

NASA TECHNICAL NOTE



NASA TN D-3395

21

LOAN COPY: RETURN TO
AFWL (WLIL-2)
KIRTLAND AFB, N MEX

NASA TN D-3395

ADAPTIVE DATA COMPRESSION FOR VIDEO SIGNALS

by

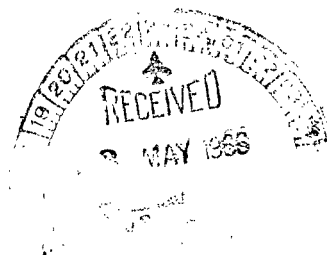
A. V. Balakrishnan

University of California at Los Angeles

and

R. L. Kutz and R. A. Stampfl

Goddard Space Flight Center



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • APRIL 1966

ADAPTIVE DATA COMPRESSION FOR VIDEO SIGNALS

By A. V. Balakrishnan

University of California at Los Angeles
Los Angeles, Calif.

R. L. Kutz and R. A. Stampfl

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 - Price \$2.00

ABSTRACT

The properties of adaptive predictors are discussed and a brief derivation of the prediction mathematics is given. The ten Tiros television cloud pictures used in the study are described and the method used to display them on the Stromberg Carlson 4020 microfilm recorder is discussed. A detailed explanation of how the technique is applied to television pictures and how it is programmed on the computer is given. The result of the compression on the quality of the TV pictures is discussed and is shown by the pictures displayed herein.

CONTENTS

Abstract	ii
INTRODUCTION	1
PROPERTIES OF ADAPTIVE PREDICTORS	2
THEORY OF ADAPTIVE COMPRESSION SYSTEM	3
TIROS VIDEO DATA	8
COMPUTER IMPLEMENTATION	11
DISCUSSION OF RESULTS	18
CONCLUSIONS	22
ACKNOWLEDGMENTS	22
References	22
Appendix A—Symbols	43

ADAPTIVE DATA COMPRESSION FOR VIDEO SIGNALS

by

A. V. Balakrishnan

University of California

R. L. Kutz and R. A. Stampfl

Goddard Space Flight Center

INTRODUCTION

The spin-stabilized Tiros and Tos weather satellites produce, along with other data, large quantities of television pictures. The two television cameras on board can televise directly when within communication range of a ground station, and, more importantly, they can store 32 TV frames per camera on magnetic tape and transmit this information when an appropriate ground command is given. A 62.5-kc video bandwidth is employed for the transmission of 500 line frames, each with a frame duration of 2 seconds. The gray scale quality of the pictures varies between 5 and 10 levels as a function of camera tube life and telemetry performance; 6 levels can be considered an average quality value. *

When pictures are received at ground stations, a composite mosaic consisting of overlapping individual frames is prepared from the 32 picture series with the help of the satellite attitude data available from other data channels. Meteorologists at the receiving sites transpose the meteorological information content onto weather maps by visual inspection. Cloudiness of the area viewed, types of clouds, and interpretive results such as fronts and storm centers are all noted on maps; the results are conventional nephanalyses. These nephanalyses are transmitted via facsimile links to forecasting centers or are translated into teletype messages for distribution. A sample of a mosaic and the corresponding nephanalysis and teletype message are given in Figure 1.

The glaring difference between the number of data points contained in 32 pictures or in a prepared mosaic and the true meteorological information content of the mosaic stresses the need for research aimed at practical data compression methods. Fully operational satellites delivering data continuously by covering the entire earth every 12 hours will benefit even more from the results of such studies. The present paper reports on the method employed to compress video data by applying adaptive prediction of the kind first reported by Balakrishnan (Reference 1) to a sample of ten Tiros cloud pictures. The choice of this rather elaborate method using computer implementation rather than another approach was made after a brief investigation showed that trivial approaches will yield only meager data compression. Delta modulation, a simple technique in which the amplitude difference between adjacent samples is transmitted and the sampling rate is chosen such that no more than 1 quantum step difference is expected, is objectionable because resolution is lost on high-contrast transitions. An implementation of the delta modulation method has

**SAMPLE
MISCELLANEOUS SATELLITE BULLETIN**

22/1125Z E 17.N, 57.5W, STAGE X,
DIA 5, BANDS 2, TROPICAL STORM
BEULAH, EYE NOT VISIBLE, STORM
CENTER LOCATED WITHIN BROAD
CLOUD BAND TIP AT 18N, 58W.

