = E\left[\chi/X = \left(x_{i-M}, x_{i-M+1}, \cdots x_{i-1}\right)\right] ; \tag{4}$$

In the case of discrete data, $\hat{x}_i$ is given simply by

$$\hat{x}_i = \sum_{j=1}^{Q} x_j \, P\left[\chi = x_j/X = \left(x_{i-M}, x_{i-M+1}, \cdots x_{i-1}\right)\right] \tag{5}$$

where x_j is a possible successor to X and $P\left[\chi = x_j/X = \left(x_{i-M}, x_{i-M+1}, \cdots x_{i-1}\right)\right]$ is the probability that $\chi = x_j$ given $X = \left(x_{i-M}, x_{i-M+1}, \cdots x_{i-1}\right)$. If the data are assumed to be a long sample from an ergodic process, equation 5 represents the "best" RMS predictor, since the mean is that point about which the second moment is minimized.

$$X_1, X_2, \ldots X_{i-L-1}, X_{i-L}, X_{i-L+1}, \ldots X_{i-M-1}, X_{i-M}, X_{i-M+1}, \ldots X_{i-1}, X_i, X_{i+1}, \ldots X_n$$

Memory
M

Learning period
L

Past data ← → Future data
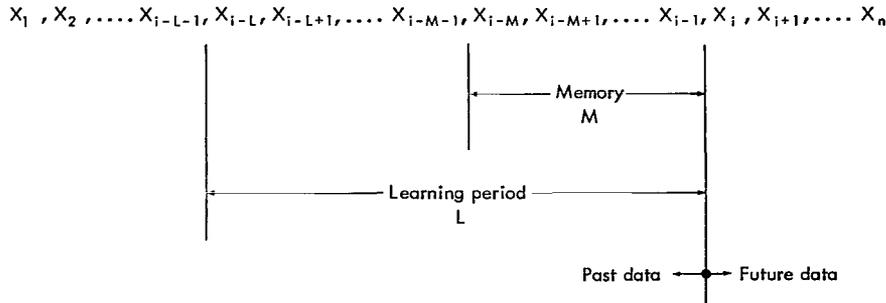
Figure 1—Sequence of discrete data samples.

To illustrate by example, assume that k observations of a particular X are made, and at each observation the value of the immediate successor to X is recorded. Suppose a prediction is required for the immediate successor to the $(k+1)$st observation of this particular X. According to Equation 5 the optimal prediction is the mean of the sample of past successors to X and is given by

$$\hat{x}_{k+1} = \sum_{j=1}^{Q} x_j \frac{k_j}{k}, \qquad (6)$$

where x_j is a possible successor to X, $1 \leq j \leq Q$, and k_j is the number of times x_j was observed.

Actually one could choose a statistic other than the mean, and correspondingly minimize some prediction error criterion other than the mean square error. The mode, for example, could be used as the prediction for the immediate successor to the random variable X. Utilizing the mode as the predictor minimizes the probability of error. In order to implement the mode predictor, histograms representing the distribution of the immediate successor to each particular X in the sample space are constructed. The most frequent successor then becomes the prediction.* One could also choose the median as the prediction; choice of the median minimizes the absolute error. It is interesting to note that if the data were both Gaussian and stationary then the mode, median, and mean would produce identical prediction results.

COMPUTER SIMULATION OF CONDITIONAL EXPECTATION PREDICTOR

The results of this work were obtained from simulations on the IBM 7094 computer using Tiros TV cloud-cover picture data as the information source. Reference 4 contains a good deal of background information on these data, including their origin and subsequent formating for computer simulation. A Tiros TV picture is nominally a 500-scan-line picture with each line composed

---

*This technique was implemented by Davisson of Princeton during his participation in the 1965 Goddard Summer Workshop (Reference 5). His results in some cases were somewhat better than those using the mean as the prediction.

3

(a)
Q=16, M=1, S=16

(b)
Q=16, M=2, S=256

(c)
Q=16, M=3, S=4096
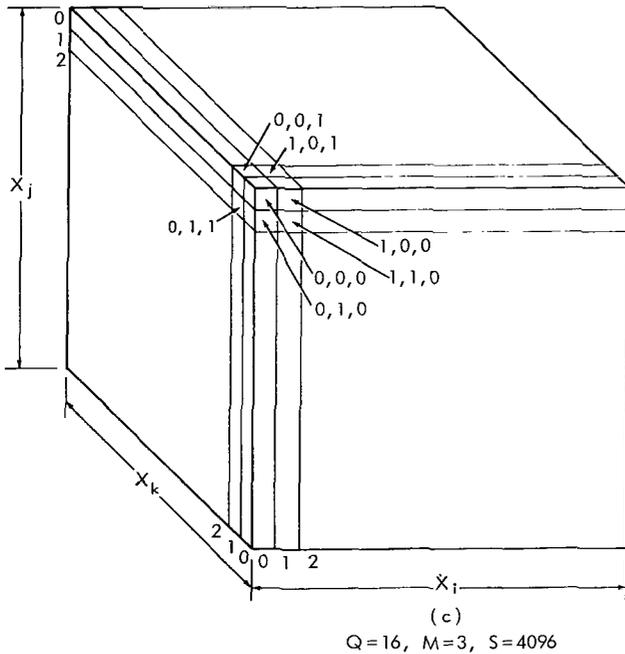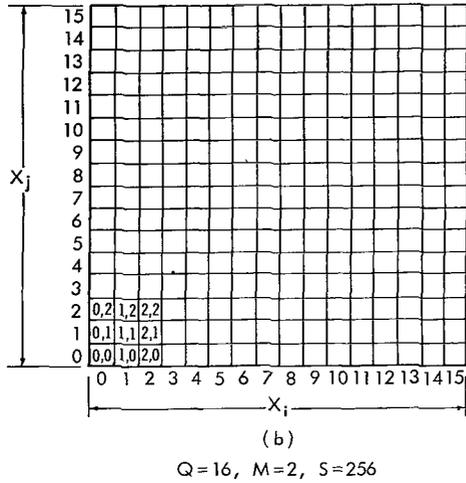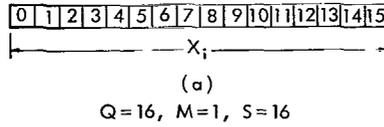
Figure 2—Memory cell geometries for Q = 16 and M = 1, 2, and 3.

of 500 TV picture elements. This study has been made on 10 meteorologically significant Tiros TV cloud-cover pictures (Figures 12 to 21); these pictures are the same pictures as those used for the study reported in Reference 4. Results have been obtained with each TV element quantized to 4 and 6 bits.

Assume that the video data are to be scanned an element at a time from left to right and top to bottom, beginning with the top leftmost TV element. The choice of the parameter M (memory size) determines the number of M-dimensional cubes (called M-cubes in this paper) which are required to store the statistical structure of the data. For example, if the data are quantized to 16 levels (Q) and a memory size M of 2 is chosen, then there must be exactly Q^M or $(16)^2 = 256$ 2-cubes. Figure 2 shows memory cell geometries for Q = 16 and M = 1, 2, and 3.* The process begins by scanning the data one element at a time and observing the random variable X. At each observation of X the prediction for its immediate successor is computed from the statistics stored in the M-cube associated with the particular X under observation. The prediction error is given by

$$E_p = |x_a - x_p| , \qquad (7)$$

where x_a is the actual value and x_p is the predicted value. If $E_p \leq T$ where T is a preset allowable error threshold, the element is predictable and need not be transmitted. If, however, $E_p > T$ this particular element is not predictable and must be transmitted in unmodified form.

The data compression system at the transmitter end must provide the receiver with the data necessary to reconstruct the original

---

*For memory sizes larger than 3, the number of storage locations required becomes unwieldy for practical computer simulations.

message within the allowable error threshold T. To accomplish this, the predictor at the transmitter end must operate on exactly the same data which it will send to the receiver for reconstruction. Therefore, if an element is predictable $(E_p \leq T)$ it need not be transmitted, but the predicted value is treated as though it were the actual value and is also used to update the statistics stored in the M-cube defined by the X under observation. If, however, an element is not predictable $(E_p > T)$, the actual value is used to update the statistics. This is called "closed-loop" operation. The prediction mechanism could be evaluated in the "open-loop" mode. In open-loop operation predicted values do not replace actual values; thus the predictor operates on raw data only. The studies described in this report, however, were done in the closed-loop mode.

The example given previously described the formulation of the sample mean in terms of Equation 6. In the computer simulation it is not necessary to keep track of the relative frequency terms k_j/k because each M-cube defined by X can be composed of two storage locations, a sum location, and a counter location. At each observation of X, the sum location corresponding to this X is updated by adding to the existing sum either the actual or predicted value of the successor to X depending on whether the element is predictable. At the same time the corresponding counter location is incremented by one count for each observation of X. The k_j terms of Equation 6 are implicitly contained in the sum at all times. Therefore the prediction computation need be performed only when a prediction is required and is easily obtained by dividing the sum by the counter.

Because the learning period includes only a finite amount of past data, a prediction for the successor to a particular value of the random variable X could frequently be indeterminate because of a complete lack of past information; this is especially true at the beginning of the learning process. One solution might be to determine a prediction from the statistics contained in the M-cubes neighboring the particular M-cube defined by the X under observation. This approach, however, does not solve the problem at the beginning and in the very early stages of the learning period. The obvious solution then is to make some initial assumption for the successor to each of the values which the random variable X can assume before the learning process begins. If it turns out that the initial assumption was a poor one, it will affect the efficiency of the prediction mechanism less and less significantly as more and more of the data are observed. This scheme was utilized in the simulation of the conditional expectation predictor, and the choice of the prelearning assumption was made based on the results of the zero-order hold predictor (Reference 4).

Experiments with the zero-order hold predictor showed that very frequently an element intensity was within ±1 or ±2 quantum levels of its predecessor. Thus, if the random variable X associated with the conditional expectation predictor is a 2-dimensional random vector

$$X = (x_r, x_{r+1}) , \tag{8}$$

the prelearning assumption for the successor to the $(r + 1)$st element is the $(r + 1)$st element. Similarly, if $X = (x_{r-1}, x_r, x_{r+1})$, the prelearning assumption would again be x_{r+1}.

Quite often a suitable prediction can not be derived from the statistics contained in the particular cube defined by the X under observation. When this occurs, it is possible to utilize the neighborhood statistics as the source of a secondary prediction. For example, (see Figure 3) suppose the random variable under observation is

$$X = (i, j) \qquad 1 \leq i \leq Q, \ 1 \leq j \leq Q$$

where i and j are values of element intensity specifying the coordinates of a specific 2-cube in the memory array. Suppose that the conditional expectation calculated from the statistics contained in (i, j) is inadequate; that is, $E_p > T$. As soon as it is determined that the prediction error E_p exceeds the threshold T, a secondary prediction is provided by computing the mean of the statistics contained in the 2-cubes in the neighborhood of cube (i, j). The boundaries of the neighborhood are governed by the allowable prediction error threshold T so as to accommodate the fidelity criterion. For example, if T is ± 1 quantum level and a suitable prediction cannot be made from cube (i, j), the cubes which are not more than ± 1 quantum level away from (i, j) are those from which the secondary prediction is determined. The concept of providing a secondary prediction if the primary prediction fails is in itself attractive, but this attractiveness is somewhat dulled when one considers that the use of alternate prediction modes in the same compression mechanism complicates the coding problem since the receiver must determine the source of each prediction.

The discussion thus far assumes that the TV data are observed serially one element at a time, scanning from left to right. There is some advantage, however, in observing the data not only from left to right along a TV line but also from line to line so as to take advantage of the vertical correlation in the TV data. Figure 4 depicts the geometry of the TV data. If one wishes to operate the prediction mechanism only on data scanned serially from left to right, the random variable X would take the form of an ordered pair of adjacent elements on the same line $\left(\text{e.g., typically, } X = \left[x_{i,j}, x_{i,j+1} \right] \right)$. If, however, one wishes to take advantage of line-to-line
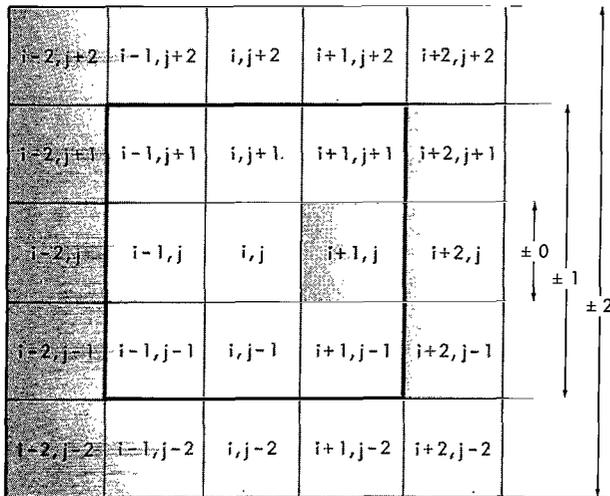
Figure 3—Two-dimensional memory cube and
its neighboring cubes.
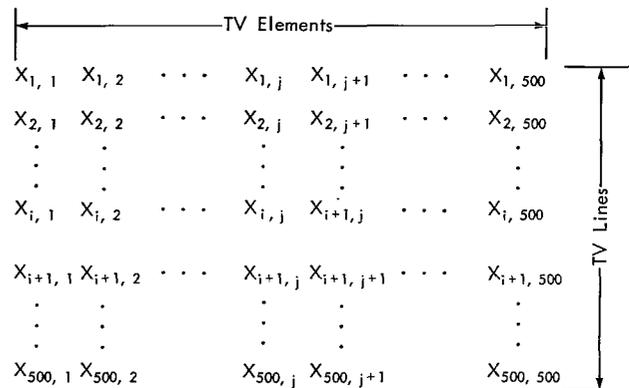
Figure 4—Geometry of TV data.

correlation, X might consist of an ordered pair of TV elements of the form $X = \left(x_{i,j}, x_{i+1,j+1}\right)$ where the element to be predicted is $x_{i,j+1}$. This scheme might be termed an elementary, two-dimensional predictor.