25/1315Z AUG GWBC

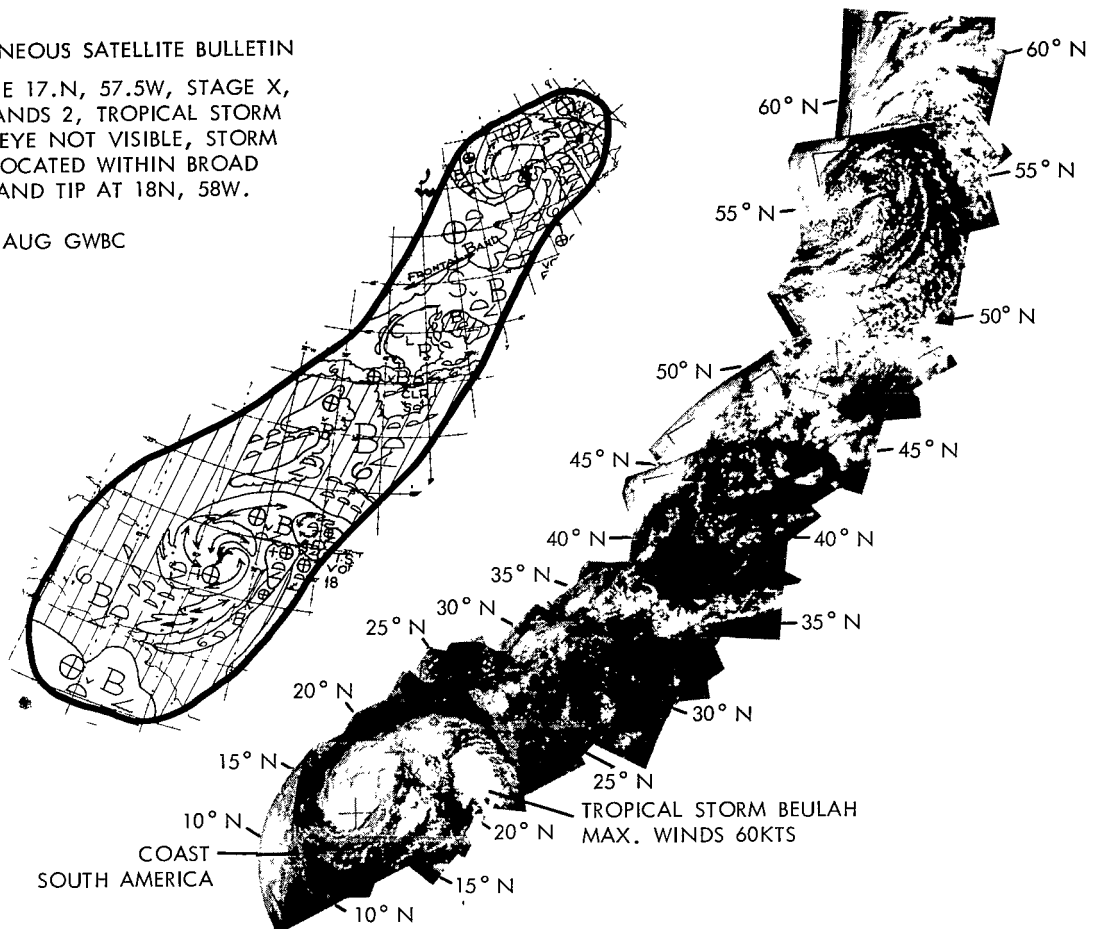


Figure 1—Sample of a mosaic and the corresponding nephanalysis and teletype message.

the most attractive circuitry because of its simplicity but visual inspection of cloud pictures confirms objectionable loss in picture resolution. Even though pictures can be transmitted by a four-bit code because of the limited number of gray levels, the total number of bits per frame is seldom reduced more than two to one when delta modulation is used. Run-length coding, a technique where the number of consecutive elements with constant attributes are counted, has also been investigated; there were no conclusive results but the need was shown for computer simulation so that a variety of methods can be tried with minimum investment.

PROPERTIES OF ADAPTIVE PREDICTORS

The technique used in this study is adaptive and has been presented rigorously in the literature (see Reference 1), so an analysis of the basic features of the system shall suffice. An adaptive

system is distinguished primarily by an explicit or implicit learning mode; however, to be truly adaptive the system must have other features as well.

At this point it is well to present some definitions: *Adaptive*: Implies modification to meet new conditions. *Adjusting*: Implies a bringing into as exact or close a correspondence as exists between the parts of a mechanism, but suggests more tact or more ingenuity in the agent. "Adjusting" appears to be a more exact description of the system than "adaptive," but the word "adaptive" has become general usage. Actually, it is impossible to draw a sharp distinction between what is adaptive and what is not. The word adaptive does signify something more than a self-adjusting mechanism (such as, say, an AGC or AFC system where there are no conditions not envisaged at the time of design). Without getting involved in semantics, one can say that an adaptive system is one which monitors its own performance; the system learns what the new conditions are and changes its structure accordingly. (In a biological system where the adaptivity is more easily described, the performance criterion would be self-preservation.)

In a communication system the new conditions consist of parameters changing in some *a priori* unknowable fashion. These parameters have to be learned or measured and appropriate changes must be made in the system to maintain a prescribed level of fidelity, or some other chosen criteria. A simple example is the usual Shannon-Fano noiseless coding for minimal length, which requires complete specification of the message probabilities. In many cases (strictly speaking, in all cases) these probabilities are unknown and have to be measured. Moreover, if the probabilities are also subject to change, we need a learning phase, whether it is or is not separate from the coding phase. One cannot call such a system "adaptive" since there is no monitoring of performance involved. In a two-way or feedback communication system (such as a data transmission system) the receiver monitors quality of performance (with an error-detecting code) and the transmitter adjusts by reducing the transmission rate or stopping transmission altogether. There is no learning involved here so this system is not quite adaptive in the sense in which we construe it. The learning phase is triggered by a degradation in performance. Now a radar system which simply measures, say, the clutter or reverberation parameters and then changes the detection circuitry parameters is not quite adaptive since there is not a self-monitoring criterion. If, for example, the transmission system learns or measures channel characteristics and changes its mode of transmission accordingly to restore the performance quality, it would be a truly adaptive system.

Actually, the term "adaptive" is beginning to be used whenever there is a learning or measurement feature (implicit or explicit) even if there is no self-monitoring. (The terms *self-regulating* and *self-adjusting* are used for systems with the monitoring feature but with no learning feature.)

THEORY OF ADAPTIVE COMPRESSION SYSTEM

Note that the compression system is truly adaptive in the sense that it features a self-monitoring function in addition to the learning and adjusting functions.

Let $x(t)$ represent a continuous or discrete parameter source. It is customary to assume that $x(t)$ can then be regarded (at least for analytical purposes) as a stochastic process. If the

time parameter t is continuous, it is often possible to assume that

$$x(t) = \int_{-B}^B e^{2\pi i f t} d\psi(f)$$

in a suitable sense (that is, depending on whether we adopt the stochastic or nonstochastic viewpoint) for sufficiently large values of B , the bandwidth. By use of the well known "sampling principle" one can then represent the continuous waveform using periodic samples taken at $t = n/2B$. In most communication systems using sampled data of this kind, the sampling rate is determined by the nominal highest or cut-off frequency B expected in the data. However, in the case of many kinds of data—such as space-telemetry data—the actual cut-off frequency will be much smaller for most of the time, so that there is no need to sample at the nominal rate of $2B$. In other words, it is possible to "compress" the data sampling rate or channel bandwidth.

A method of achieving such compression in PCM systems is to exploit the redundancy or predictability of the data. Thus, we employ a "predicator" at the transmitter which predicts the data at time $t + \Delta$ based on the past data up to time t . The actual value sampled is then compared with the predicted value. If the difference exceeds (in absolute value) a preset threshold, then the actual sample is transmitted. If it is below the threshold, a prearranged code word using say, one digit is transmitted instead of the m digits, thus reducing the number of digits transmitted per second. A comma-free code may be used to sort out the two kinds of words unambiguously. The adaptive or learning feature comes in the predictor mechanism. The predictor is not updated until its threshold is exceeded; this exhibits the self-monitoring feature. The predictor itself is based on a learning phase, and the prediction operator is adjusted accordingly. For details on the predictor itself see Reference 1. The significance of the adaptive prediction feature lies in the fact that no *a priori* statistics or other assumptions concerning the data are required, and, in particular, it is realized that there may be periods in the data where prediction (to the quality set) may not be possible.