So far not too much has been said about the learning operation. Actually, the learning operation of the conditional expectation predictor (Method II of Reference 3) is not so explicit as that of the linear predictor (Method I of Reference 3). In Method I the learning period is composed of about 20 data samples preceding the elements to be predicted. The function of this learning period is to develop an optimal operator based on these 20 previous points. In Method II, however, the function of the learning period is to determine the optimal operation to predict the successor to the present observation of the random variable X. This is the basic difference between Method I and Method II. Method I determines an *optimal operator* based on a few points preceding the elements to be predicted, while Method II determines the *optimal operation* based on previous observations of the successor to the particular X under observation. Also, in Method I, either a linear or a nonlinear operation is explicitly chosen. For example, Method I as it is described in Reference 4 is very obviously linear. Method II, however, does not distinguish between linear and nonlinear operations. The prediction mechanism simply proceeds to the optimum operation without restriction to either linear or nonlinear operation.

Since the learning period of the linear predictor is used to determine an operator over a fairly small number of previous data samples, and the learning period of the conditional expectation predictor is used to observe occupancies of a relatively large number of M-cubes, it seems reasonable that the second method should require a much larger learning period than that required by the first method. Results from computer simulations included in the discussion of results appear to support this argument. It is important to note that in Method I the optimal operator is found over a learning period just preceding the sequence of elements to be predicted, and a new learning process does not begin until the mean square prediction error exceeds a preset threshold. In Method II, however, the learning process is more continuous in nature, and prediction and learning take place almost simultaneously.

Method I as described in Reference 4 utilizes two thresholds. The first is the threshold T, which is the allowable error between true and predicted values of a data sample. The second threshold is associated with the mean square prediction error which is calculated periodically to determine the prediction ability of the present operator. When the mean square prediction error exceeds this threshold, the prediction mechanism is signaled to restart its learning operation. Thus far in the description of Method II only one threshold has been mentioned. This is the threshold T which corresponds exactly to the first threshold of Method I. Method II in its present configuration does not employ a second threshold equivalent to that of Method I.

In the first simulation of the conditional expectation predictor, the learning-period length was chosen based on parametric trials with element compression ratio serving as the figure of merit. Figure 5 shows these results with the data quantized to 6 bits per TV element, the memory M = 1 TV element, and an allowable error threshold T of ±2 quantum levels, where cumulative element compression ratio is the 4800-line (10-TV-picture) average compression ratio. This is not an
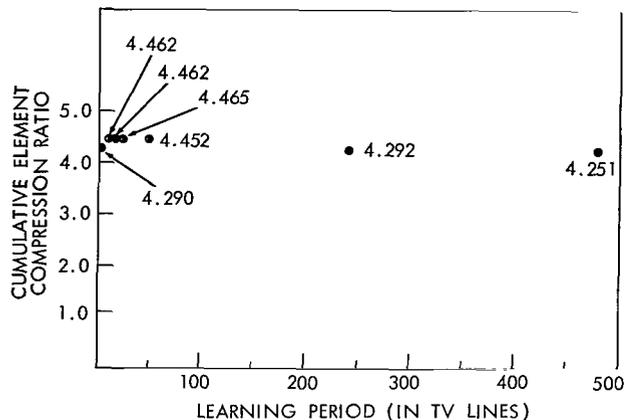
Figure 5—Cumulative element compression ratio versus learning period (in TV lines) with Q = 64 (6 bits/element), M = 1, T = ±2 quantum levels.

ideal way to handle the learning operation. It might be worthwhile to monitor the mean square prediction error and introduce a second threshold as in Method I. The problem with this, as in the present configuration, is that the start of a new learning period causes a large instantaneous drop in the amount of statistical data available with which to make predictions.

A solution free from this problem is to allow the statistical structure to decay slowly to some effective N element average. Each M-cube of the memory array is composed of a summer and a counter. Suppose the counter is allowed to build up freely to N observations and future observations are handled as follows: Let

$$\sigma_N = \text{Sum contained in the sum location after N observations}$$

$$\rho_{N+1} = (N+1)\text{st sample.}$$

Then at the (N + 1)st observation σ_N is replaced by

$$\sigma_{N+1} = \left(\sigma_N + \rho_{N+1}\right)\left(\frac{N}{N+1}\right). \tag{9}$$

Furthermore,

$$\sigma_{N+2} = \left(\sigma_{N+1} + \rho_{N+2}\right)\left(\frac{N}{N+1}\right) = \left(\sigma_N + \rho_{N+1}\right)\left(\frac{N}{N+1}\right)^2 + \rho_{N+2}\left(\frac{N}{N+1}\right),$$

$$\sigma_{N+3} = \left(\sigma_{N+2} + \rho_{N+3}\right)\left(\frac{N}{N+1}\right),$$

and so on. Thus the most recent observation is weighted most significantly; the second most recent observation, the second most significantly; and so forth.

DISCUSSION OF RESULTS

Figures 12 to 21 are copies of the Tiros TV cloud-cover pictures used in this study. These 10 pictures are the same as those used in the study reported in Reference 4. The background on the original analog data, the construction of the unmodified digital pictures, and the description of the display of these same pictures after processing with the prediction mechanism are also contained in Reference 4. The pictures which appear in this report probably will have lost some of the

8

linearity of the gray scale because of the reproduction process but their overall quality should not be degraded because of the large number of gray scales present.

The complete data compression system embraces two problems, the prediction problem and the coding problem. Although they are not independent, it is possible to think of them as two distinctly separate problems. In order to separate them, one needs to impose the constraint on the prediction mechanism that it at least does not hamper the coding mechanism in reasonably representing the data. With this consideration in mind, the results obtained so far can be presented in two parts. The first section will deal with the characteristics of the prediction mechanism with *element compression ratio* as the standard of comparison. The second section presents some possible approaches to the coding problem with *bit compression ratio* as the standard of comparison.

Results of Simulations of Prediction Mechanism

Since it was assumed that the prediction problem and the coding problem were separate, the objective of the simulation of the prediction technique was to maximize the element compression ratio. Element compression ratio is defined as the ratio of the total number of TV elements in the original unmodified picture to the total number of unmodified TV elements which must be transmitted after the picture is processed by the prediction mechanism. Element compression ratio then is simply a measure of the "predictability" of the data, and certainly does not include coding considerations. The objective of the initial work was to simulate the technique and evaluate the results with element compression ratio serving as the figure of merit.

Learning Period Considerations

Table 1 shows element compression ratios for the basic prediction scheme with data quantized to 6 bits per TV element, $M = 1$, $T = \pm2$ quantum levels, and learning periods varying from 2 TV

Table 1

Various Learning Periods with $Q = 64$ (6 Bits/Element); $M = 1$ and $T = \pm2$.

Figure Number	Element Compression Ratio for Learning Period L of —						
	480 TV lines	240 TV lines	48 TV lines	24 TV lines	16 TV lines	10 TV lines	2 TV lines
12	5.478	5.740	5.970	5.978	5.935	5.964	5.601
13	4.814	4.820	5.023	5.024	5.006	4.967	4.725
14	4.322	4.432	4.725	4.720	4.709	4.699	4.479
15	4.873	4.935	5.063	5.082	5.094	5.036	4.878
16	3.908	3.962	4.053	4.053	4.042	4.062	3.897
17	3.688	3.710	3.822	3.814	3.810	3.793	3.660
18	4.206	4.205	4.329	4.335	4.362	4.400	4.300
19	2.376	2.385	2.450	2.460	2.452	2.447	2.381
20	4.256	4.237	4.461	4.533	4.527	4.559	4.336
21	9.550	9.527	10.103	10.205	10.381	10.453	10.025
Cumulative Element Compression Ratio	4.251	4.292	4.452	4.465	4.462	4.462	4.290

lines to 480 TV lines in one picture. There certainly are no significant gains in compression ratio for any of the learning periods used. However, as a first choice one might pick a learning period length of 256 samples (approximately 1/2 TV line) for this case. The reasoning for this is quite simple. Consider the general case with a memory size M, with the data quantized to Q quantum levels. As described earlier in the report, prediction depends on the conditional expectation of the successor to a random variable X whose sample space size depends on M. In particular the sample space size $S = Q^M$. Thus, if $M = 1$, $Q = 16$, and the objective is to predict X_p when X_{p-1} is known, there are exactly $(16)^2$ or 256 possibilities for the set (X_{p-1}, X_p). Therefore, if all cases were equiprobable, one would have to allow the learning period to cover 256 samples to be sure that each case was observed at least once. Thus for the general case of Q and M, one might choose as the minimum learning period length $L = Q^{(M+1)}$. This certainly does not represent the optimum learning period length, but it does provide a guideline as to the minimum learning period length. One might govern the upper bound of the learning period size by investigating the changes in the structure of the statistics as more and more samples are observed. In any case it is advantageous to keep the learning period size as small as possible, since the data are suspected to be somewhat nonstationary.

The results of successive experiments designed to test the performance of the conditional expectation prediction on both 6- and 4-bit data are contained in Tables 2 and 3. Figures 6 and 7 depict these results as bar plots. A few conclusions can be drawn from these results:

(1) There is essentially no difference between the results for $M = 1$ and $M = 2$ without the statistical-neighborhood and two-dimensional prediction modes.

Table 2

Element Compression Ratios with $Q = 64$ (6 Bits/Element) and $T = \pm 2$.

Figure Number	Element Compression Ratio for —				
	$M = 1$; $L = 480$ TV lines	$M = 1$; $L = 16$ TV lines	$M = 1$; $L = 16$ TV lines with SNP[1]	$M = 2$; $L = 16$ TV lines with SNP[1]	$M = 2$; $L = 16$ TV lines with both SNP[1] and EAP[2]
12	5.478	5.935	7.420	8.712	11.425
13	4.814	5.006	5.982	7.120	9.734
14	4.322	4.709	5.686	6.703	8.134
15	4.873	5.094	6.249	7.217	8.444
16	3.908	4.042	4.826	5.521	6.413
17	3.688	3.810	4.424	4.994	5.913
18	4.206	4.362	5.295	6.543	6.863
19	2.376	2.452	2.825	3.316	3.297
20	4.256	4.527	5.469	6.874	7.875
21	9.550	10.381	12.623	15.941	17.550
Cumulative Element Compression Ratio	4.251	4.462	5.330	6.301	7.196

[1] Statistical neighborhood predictor.
[2] Element area predictor.

(2) A slight improvement in compression ratio was achieved when the learning period L was reduced from 480 to 16 TV lines.

(3) Significant improvements in compression ratio were achieved with the addition of the neighborhood and two-dimensional predictors.

Table 3

Element Compression Ratios with Q = 16 (4 Bits/Element) and T = ±1.

Figure Number	Element Compression Ratio for —				
	M = 1; L = 480 TV lines	M = 1; L = 16 TV lines	M = 1; L = 16 TV lines with SNP [1]	M = 2; L = 16 TV lines with SNP[1]	M = 2; L = 16 TV lines with both SNP[1] and EAP[2]
12	7.443	7.846	9.697	9.762	14.267
13	6.355	6.725	8.182	7.995	12.705
14	6.092	6.410	7.777	7.566	10.390
15	6.256	6.721	8.015	7.786	10.356
16	4.904	5.258	6.331	6.523	8.146
17	4.611	4.896	5.575	5.753	7.379
18	5.457	5.588	6.870	7.051	7.824
19	2.930	3.055	3.593	3.750	3.890
20	5.418	6.334	7.789	7.923	9.690
21	13.442	13.449	17.127	15.550	19.954
Cumulative Element Compression Ratio	5.494	5.835	7.009	7.071	8.787

[1] Statistical neighborhood predictor.
[2] Element area predictor.