The basic prediction philosophy may be indicated briefly. We are given a waveform of duration T (i.e., T is a function $x(t)$, $0 \leq t \leq T$) and we are required to predict the value at time $T + \Delta$. Any prediction operation is to be based on the data alone, no additional *a priori* knowledge being available. This is, of course, an ancient problem, but here we wish to treat it strictly in the telemetry data processing context and to note two major methods of dealing with it. One, which may be considered the numerical-analysis method, consists in assuming that certainly any physically realized waveform must be analytic and can thus be approximated by polynomials. The data may thus be "fitted" to a polynomial of high enough degree and we then simply use this polynomial to predict the future values. The other and more recent method is the statistical view in which we assume (perhaps with good reason) that the data are finite samples of a stationary stochastic process; the process' average properties, such as moments and/or distributions, are known or calculable from the data. For a process of given description we can apply the well-developed mean-square prediction theory. Perhaps the main advantage of this method is that it gives us a

quantitative, albeit theoretical, notion of the error in prediction. The problem of measuring the average statistics from a finite sample can be quite delicate, however. In the polynomial fitting method, the interpretation of the prediction error—which is, after all, the crucial point—is more nebulous and is bound up with the degree polynomial to be used and the portion of the data to be fitted. Moreover, as ordinarily used, the fitting operations on the data are linear.

We adopt a rationale for prediction which is free from *a priori* assumptions concerning the "model." In a general sense what is involved in both the above methods is, first, "model-making" consistent with the data and, second, using the numbers derived therefrom to perform some optimal operations. If an understanding of the mechanism generating the model is desired, the first step is essential. If prediction is the object, it is possible to proceed directly (and hence more optimally) to the best prediction without the intermediate step of model-making. It may be, of course, that several approaches lead to the same operations on the data; however, the present method offers some practical advantages and it is desirable to use the approach that requires the fewest prior assumptions.

We note first that any prediction is an operation or operator on part of the data, and in our case, the finite part is all that is available. The main point of departure in our view is that if we have a prediction operator, which based on all the available input data, functions optimally in the immediate past of the point where the prediction is required, this is all that we can ask for meaningfully as a solution to the prediction problem. Thus, the only basis on which we can *a priori* judge any prediction method is to "back off" slightly from the present and compare the actual available data with the predicted value using the given prediction operator.

Let us see how this can be formulated analytically. Let the total available data be described as a function $x(t)$, where $0 \leq t \leq T$, and let the prediction of the value at $T + \Delta$, (here $\Delta > 0$, being a small fraction of T) be required. Let us next consider the data in the interval $0 < t < (T - \Delta)$. If for any t_0 in this interval we should choose to make a prediction of the function value Δ ahead from the past values up to t_0 and denote the predicted value by

$$x^*(t_0 + \Delta)$$

we can explicitly observe the error

$$x^*(t_0 + \Delta) - x(t_0 + \Delta).$$

Let us next note that the general prediction operator will be a function of a finite segment of the past. Let us denote the length of this segment by M . Then, of course,

$$x^*(t_0 + \Delta) = O \{ x(\sigma) : t_0 - M < \sigma < t_0 \},$$

where O represents the prediction operator.

Next we have to specify the error criterion. We choose the mean square error because it is both simpler analytically and almost universally used and so comparison with other methods is possible. The kind of solution presented is definitely not analytic, but, rather, is based on successive approximation so other measures of error can be used at the risk of greater complexity. As far as the rationale of the method is concerned, this is largely a matter of detail rather than principle. We can thus use to determine optimality:

$$\frac{1}{L-M} \int_{T-\Delta-L+M}^{T-\Delta} [x^*(t+\Delta) - x(t+\Delta)]^2 dt = \epsilon^2 \quad (3)$$

where $M \leq L$, and we proceed on the basis that the operator O is best which minimizes Equation 3. We can, of course, generalize Equation 3 as

$$\frac{1}{L-M} \int_{T-\Delta-L+M}^{T-\Delta} \rho(t) C[x^*(t+\Delta) - x(t+\Delta)] dt, \quad (4)$$

where M is the domain of O for a prediction at t_0 , $\rho(t)$ is a positive weight function, and the function C is, say, a symmetric positive cost function (a complete list of symbols is given as Appendix A). Before we consider the problem of determining the optimal operator O , let us note an important consistency principle. Suppose the data are regarded as one long sample of an ergodic process. Then Equation 3 yields exactly the optimal operator in the statistical sense. However, we have not needed to make any such assumption concerning the data; nor have we had to compute average statistics first. The point is that while the data may not be enough for determining, let us say, the spectrum, it may be quite adequate for the prediction itself. Unlike the polynomial fitting, the operations on the data can be as nonlinear as necessary, and at the same time Equation 3 normalized to

$$\epsilon^2 / \left\{ \frac{1}{L-M} \int_{T-\Delta-L+M}^{T-\Delta} [x(t+\Delta)]^2 dt \right\} \quad (5)$$

yields a quantitative measure of the prediction error on which to judge how good the prediction will be. For actual methods of determination of the optimum predictor see Reference 1.

So far we have determined the optimal adaptive structure, but there still remains the problem of evaluating the system. For instance, it is natural to ask what is the average amount of compression it is possible to obtain in this way. Here we have to postulate some structure for the data source and then calculate the resulting compression using the adaptive system. We shall now briefly indicate what an analysis of this type involves. Not to complicate the analysis unduly, let us consider the prediction operation in the transmitter based only on the actual samples. Let us denote the data samples by x_n . We assume that the data can be taken as a stationary stochastic process. We may further assume that the process is Gaussian since the maximum prediction error

(and hence the minimum compression) occurs in this case because the prediction operation includes only the linear case. Let us consider the case where the adaptive predictor is also constrained to be linear. In this case the optimal filter weights α_j are determined by minimizing

$$\frac{1}{L-M} \sum_{n=-L+1+M}^0 \left(x_n - \sum_{j=1}^M \alpha_j x_{n-j} \right)^2 \quad (6)$$

where the past available data consist of L samples and the prediction is based on M samples. We are, of course, considering the analog method above. The corresponding (squared) error is then

$$\epsilon_0^2 [M; L] = \left(x_1 - \sum_{j=1}^M \alpha_j x_{1-j} \right)^2. \quad (7)$$

The transmission is based on Equation 7 exceeding a threshold τ_1 . Hence, what we want first is to determine the statistics of Equation 7. We note first that the optimal $\{\alpha_k\}$ that minimizes expression (6) will satisfy:

$$\sum_{n=-L+1+M}^0 x_n x_{n-k} = \sum_{j=1}^M \sum_{n=-L+1+M}^0 \alpha_j x_{n-j} x_{n-k}, \quad k = 1, \dots, M.$$

Let Y_n be the vector

$$[x_n, x_{n-1}, \dots, x_{n-L+1+M}].$$

Then the minimum of expression 6 becomes

$$\frac{1}{L-M} \frac{D_{M+1}}{D_M},$$

where D_{M+1} is the determinant of the $M+1$ by $M+1$ matrix with entries

$$Y_i \cdot Y_j, \quad i, j = 0, -1, \dots, -M$$

and D_M is the determinant of the M -by- M matrix with entries

$$Y_i \cdot Y_j, \quad i, j = -1, -2, \dots, -M.$$

On the other hand, we are interested in the error (7), which is

$$\epsilon_0^2 [M; L] = \left(\frac{D'_{M+1}}{D_M} \right)^2 \quad (8)$$

where D'_{M+1} has first row

$$x_1, x_0, x_{-1}, \dots, x_{-M+1}$$

and is otherwise the same as D_{M+1} . Our first interest is in the statistics of Equation 8. We need to know

$$P_r \left[\epsilon_0^2 (M; L) \geq T_1 \right],$$

which is the probability of exceedence of the threshold and relates directly to the attainable compression. Actually Equation 8 determines only the "open-loop" error, and to evaluate the complete system we have to consider the "closed-loop" operation where actual samples are replaced by predicted samples when the former fall within the set threshold.

Equally important, we need to know what the quality of the compressed data will be. One may, for instance, use the cross-correlation coefficient or another equivalent quantitative index and evaluate this as a function of T_1 , M , and L . A complete system evaluation requires that channel noise be also taken into account. Such an analytic evaluation is necessary to provide a logical basis for choosing optimum design parameters. Indeed, if the channel noise characteristics are known, appropriate coding or signal selection has to be considered in arriving at the total compression possible.

TIROS VIDEO DATA

Many of the thoughts expressed here go beyond the scope of the present investigation, an application to Tiros television. It is reasonable to expect remarkable future progress in applications of the theory when one considers the recent progress in the field.

Application of the technique to Tiros TV data requires appropriate conversion of the original analog magnetic tapes to computer compatible tapes. This data compression technique was chosen for investigation instead of another because of the high expectation that other techniques could do no better and because this technique permits comparison with a nearly ideal situation. In applying the method, the three picture elements preceding the data point to be predicted are usually used. They are called the memory M in this report. A weighting function is derived from a chosen sample of L picture elements, here referred to as the learning period (see Figure 2). The learning period $L = 20$ was used for many of the computer runs; however, other intervals can be chosen and data compression has been attempted with $M = 3$ and $L = 20$, $M = 0$ and $L = 20$, and $M = 0$ and $L = 1$.

Video data for this study were obtained from the standard Tiros FM subcarrier modulated magnetic tapes; the data were converted by use of the A/D converter to six-bit characters and were recorded on magnetic tape at a density of 556 characters per inch in one physical record of 40,000 words with 36 bits per word.