M = 1, L = 480 TV LINES

M = 2, L = 480 TV LINES

M = 1, L = 16 TV LINES

M = 1, L = 16 TV LINES, WITH NEIGHBORHOOD PREDICTOR

M = 2, L = 16 TV LINES, WITH NEIGHBORHOOD PREDICTOR

M = 2, L = 16 TV LINES, WITH BOTH NEIGHBORHOOD AND AREA PREDICTORS

1.0   2.0   3.0   4.0   5.0   6.0   7.0   8.0   9.0

TEN PICTURE (4800 Line Average) ELEMENT COMPRESSION RATIO
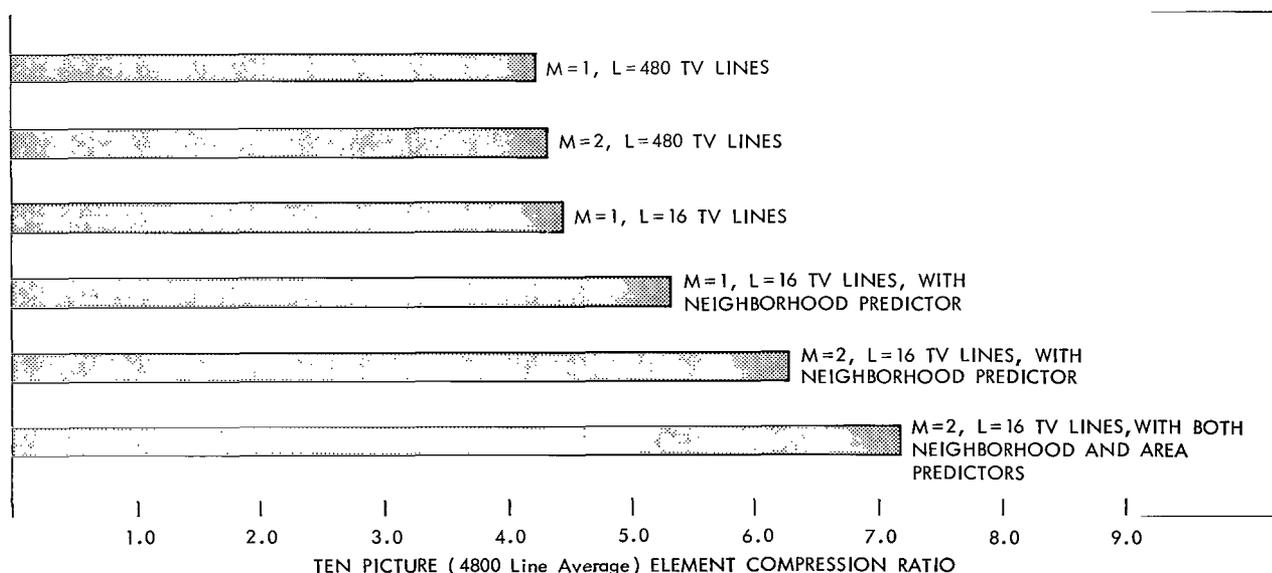
Figure 6—Performance bar plot for conditional expectation predictor with Q = 64 (6 bits/element) and T = ±2 quantum levels.
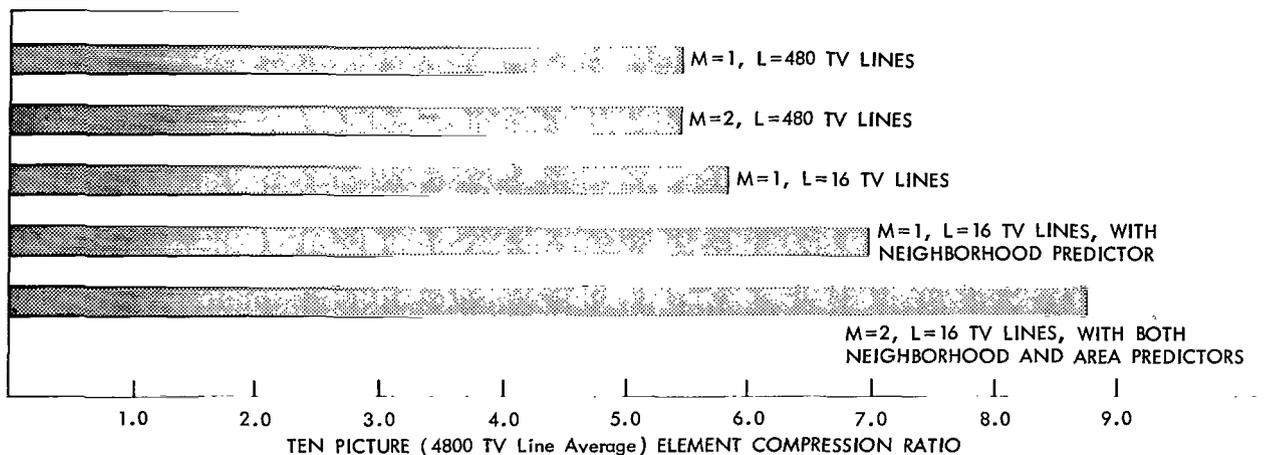
Figure 7—Performance bar plot for conditional expectation predictor with Q = 16
(4 bits/element) and T = ±1 quantum level.

Comparison with Other Techniques

Figures 8 and 9 summarize in bar-plot form the relative performance of:

(1) The zero-order hold predictor,

(2) The linear predictor of Reference 4 (Method I - Reference 3),

(3) The conditional expectation predictor (Method II - Reference 3).

The linear predictor (Method I) produces a 10-picture cumulative element compression ratio of about 3:1 for 6 bits per element and T = ±2. The zero-order hold predictor provides a compression ratio of about 4.2:1 for 6 bits per element and T = ±2 and one of about 5:1 for 4 bits per element and T = ±1. The conditional expectation predictor in its most elementary form (without neighborhood and two-dimensional predictors) performs slightly better than does the zero-order hold. The conditional expectation method along with the neighborhood and two-dimensional predictors shows a significant gain with a ratio of more than 7:1 for 6 bits per element and T = ±2 and a ratio of nearly 9:1 for 4 bits per element and T = ±1. One might reason that the zero-order hold does very well with respect to the other two methods cited, when the relative complexity of the schemes is considered. The only explanation as to why the zero-order hold predictor does this well is that the information source is nonstationary. Note, however, that the conditional expectation predictor does about 140 percent better than the linear predictor and about 70 percent better than the zero-order hold. One reason that the conditional expectation predictor does so much better than the linear predictor is that the former is not restrictive with respect to linear or nonlinear operations and therefore is able to predict well despite the nonstationary character of the data.

It was mentioned earlier that the incorporation of the statistical neighborhood predictor as an alternate prediction mode contributes to the coding costs since the receiver must determine which

LINEAR PREDICTOR, M=0, L=20, T=±2.5

LINEAR PREDICTOR, M=3, L=20, T=±2.5

ZERO-ORDER HOLD PREDICTOR, T=±2

CONDITIONAL EXPECTATION PREDICTOR, M=1, L=480 LINES, T=±2

CONDITIONAL EXPECTATION PREDICTOR, M=2, L=480 LINES, T=±2

CONDITIONAL EXPECTATION PREDICTOR, M=1, L=16 LINES, T=±2

CONDITIONAL EXPECTATION PREDICTOR, M=1, L=16 LINES, T=±2, WITH NEIGHBORHOOD PREDICTOR

CONDITIONAL EXPECTATION PREDICTOR, M=2, L=16 LINES, T=±2, WITH NEIGHBORHOOD PREDICTOR

CONDITIONAL EXPECTATION PREDICTOR, M=2, L=16 LINES, T=±2, WITH BOTH NEIGHBOR-HOOD AND AREA PREDICTORS

1.0   2.0   3.0   4.0   5.0   6.0   7.0   8.0   9.0
TEN PICTURE (4800 Line Average) ELEMENT COMPRESSION RATIO

Figure 8—Performance of conditional expectation predictor relative to linear predictor of reference 4 and zero-order hold predictor with Q = 6 bits per element.

ZERO ORDER HOLD PREDICTOR, T=±1

CONDITIONAL EXPECTATION PREDICTOR, M=1, L=480 LINES, T=±1

CONDITIONAL EXPECTATION PREDICTOR, M=2, L=480 LINES, T=±1

CONDITIONAL EXPECTATION PREDICTOR, M=1, L=16 LINES, T=±1

CONDITIONAL EXPECTATION PREDICTOR, M=1, L=16 LINES, T=±1 WITH NEIGHBORHOOD PREDICTOR

CONDITIONAL EXPECTATION PREDICTOR, M=2, L=16 LINES T=±1 WITH BOTH NEIGHBORHOOD AND AREA PREDICTORS

2.0   4.0   6.0   8.0   10.0   12.0   14.0   16.0   18.0
TEN PICTURE (4800 Line Average) ELEMENT COMPRESSION RATIO

Figure 9—Performance of conditional expectation predictor compared with performance of zero-order hold predictor with T = ±1 and Q = 4 bits per element.

13

statistics were used to make the prediction. One way to overcome this problem would be to constrain the transmitter always to make predictions from the neighborhood statistics. Preliminary results with the technique have shown that the mechanism is not able to predict nearly so well when only neighborhood predictions are permitted.

An explanation for the basis of the choice of the allowable prediction error T seems to be necessary at this point. Obviously the selection of T is very important to the performance of the prediction mechanism. Since the information source here is video data representing cloud-cover pictures, the effect of the choice of T can be easily observed when the data processed by the prediction mechanism are displayed. The problem here is that a judgment of the quality of a compressed picture must be made subjectively by eye; thus the only solution is to try different thresholds until the maximum threshold which allows retention of minimum acceptable picture quality is determined. The choices of T = ±2 quantum levels for the 6-bit case and T = ±1 quantum level for the 4-bit case were made after experimenting with a number of thresholds. Figure 10 is a picture showing the effects of too large a value of T with the data quantized to 64 levels and T = ±4.
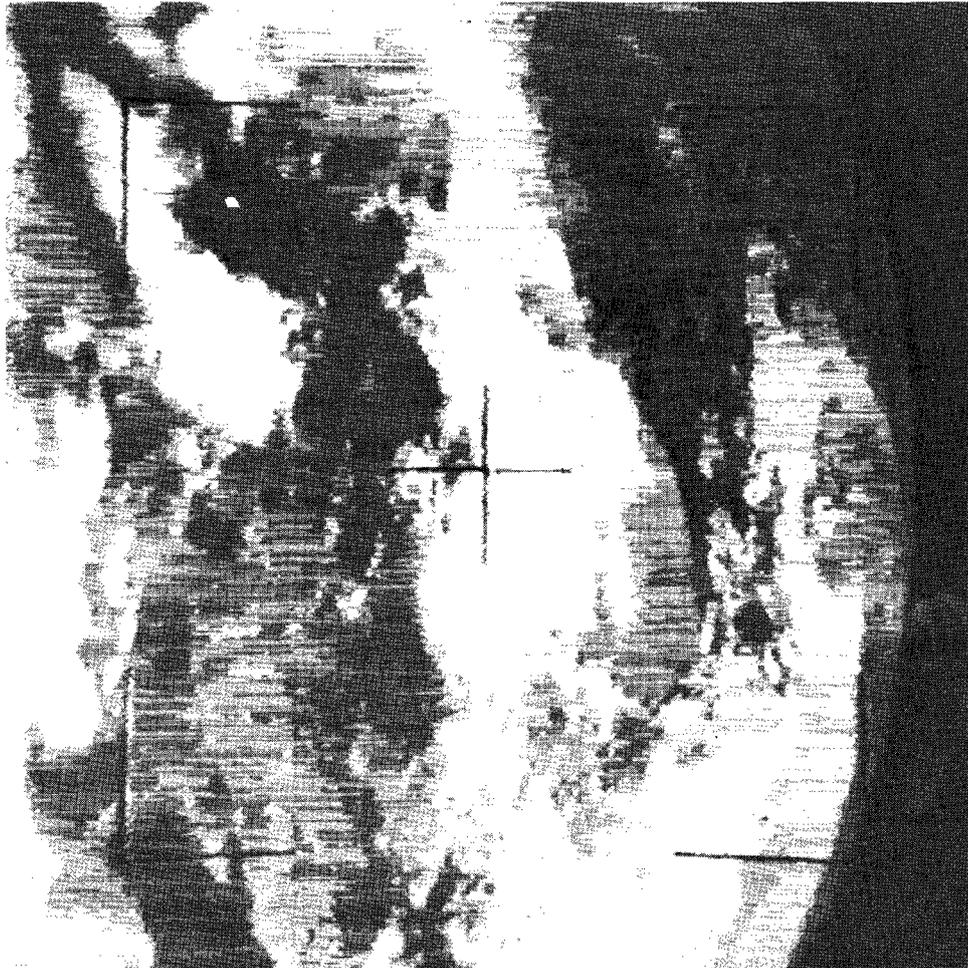
Figure 10—Effect of choosing too large a value of T (allowable prediction error). In this case Q = 64 (6 bits/element) and T = ±4 quantum levels.

Coding Considerations

The prediction problem, while not completely defined, has certainly been investigated more thoroughly than has the coding problem. The most important question is, "After prediction what does the transmitter send to the receiver"? This report will not deal explicitly with the coding problem, but will offer a few observations about it. Actually, the problem of coding for a data compression system is not an easy one, and very little work has been done in this area.

The problem with most standard coding schemes is that they require knowledge of the statistics of the data. The prediction philosophy clearly states that no *a priori* knowledge of the statistics is necessary. It therefore seems reasonable that the coding philosophy should not be constrained by this requirement either.*

In order to evaluate any hypothesis adequately, it is helpful to have some standard of comparison which is optimum in some sense. Suppose that P is the probability of making an accurate prediction and also that each of Q levels is equally likely when accurate prediction is not possible. If it is also assumed that the ability to predict is sample-to-sample independent, then theory explains that in the noise-free case a bit compression ratio (including coding costs) of

$$\frac{\log_2 Q}{P \log_2 \left(\frac{1}{P}\right) + (1-P) \log_2 \left(\frac{Q}{1-P}\right)} \tag{10}$$

can be approached with optimum coding. Figure 11 is a family of curves of bit compression ratio C_B versus element compression ratio C_E with $\log_2 Q = 4$ and 6. The probability of predicting P accurately is related to the element compression ratio C_E by

$$P = 1 - \frac{1}{C_E} . \tag{11}$$

Table 4 provides examples of resultant bit compression ratios for each of the 10 pictures with $Q = 64$ and $T = \pm 2$, and with $Q = 16$ and $T = \pm 1$. It must be made clear that these results are by no means quotations of bit compression ratios one could obtain in practice for the following two reasons:

(1) The results assume optimum coding which would probably not be attainable in practice.

(2) These results apply to the noiseless channel and do not account for necessary error-correction coding.

These data are presented solely to provide guidelines to those who demand results which are in line with practical arguments.
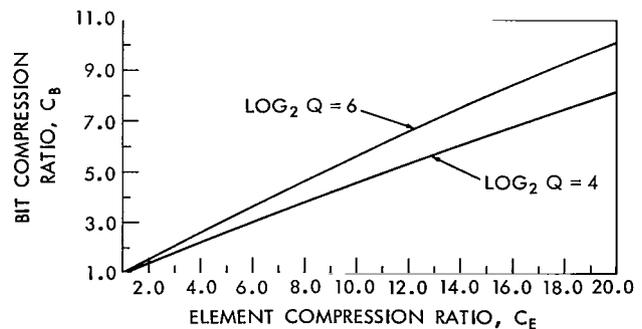
Figure 11—Bit compression ratio C_B versus element compression ratio C_E for $\log_2 Q = 4, 6$.

---

*Reference 4 shows some examples of coding the compressed data with variations of run-length coding. These results are interesting and similar simulations might be made with the conditional expectation predictor.

Table 4

Element Compression Ratios and Corresponding Bit Compression Ratios
for Conditional Expectation Predictors.

Figure Number	Q = 64 levels; T = ±2; M = 2; L = 16 TV lines		Q = 16 levels; T = ±1; M = 2; L = 16 TV lines	
	Element Compression Ratio	Bit Compression Ratio	Element Compression Ratio	Bit Compression Ratio
12	11.425	6.285	14.267	6.189
13	9.734	5.481	12.705	5.607
14	8.134	4.705	10.390	4.757
15	8.444	4.851	10.356	4.729
16	6.413	3.846	8.146	3.888
17	5.913	3.593	7.379	3.705
18	6.863	4.064	7.824	3.762
19	3.297	2.218	3.890	2.160
20	7.875	4.576	9.690	4.473
21	17.550	9.127	19.954	8.219
Cumulative Compression Ratio	7.196	4.238	8.787	4.136

Comments on the TV Pictures

Each of Figures 12 to 21 contains in the following order:

(1) A photograph of the original analog picture.

(2) A photograph of the original digital picture constructed from the analog data.

(3) Two photographs of the digital data redisplayed after processing by the conditional expectation predictor.

Reference 4 contains a good deal of information on the history and specific characteristics of many of these pictures, as well as a description of the techniques used to display them.

The author will not attempt to give a detailed meteorological analysis for each picture but will rather provide a general comparison of the pictures processed by the conditional expectation predictor with the unmodified pictures as well as with those processed by other compression techniques. It is impossible for the untrained eye to pass judgment as to the retention of meteorological fidelity of the compressed pictures.* The only alternative for the layman is to compare the compressed pictures subjectively with the originals and to estimate the loss of apparent picture quality.

---

*Reference to a "compressed" picture does not imply that the picture geometry is made smaller or more compact in any way. It is simply true that the amount of data required to transmit a "compressed" picture over a communications link is less than the amount of data required to send the original picture.

When the pictures processed by the conditional expectation predictor are compared with the digital originals, the loss of picture quality is obvious but not objectionable. Contouring or "streakiness" in highly detailed regions seems to be the most popular complaint. This contouring is caused by the ability of the prediction mechanism to predict long sequences of elements at the same level successively. This effect becomes more pronounced as T is increased. A technique which might partially solve this problem is the use of a weighted prediction error criterion where prediction errors are accumulated until a present threshold has been exceeded.*

The pictures quantized to 4 bits per element with T = ±1 (Figures 12(c) to 21(c)) exhibit a higher degree of picture quality degradation than do the pictures quantized to 6 bits per element with T = ±2 (Figures 12(d) to 21(d)). The reason for this is that a threshold of ±1 quantum level at 4 bits per element is a larger percentage error than a threshold of ±2 quantum levels at 6 bits per element. Both the 6-bit and the 4-bit pictures are displayed with 16 shades of gray. The 6-bit compressed pictures with T = ±2 are acceptable while the 4-bit compressed pictures with T = ±1 seem to be at the threshold of acceptability. Perhaps the best compromise would be to use data quantized to 5 bits per element and allow T = ±1, which is the same percentage error as 6 bits per element with T = ±2. Thus one would expect the element compression ratios for the 5-bit, T = ±1 case to be about the same as those for the 6-bit, T = ±2 case. If these 5-bit pictures were also displayed with 16 gray shades, then they would possess about the same quality as the 6-bit pictures. The first-order entropies of the unmodified digital data quantized to 4, 5, and 6 bits per element are given in Table 5. The entropies for the 5- and 6-bit pictures are almost exactly the same, while the entropies for the 4-bit pictures are somewhat smaller.

The reader may find it interesting to compare the pictures processed by the conditional expectation predictor with those processed by the zero-order hold and linear predictors which are discussed in Reference 4. In general, the pictures processed by the conditional expectation predictor are of slightly better quality than zero-order-hold-predicted pictures. The pictures processed by the linear predictor are of higher quality than those processed by either of the other two methods.

Table 5

First-Order Entropies.

Figure Number	First-Order Entropy for —		
	Q = 64 levels (6 bits/TV element)	Q = 32 levels (5 bits/TV element)	Q = 16 levels (4 bits/TV element)
12	4.510	4.472	3.512
13	4.736	4.622	3.647
14	4.777	4.678	3.701
15	4.728	4.656	3.689
16	4.257	4.201	3.227
17	4.606	4.566	3.578
18	4.516	4.449	3.483
19	4.649	4.561	3.590
20	4.891	4.813	3.829
21	4.478	4.443	3.532

*This scheme was implemented by Davisson of Princeton and described in the report of his work in the 1965 Goddard Summer Workshop (Reference 5).
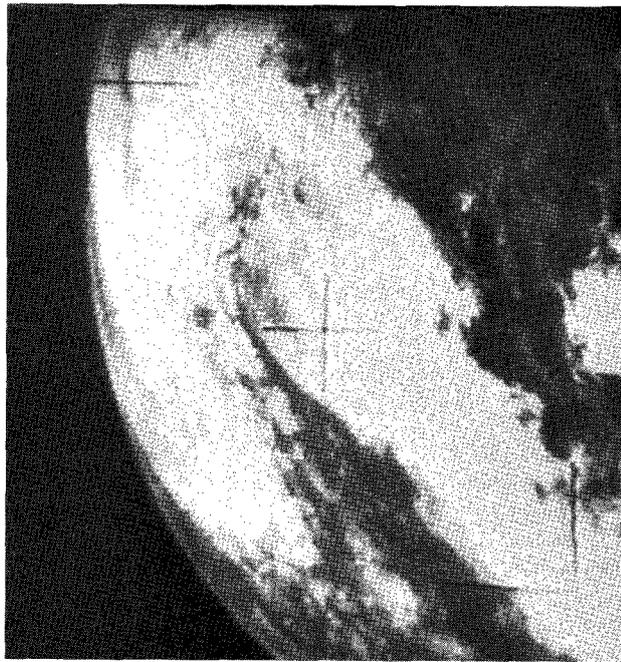
(a) Analog original.
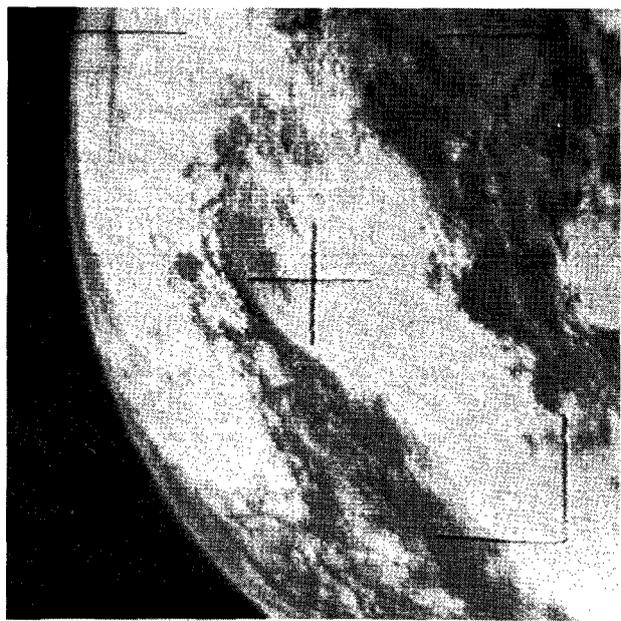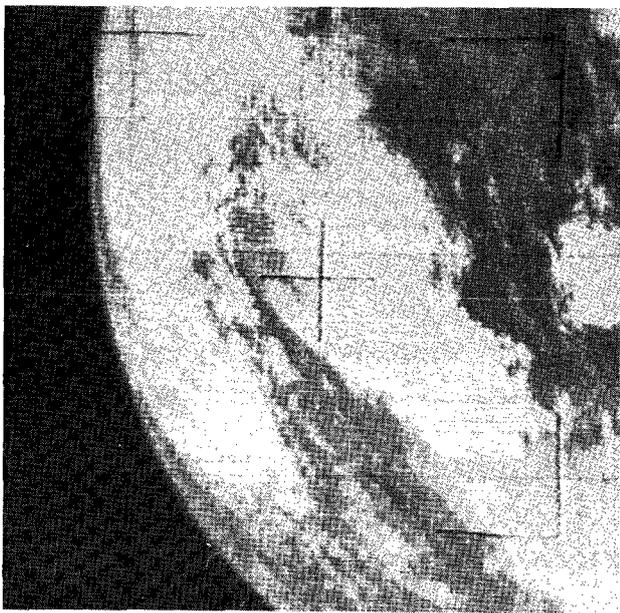
(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L = 16 TV lines. Element compression ratio, 14.267; bit compression ratio, 6.189.
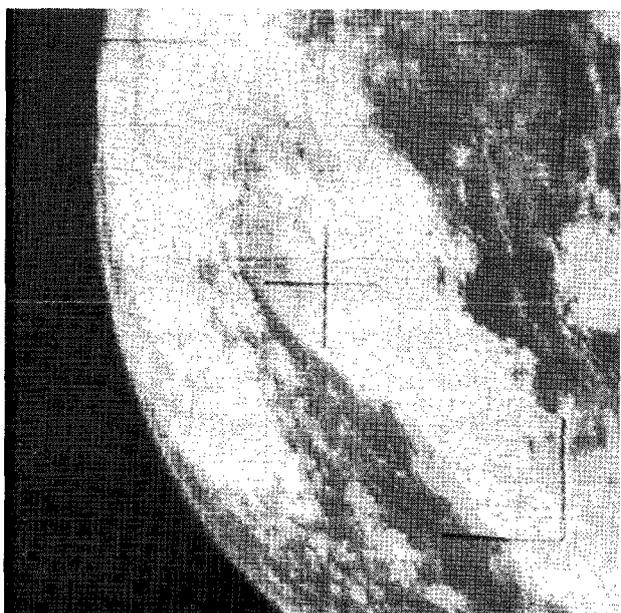
(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L = 16 TV lines. Element compression ratio, 11.425; bit compression ratio, 6.285.