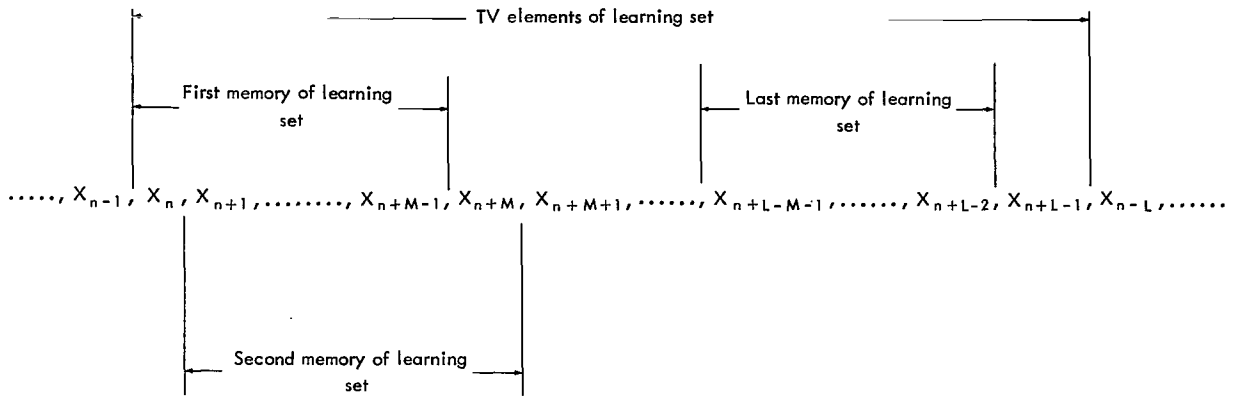


Figure 2—Relation of learning and memory sets.

The video input to the A/D converter included horizontal sync pulses to provide line sync in the conventional analog Tiros display system. A program for the IBM 7094 computer read each 40,000-word record in two passes, since the core capacity is only 32,000 words, found line sync, and reformatted the data into 24 records of 20 lines each. Each line of TV data now contained 468 elements with 480 lines in each TV cloud picture; this was a reasonably good match to the 500-element, 500-line analog original pictures. The reformatting and sync program for the 7094 did not find the start of each TV line with perfect consistency and for this reason some of the TV lines were registered incorrectly, as is evident by inspection of the TV cloud pictures used in the study. The dynamic range of the analog data entering the A/D converter was adjusted so that it was smaller than a six-bit code would permit. A tolerance was allowed at the minimum and maximum amplitudes of the analog video data so that only approximately 32 of the possible 64 amplitude levels were used in the converter. The TV data used in this study should therefore be considered five-bit data.

A number of typical Tiros pictures (Figures 6 to 15) have been selected; these pictures show a variety of cloud patterns so that individual interpretation after processing permits general judgment as to acceptable quality. Pictures from both Tiros operating modes have been analyzed, that is, those recorded in the satellite and played back later and also those transmitted directly when within communication range. The latter pictures are preferred because they show slightly better quality than do the taped ones.

The first four pictures (Figures 6(a) to 15(a)) are part of a direct camera sequence of Tiros III, orbit 4, when the satellite was moving east-south-east over the central U.S. They were taken on July 12, 1961, and the picture principal point, which is near the cross fiducial mark, lies in the SW corner of Minnesota in Figure 6(a) and in North Central Iowa in Figure 9(a). The data log indicates for this orbit that the satellite camera was looking in a northwesterly direction. Because of the spin of the satellite the north direction may be different for each picture. Visualization of the geometry is simplified by the fact that the spin axis was practically in the orbital plane on the first day when orbit 4 occurred. Heavy cloud masses in a very distinct pattern can be seen, as well as large clear areas and regions of scattered overcast. The brightness levels viewed by Tiros cover the entire range of cloud cover intensities from black to full brightness.

The next two pictures (Figures 10(a) and 11(a)) were obtained again from Tiros III but are in the tape mode and are successive frames from orbit 102. The principal point lies near the Ivory Coast of Africa in Figure 10(a) and Liberia, Africa, in Figure 11(a). They were taken approximately 1 week after the preceding series. These two pictures were selected because of the random nature of the cloud pattern. It is also interesting to note the strong effect of viewing angle under the same sun illumination by comparing the two adjacent frames. The next picture processed (Figure 12(a)) was taken by Tiros V on orbit 3143 in the direct mode after 219 days orbital camera life in late January 1963. It shows distinct streakiness in the clouds grouped into two large sections in the picture. The clear long banded areas are also distinct features.

The remaining three pictures included in this study (Figures 13(a) to 15(a)) were taken by Tiros VI in December 1962, September 1962, and May 1963, respectively. Figure 13(a), in the direct mode, shows a tropical storm under oblique view in the Atlantic off the coast of Miami; Figure 14(a), in the tape mode, shows a large cloud area over the Atlantic located south of Greenland; and Figure 15(a), in the tape mode, shows a mature tropical storm located in the Atlantic northeast of the Bermuda Islands.

The digitized pictures (Figures 6(b) to 15(b)) were reproduced as four-bit encoded signals because of the limited dynamic range of the Vidicon and the display. As was to be expected, no noticeable difference can be detected. Reproduction of the pictures was accomplished on a Stromberg Carlson 4020 microfilm recorder. The input for this machine came from digital magnetic tape produced as an output of the data-compression programs. This machine has the capability of plotting symbols on a 1024×1024 grid with 16 different intensities including the zero or no-plot intensity (see the scale at the bottoms of the figures). The plotting symbol used here was the dot. This symbol requires the least area and does not overlap if odd rows and columns in the plotting grid are skipped. Thus, a 512×512 grid is available for picture plotting, gray scale plotting, and the printing of identification marks. The fifteen intensities which cause darkening of the paper are produced by unblanking the exposure beam from one to fifteen counts, in correspondence with a four-bit control word. This method of exposure does not produce a gray scale which is linear to the eye.

In order to produce a gray scale which would appear approximately linear, a power rule for the exposure time was used. Although a standard gray wedge has intensities which change by a ratio of $\sqrt{2}$, a smaller change is still easily seen by the eye. An exposure ratio of 1.25:1 was used here. In this manner a gray wedge was produced by starting with an exposure time of four counts and increasing the exposure time in steps corresponding to a ratio of 1.25:1 until 91 counts were reached. This provides fifteen gray steps with differences in intensity which look equal to the eye. The SC-4020 uses a lens with a variable diaphragm controlled by the f setting to project the CRT image onto the display medium. It was determined by eye that $f16$ produced the same shade change between the background and the first step, which is a count of four, as the shade change produced by the 1.25 exposure time rule. Another factor controlling the choice of the f setting was that below $f16$ the range of the display medium would have limited the gray scale on the dark end.

Sixteen levels of gray changing in steps which look equal to the eye were produced in this manner to correspond to four-bit data. When five-bit data were displayed the least significant bit was dropped. The pictures which appear in this report probably will have lost some of the linearity of the gray scale because of the reproduction process but the overall quality should not be degraded because of the large number of gray shades present. The difference in quality between the analog original copies and the digitally reproduced versions is quite pronounced and merely a consequence of the particular properties of the equipment used. Better reproduction techniques can readily be designed when specific methods are to be implemented. Contouring due to the finite number of quantum steps is visible but less objectionable than the appearance of the reproduction process. Superposition of random noise can eliminate the contouring and thereby cope with this psychological objection.

COMPUTER IMPLEMENTATION

In the practical implementation of the method previously described, it is necessary to generate a set of weights which will produce the best estimate of the most future data samples when operating on past data samples. In an actual telemetry system only the unpredictable elements need to be transmitted. This condition of unpredictability is said to exist when the absolute difference between a prediction and the data element being predicted exceeds a preset, chosen threshold T_1 .

In such a telemetry system, which will eventually be designed, the receiver and transmitter will use the same set of weights and will operate on identical data elements so they can produce the same estimates of future points. The weights could also be calculated at the transmitter and telemetered to the receiver, but it seems wiser to send data rather than weights which are derived from the data. It might also take longer to transmit the weights than to transmit the data since the weights may have to be approximated to many places in order to get satisfactory results.

Here the set of weights is found from a set of TV elements called the learning set. Let the number of elements in the learning set be denoted by L . The first learning set is taken from the beginning of the first TV line. The first TV line has been taken here to be the top of the picture and its beginning is on the left-hand side so that the picture is scanned left to right and top to bottom. Each succeeding learning set is taken to begin at those multiples of $L/2$ from the beginning of each line at which new weights are desired. A new set of weights is needed when it is found that, on the average, the old set of weights is no longer producing satisfactory estimates. The rms prediction error is checked every $L/2$ elements against a threshold T_2 to determine when new weights are required. Threshold T_2 has previously been adjusted so that if T_2 is exceeded, many of the estimates of the last $L/2$ elements predicted would have exceeded T_1 . Those estimates which exceed T_1 are discarded and for the corresponding unpredictable picture element an original data element is substituted for transmission. Therefore, each interval of $L/2$ over which prediction has been attempted will contain original data elements and/or elements estimated by the predictor depending on whether T_1 had or had not been exceeded, respectively. Each time T_2 is exceeded $L/2$ elements from the last prediction interval and the succeeding $L/2$ original data elements are used for the required learning set. This choice is made to minimize the requirement

for transmission of actual data elements since only $L/2$ (not L) original data elements must be transmitted immediately following the exceedence of threshold T_2 .