Figure 12 - Pictures from Tiros III, orbit 4, frame 2, camera 2; direct transmission from satellite; principal point, 43.6N, 95.5W; subsatellite point, 41.0N, 89.2W.
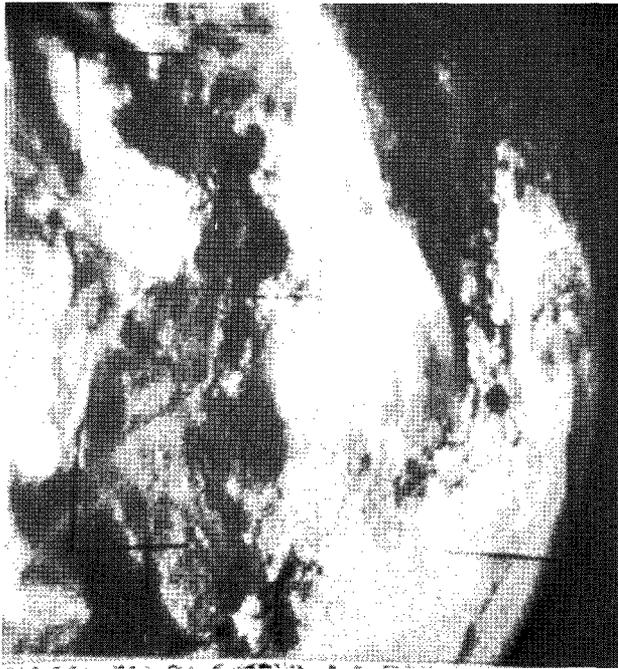
(a) Analog original.
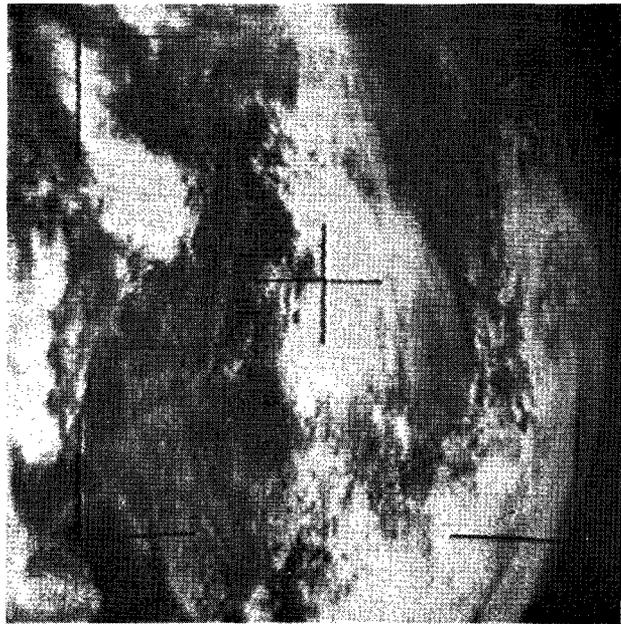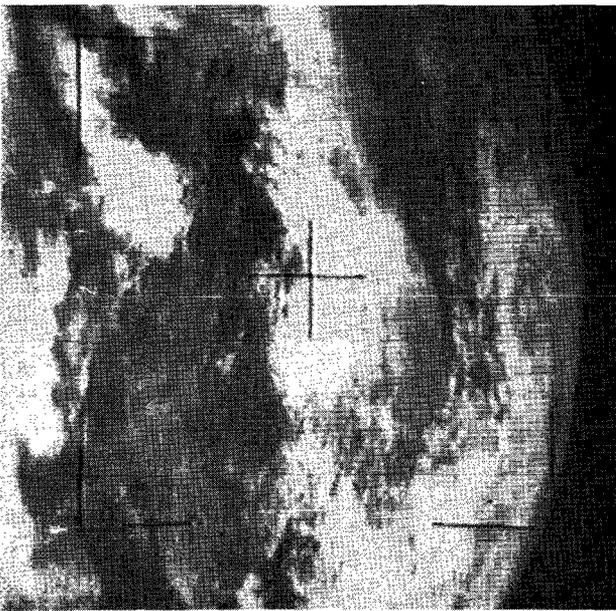
(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 12.705; bit compression ratio, 5.607.
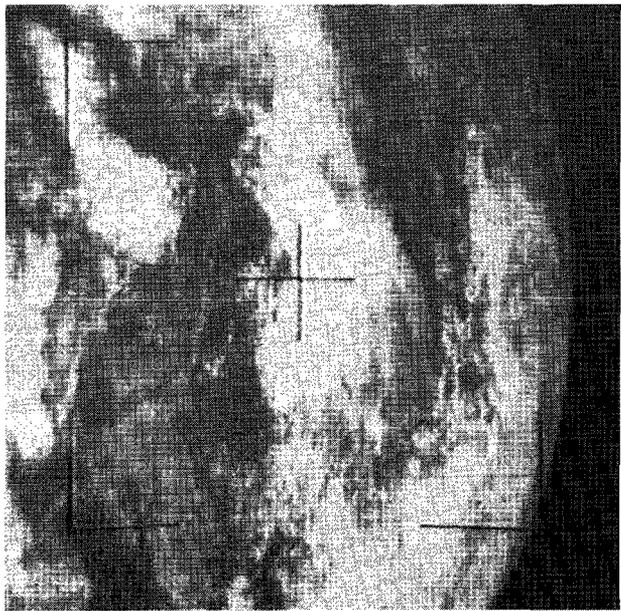
(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 9.734; bit compression ratio, 5.481.

Figure 13 – Pictures from Tiros III, orbit 4, frame 3, camera 2; direct transmission from satellite; principal point, 43.4N, 95.0W; subsatellite point, 40.8N, 88.8W.

(a) Analog original.
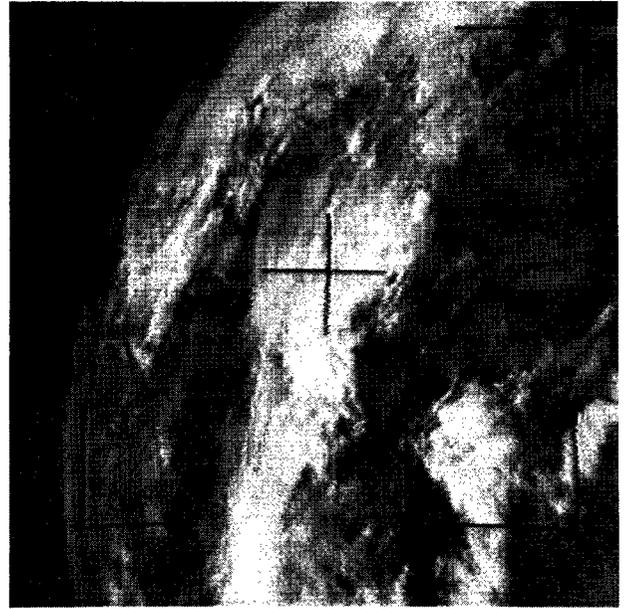
(b) Digital original.

(c) Processed copy generated by conditional expecta-
tion predictor with neighborhood and two-dimensional
predictors. Q=4 bits per TV element; T=± 1 level;
L = 16 TV lines. Element compression ratio, 10.390; bit
compression ratio, 4.757.
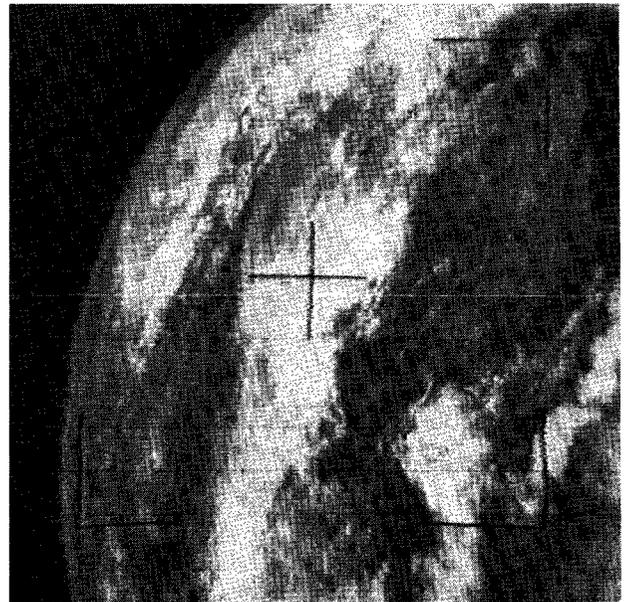
(d) Processed copy generated by conditional expecta-
tion predictor with neighborhood and two-dimensional
predictors. Q=6 bits per TV element; T=± 2 levels;
L = 16 TV lines. Element compression ratio, 8.134; bit
compression ratio, 4.705.

Figure 14 - Pictures from Tiros III, orbit 4, frame 4, camera 2; direct transmission from satellite;
principal point, 43.0N, 94.0W; subsatellite point, 40.5N, 88.1W.
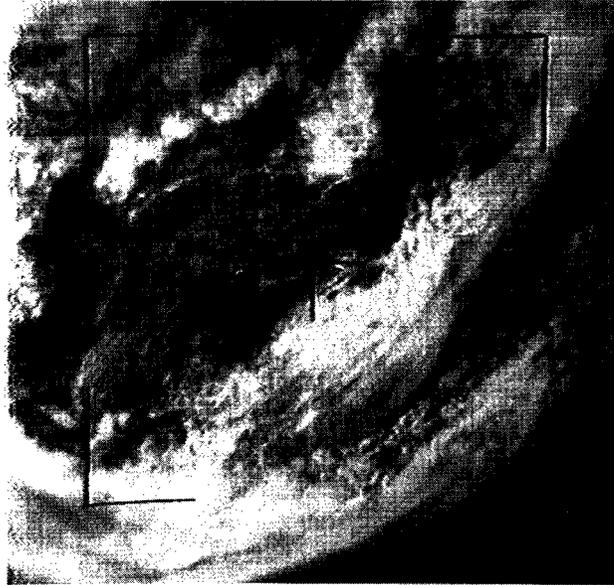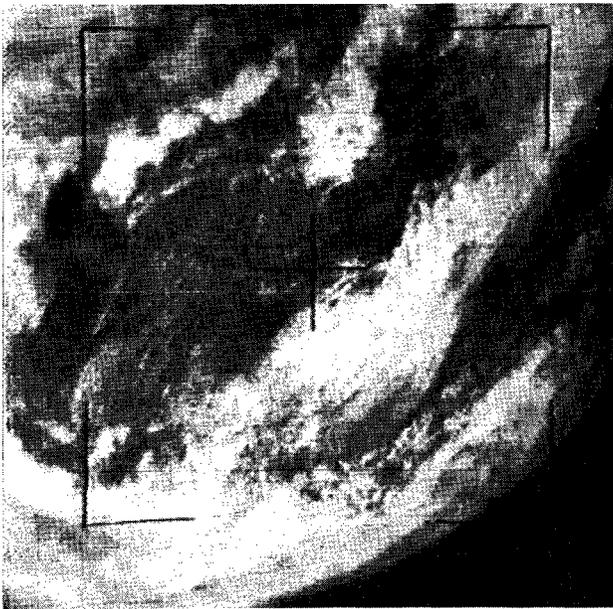
20

(a) Analog original.

(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 10.356; bit compression ratio, 4.729.
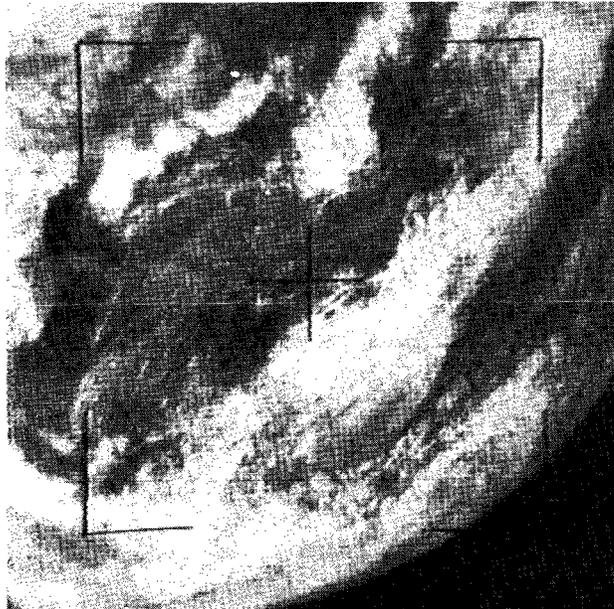
(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 8.444; bit compression ratio, 4.851.

Figure 15 - Pictures from Tiros III, orbit 4, frame 5, camera 2; direct transmission from satellite; principal point, 42.6N, 93.0W; subsatellite point, 40.1N, 87.3W.
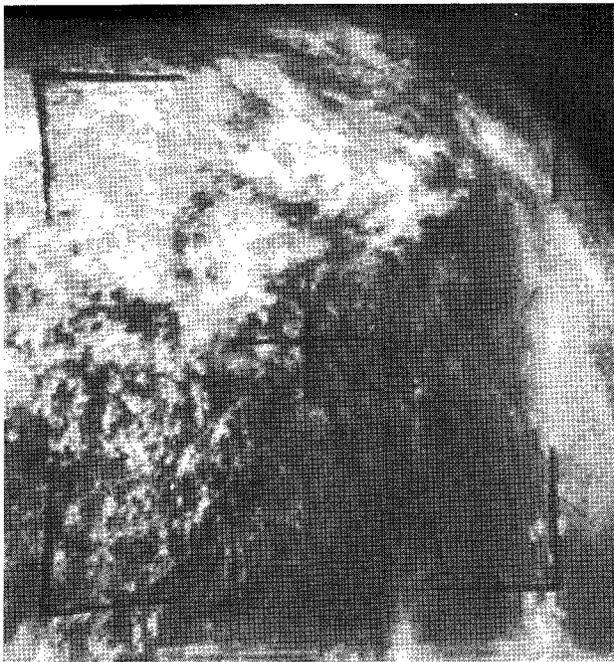
(a) Analog original.
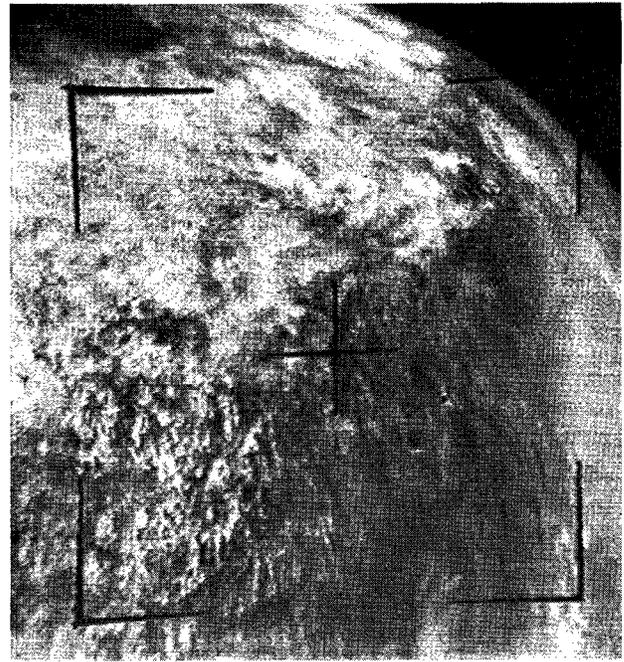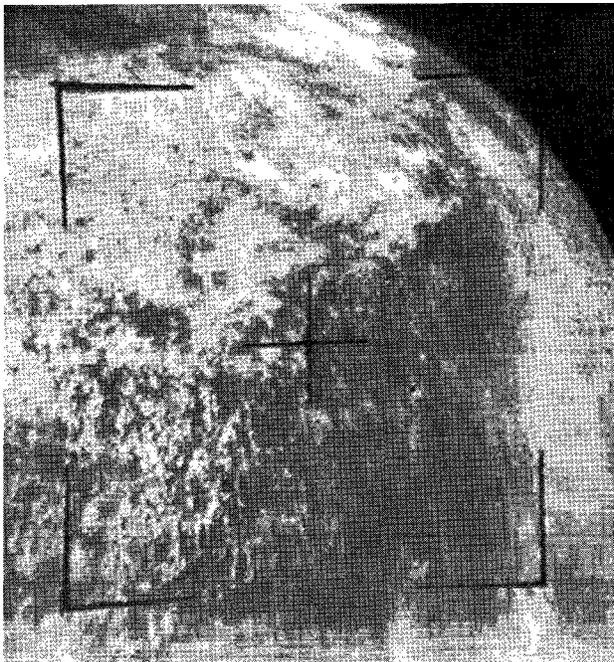
(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L = 16 TV lines. Element compression ratio, 8.146; bit compression ratio, 3.888.
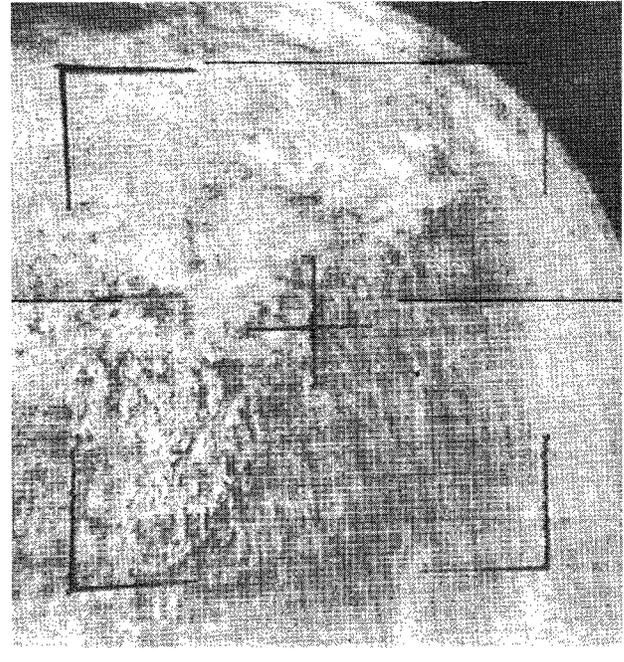
(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L = 16 TV lines. Element compression ratio, 6.413; bit compression ratio, 3.846.

Figure 16 - Pictures from Tiros III, orbit 102, frame 1, camera 1; taped before transmission from satellite; principal point, 11.5N, 4.0W; subsatellite point, 10.3N, 0.6W.
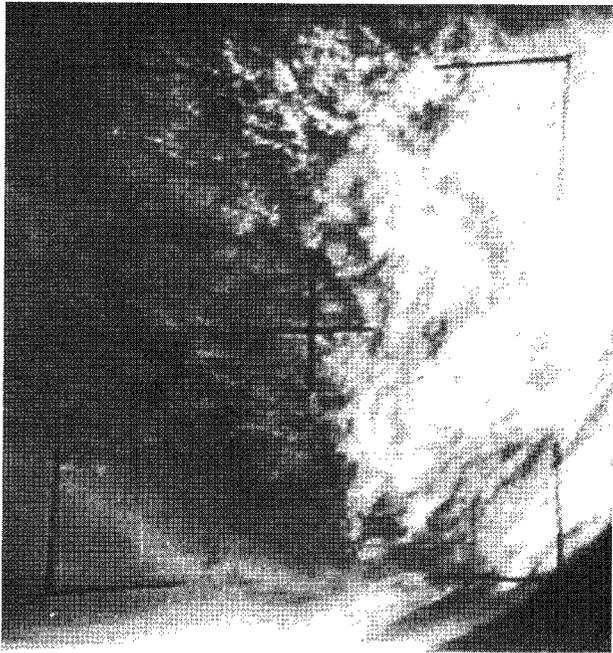
(a) Analog original.
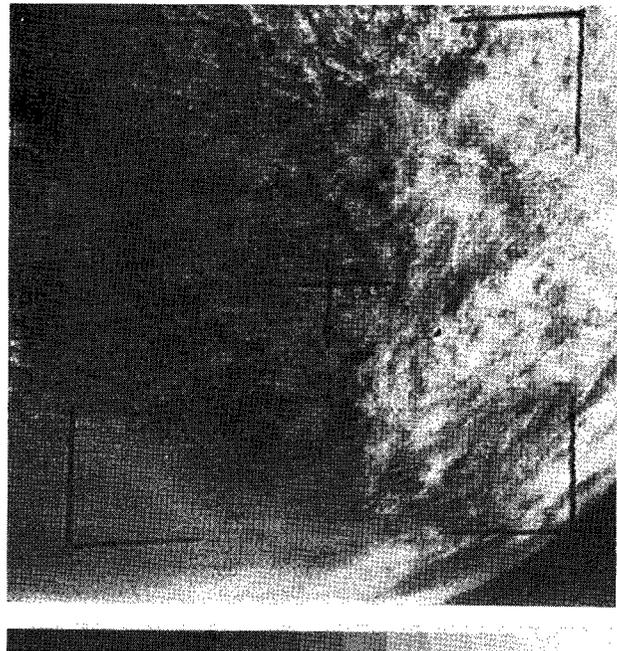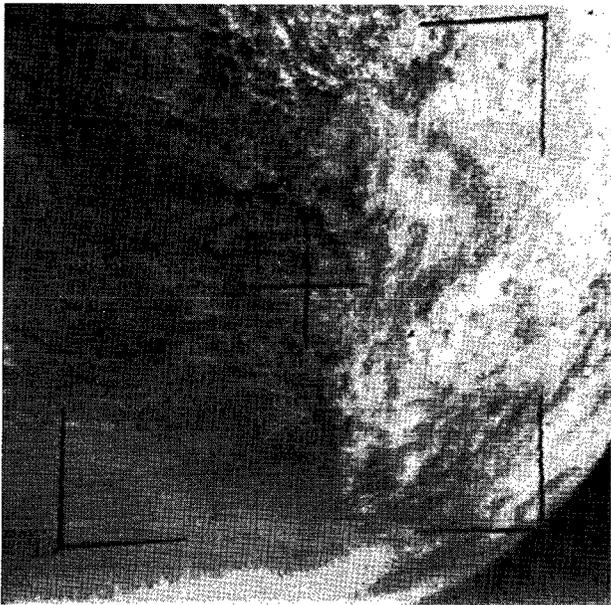
(b) Digital original.

(c) Processed copy generated by conditional expecta-
tion predictor with neighborhood and two-dimensional
predictors. Q=4 bits per TV element; T=± 1 level;
L = 16 TV lines. Element compression ratio, 7.379; bit
compression ratio, 3.705.
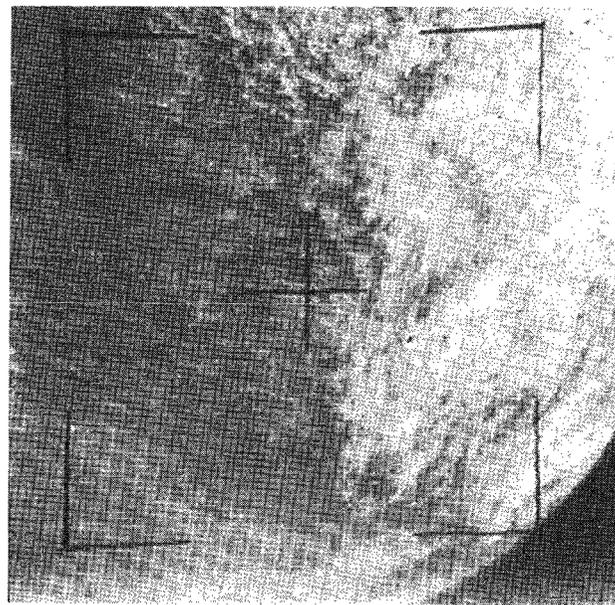
(d) Processed copy generated by conditional expecta-
tion predictor with neighborhood and two-dimensional
predictors. Q=6 bits per TV element; T=± 2 levels;
L = 16 TV lines. Element compression ratio, 5.913; bit
compression ratio, 3.593.

Figure 17 - Pictures from Tiros III, orbit 102, frame 2, camera 1; taped before transmission from
satellite; principal point, 13.3N, 5.6W; subsatellite point, 11.9N, 1.9W.

(a) Analog original.
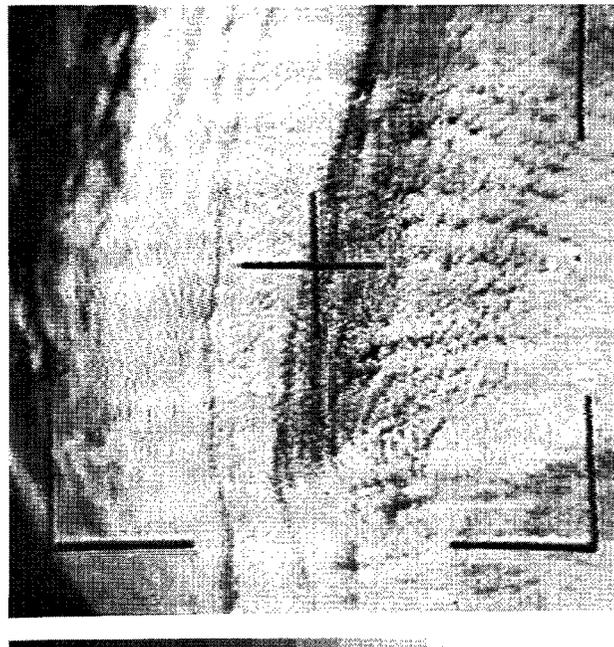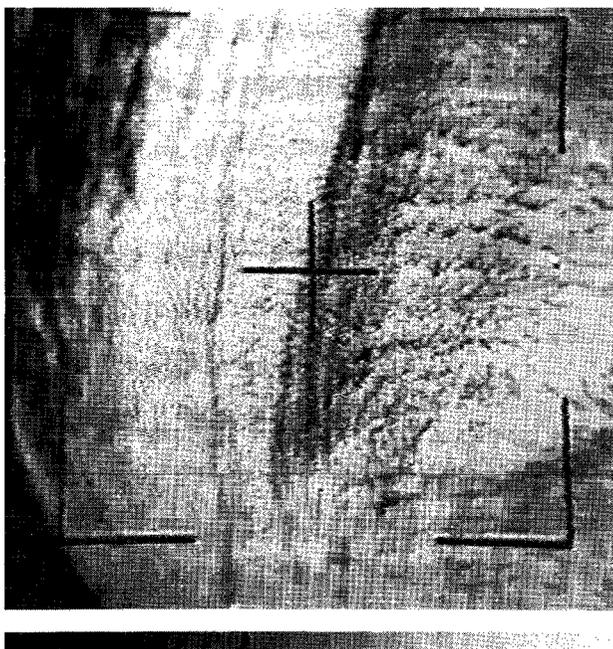
(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L = 16 TV lines. Element compression ratio, 7.824; bit compression ratio, 3.762.
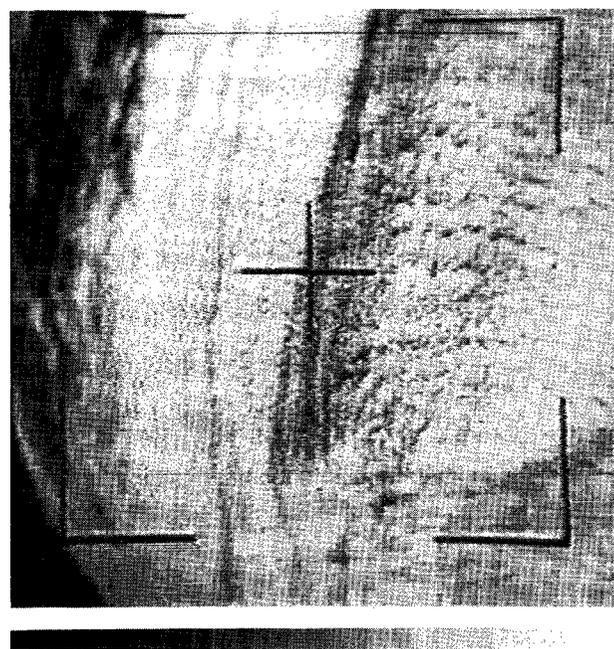
(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L = 16 TV lines. Element compression ratio, 6.863; bit compression ratio, 4.064.

Figure 18 - Pictures from Tiros V, orbit 3143, frame 6, camera 1; direct transmission from satellite; principal point, 32.4N, 69.3W; subsatellite point, 33.9N, 73.4W.
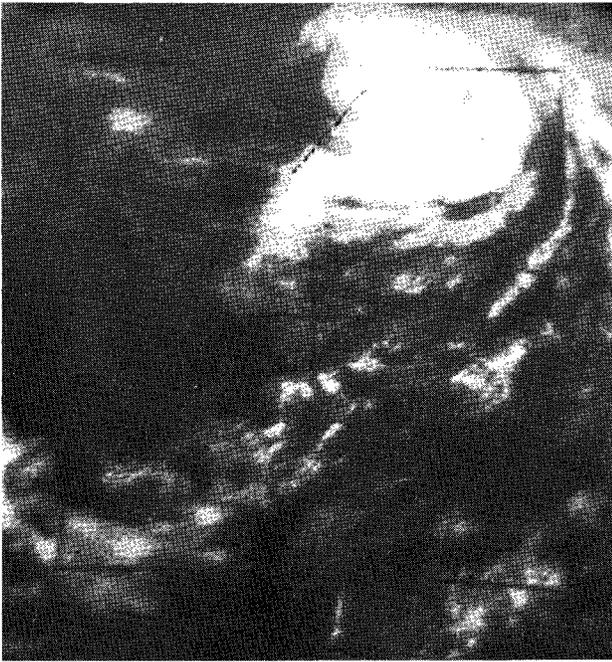
(a) Analog original.
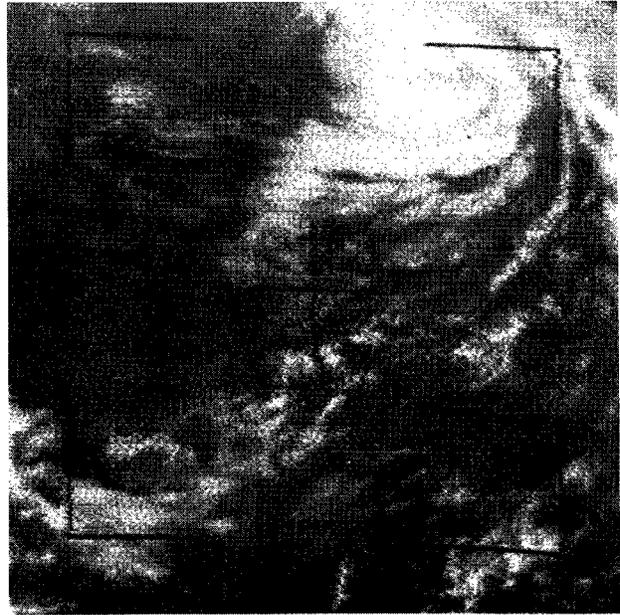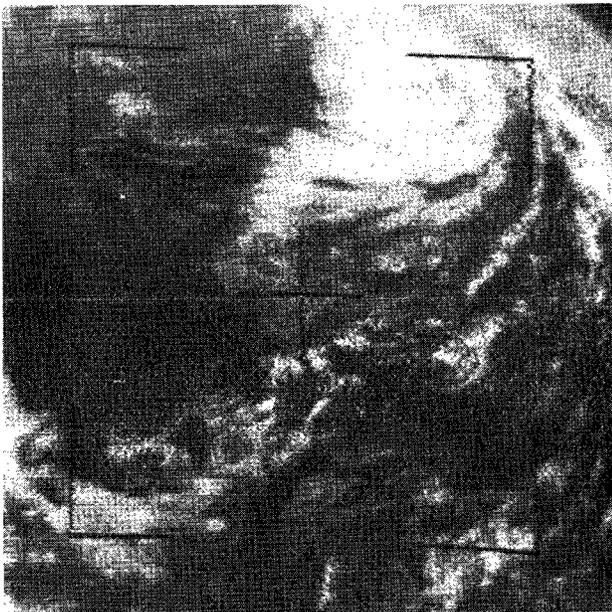
(b) Digital original.

(c) Processed copy generated by conditional expecta-
tion predictor with neighborhood and two-dimensional
predictors. Q=4 bits per TV element; T=± 1 level;
L=16 TV lines. Element compression ratio, 3.890; bit
compression ratio, 2.160.
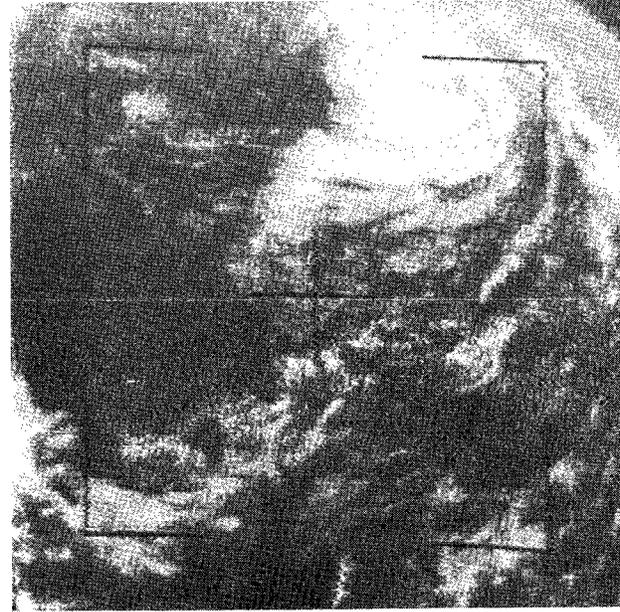
(d) Processed copy generated by conditional expecta-
tion predictor with neighborhood and two-dimensional
predictors. Q=6 bits per TV element; T=± 2 levels;
L=16 TV lines. Element compression ratio, 3.297; bit
compression ratio, 2.218.

Figure 19 - Pictures from Tiros VI, orbit 1100, frame 15, camera 1; direct transmission from satellite;
principal point, 28.3N, 79.0W; subsatellite point, 26.1N, 79.8W.
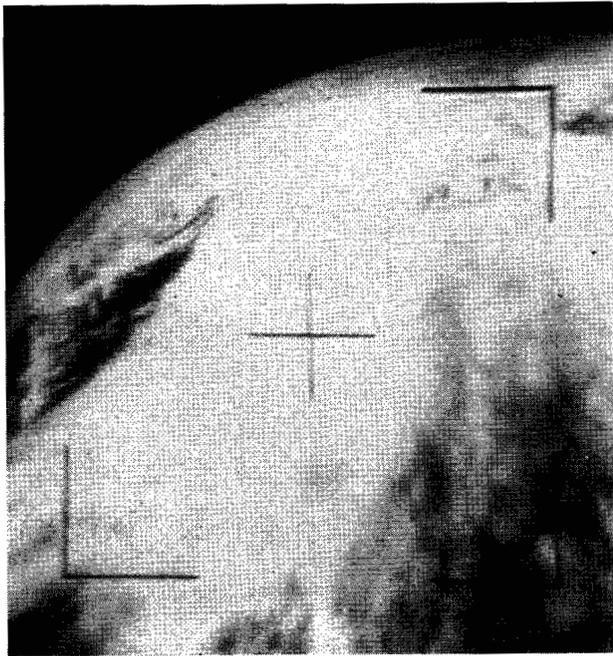
(a) Analog original.
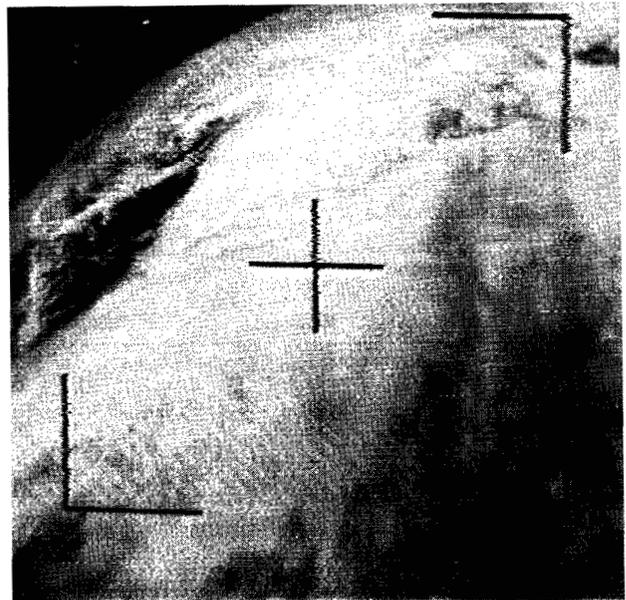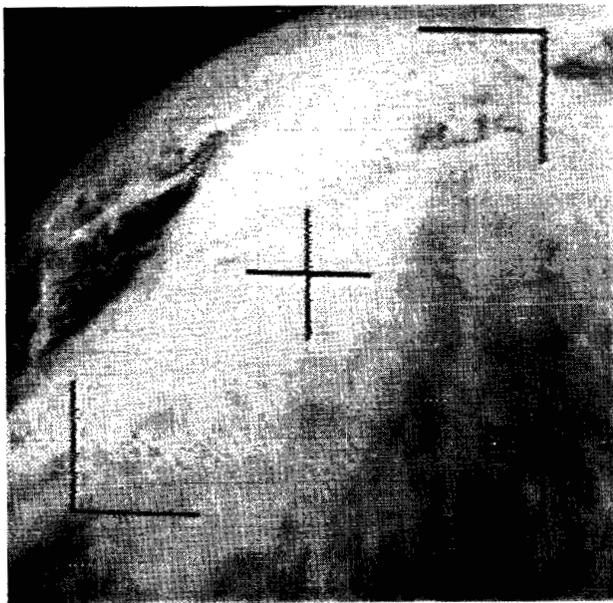
(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. $Q=4$ bits per TV element; $T=\pm 1$ level; $L=16$ TV lines. Element compression ratio, 9.690; bit compression ratio, 4.473.
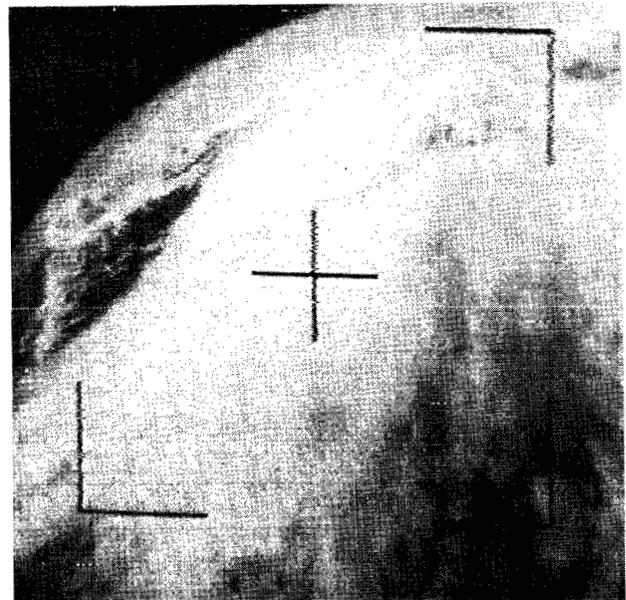
(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. $Q=6$ bits per TV element; $T=\pm 2$ levels; $L=16$ TV lines. Element compression ratio, 7.875; bit compression ratio, 4.576.

Figure 20 - Pictures from Tiros VI, orbit 18, frame 21, camera 1; taped before transmission from satellite; principal point, 52.5N, 45.2W; subsatellite point, 50.4N, 37.3W.
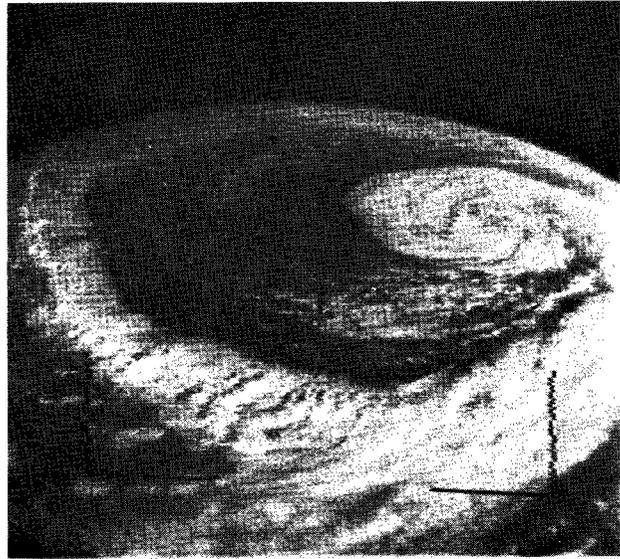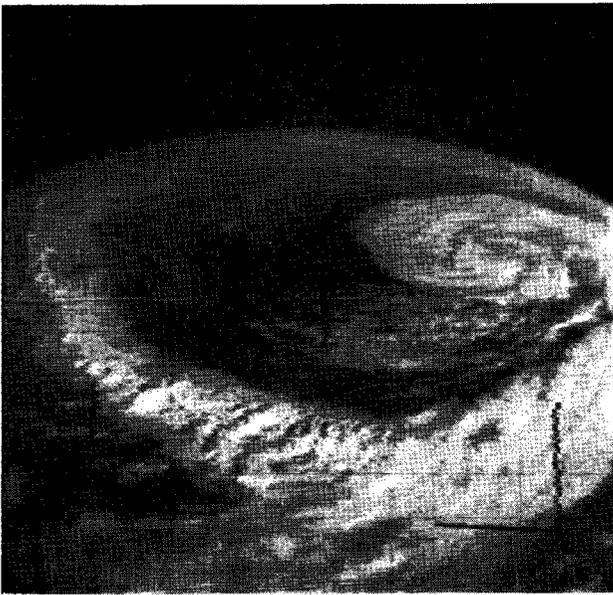
26

(a) Analog original.

(b) Digital original.

(c) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=4 bits per TV element; T=± 1 level; L=16 TV lines. Element compression ratio, 19.954; bit compression ratio, 8.219.
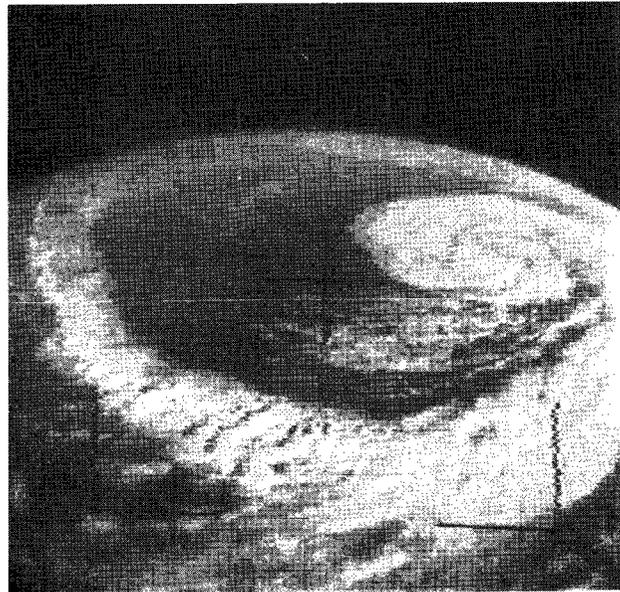
(d) Processed copy generated by conditional expectation predictor with neighborhood and two-dimensional predictors. Q=6 bits per TV element; T=± 2 levels; L=16 TV lines. Element compression ratio, 17.550; bit compression ratio, 9.127.

Figure 21 - Pictures from Tiros VI, orbit 3692, frame 31, camera 1; taped before transmission from satellite; principal point, 36.8N, 57.2W; subsatellite point, 33.1N, 48.7W.

APPLICATIONS FOR DATA COMPRESSION SYSTEMS

Some of the most obvious applications of data compression systems are in deep space communications, earth-orbiting operational spacecraft, and land-line data transmission. Figure 22 is a block diagram of both the transmitter and receiver ends of a data compression system model. At the transmitter end, the predictor accepts raw data from the information source. The predictor contains arithmetic, memory, and control functions which are arranged according to some prediction algorithm. Each raw data sample is compared to the corresponding predicted sample, and the prediction error E_p is determined. If the prediction error exceeds some preset threshold T, the raw data sample must be transmitted in unmodified form. If, however, the prediction error is less than T, the sample is predictable and need not be transmitted. The comparator output is also fed back to the predictor to update the prediction mechanism. The encoder accepts raw unpredictable samples as well as indications of predictable samples and arranges this information according to some appropriate code. The information rate at the output of the encoder is, in general, nonuniform. Since the main data-storage device would probably require a uniform read-in rate, a smoothing buffer is necessary.

At the receiver end, the decoder provides the predictor with all the data necessary to reconstruct the original message within the allowable prediction error. The predictor at the receiver is an exact copy of the predictor at the transmitter. After reconstruction, the message is transferred to the information sink.

Bit compression ratio can be a very useful parameter to a communications system designer. If C_B represents the bit compression ratio, then the designer can choose to reduce the transmission
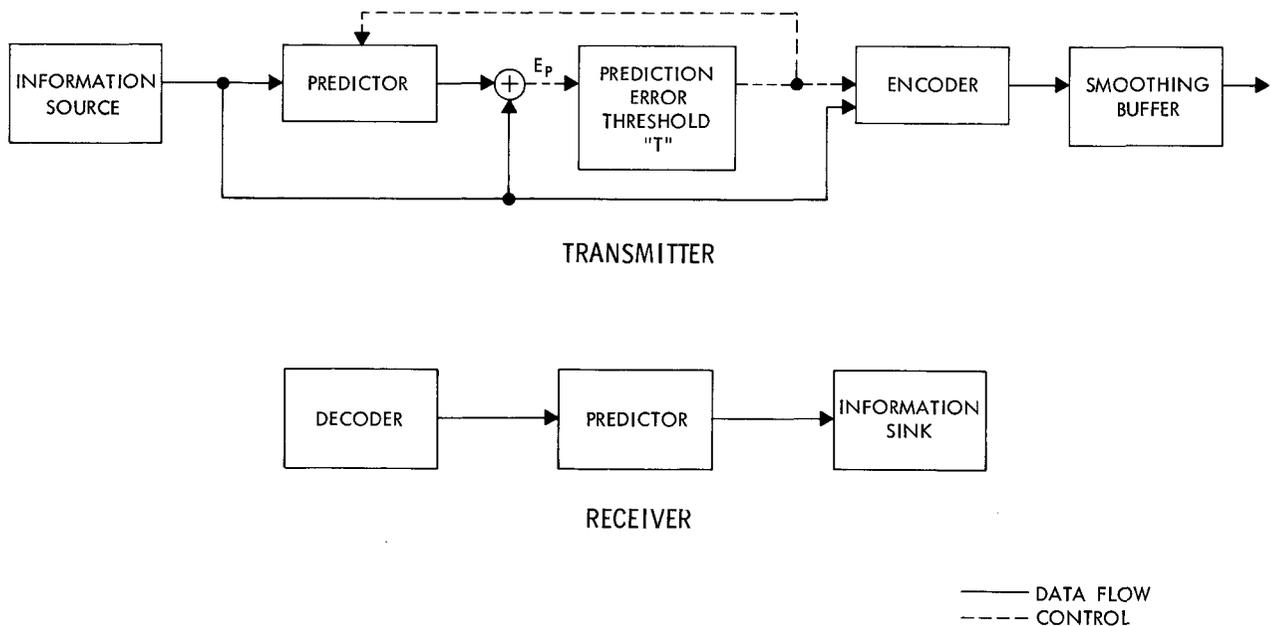
Figure 22—Block diagram of data compression system.

time to T/C_B for the same bandwidth or alternatively reduce the original bandwidth to B_W/C_B. If one desires to save power or reduce spacecraft weight by saving power, the signal power can be reduced by S/C_B without changing the S/N ratio, since the thermal noise is directly proportional to the bandwidth. In practice, however, one would probably choose to employ data compression techniques to achieve high communication channel efficiency. This can be achieved by keeping the information rate close to the channel capacity at all times. This implies a channel with the capability of adapting to the time-varying information rate.

CONCLUDING REMARKS

The most important outcome of this work was that the conditional expectation predictor produced compression ratios greater than either the zero-order hold or the linear predictors. This result is true for each of the 10 TV frames used in the study and is significant since it shows that the conditional expectation predictor yields superior compression ratios, despite the suspected nonstationary character of the information source. To summarize the numerical results, it should be noted that the conditional expectation predictor produced bit compression ratios (assuming ideal coding in the noiseless case) exceeding 5:1 on a number of single TV frames. It is also important that the cumulative compression ratio (10-picture average) exceeded 4:1 for both the cases with 6 and those with 4 bits per TV element. At the same time the compressed pictures (Figures 12 to 21) seem to retain at least an acceptable level of quality.

Many interesting problems associated with adaptive data compression systems require further investigation. The prediction mechanism itself should be further developed to include, for example, the optimal relationship between learning period and memory size. Certainly the determination of efficient coding schemes for the adaptive data compression system for the noiseless channel is the most important problem still to be solved.

Further investigations might also consist of simulating noise environments for possible missions and analyzing effects on the noiseless-case coding structure in order to develop efficient error-correction codes. Investigations of this sort would eventually allow laboratory simulation of a complete compression system from the information source to the transmitter, through the communication channel to the receiver, and finally to the information sink. This arrangement would permit feasibility studies for specific missions as well as establish system design guidelines.

ACKNOWLEDGMENTS

REFERENCES

1. Weber, D., "A Synopsis on Data Compression," Proc. 1965 National Telemetering Conf., The Institute of Radio Engineers, New York, N. Y., 1965.

2. Ellersick, F. W., "Data Compactors for Space Vehicles," Proc. Eighth Annual East Coast Conference on Aerospace and Navigational Electronics (October 23-25, 1961, Baltimore Md.), North Hollywood, Calif.: Western Periodicals Pub. Co., 1961, pp. 6.2.6-1—6.2.6-5.

3. Balakrishnan, A. V., "An Adaptive Nonlinear Data Predicator," Proc. National Telemetering Conf. (May 23-25, 1962, Washington, D. C.), The Institute of Radio Engineers, New York, N. Y., 1962, Vol. II, Suppl.

4. Balakrishnan, A. V., Kutz, R. L., and Stampfl, R. A., "Adaptive Data Compression for Video Signals," NASA Technical Note D-3395.

5. Davisson, L. D., "Data Compression and its Application to Video Signals," in: *Final Report of the Goddard Space Flight Center Summer Workshop Program in the Analysis of Space Data and Measurement of Space Environments,* GSFC Document X-100-65-407, June 15 to September 15, 1965, pp. A-35—A-55.

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Technical information generated in connection with a NASA contract or grant and released under NASA auspices.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

TECHNICAL REPRINTS: Information derived from NASA activities and initially published in the form of journal articles.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities but not necessarily reporting the results of individual NASA-programmed scientific efforts. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546