The generation of new weights in an actual transmission system is costly and is thus to be avoided. Two measures have been taken to avoid unnecessary weight generation and *a fortiori* data transmission. First, at the beginning of each line, starting with the second line, the weights from the beginning of the preceding line are used; thus some advantage is taken of line-to-line correlation. The last set of weights calculated within $2L$ elements from the beginning of a line are the weights used for the beginning of the next line. The same set of weights is used to start all successive lines until T_2 is exceeded within $2L$ from the start of some future line; at this point the cycle starts again. Original TV data elements need be transmitted only at the start of each line in such quantity that the weights can be used to make the first estimate. Second, as one approaches the end of each TV line a new set of weights becomes less important since few elements remain to be estimated. Therefore, when only L or less elements remain to be estimated near the end of a TV line T_2 is ignored and the prediction process still using T_1 continues to the end of the line. Notice that the only place L original data elements are transmitted and used in the learning set, *a priori*, is at the beginning of the first line of each TV picture.

The method of producing the weights was to minimize the mean square estimation error over the learning set. A number of tests made in the "open loop" mode showed that satisfactory results can be obtained with the chosen values of M and L .

The set of elements upon which the weights operate to produce an estimate is called the memory and consists of the M picture elements immediately preceding the element to be estimated. According to the theory each estimate not only depends on the M elements of the memory but also on a constant K . It is convenient to denote W as the weight vector containing the set of $M + 1$ weights; then the number of W 's in the set of weights is $M + 1$ corresponding to K and the M elements in the memory. The weight vector W is the unknown to be found by using the learning set. This is accomplished by solving the matrix equation $SW = P$, where S is a matrix that has its rows made up of a memory set plus K , and P is a column vector of TV elements to be estimated from the corresponding row of the memory matrix S .

For example, let $x_n, x_{n+1}, x_{n+2}, \dots, x_{n+L-1}$ be a learning set of elements for some $n = 1, 2, \dots, 468 - (3L/2)$ on the first line. Then:

$$W = \begin{bmatrix} W_0 \\ W_1 \\ W_2 \\ \vdots \\ W_M \end{bmatrix} S = \begin{bmatrix} K & x_n & x_{n+1} & \cdots & x_{n+M-1} \\ K & x_{n+1} & x_{n+2} & \cdots & x_{n+M} \\ K & x_{n+2} & x_{n+3} & \cdots & x_{n+M+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K & x_{n+L-M-1} & \cdots & x_{n+L-2} \end{bmatrix} P = \begin{bmatrix} x_{n+M} \\ x_{n+M+1} \\ x_{n+M+2} \\ \vdots \\ x_{n+L-1} \end{bmatrix}$$

The method given on page 422 of Reference 2 was used for solution of the matrix equation $RW = G$ where $R = S^T S$, $G = S^T P$, and S^T is the transpose of S .

This method of solution gives W without error in $M + 1$ iterations if there is no round-off error. The iterative loop is stopped if either $|G - RW_i|^2$ drops below a preset threshold, here $|G - RW_0|^2 \times 10^{-4}$, or the number of iterations exceeds a preset count. (Here i is the iteration index and $|Y|^2 = \sum_{j=1}^n y_j^2$ is the squared length of Y , a real vector.)

The implementation of the program can best be followed by reference to the three flow diagrams in Figures 3 to 5. Two threshold values T_1 (± 2.5 and ± 4.5 quantum steps) have been used and the zero-order hold case ($M = 0$) has also been programmed. Although the problem of obtaining the maximum number of good picture element predictions is of foremost importance, there remains the problem of efficient coding once the prediction problem is solved. The report does not deal with this second problem, but coding methods have been tried with the zero-order hold method. It can be seen from Figure 3 that the program for the linear data predictor starts with five input control parameters which are read from a control card. These parameters are the prediction error threshold T_1 , the periodic average prediction error threshold T_2 , the number of TV elements L used to develop new weights, the number of TV elements M used to make each prediction, and the constant K associated with each prediction.

Most of the housekeeping work of the program is contained in the program's upper loop which cycles once for each TV line processed. This loop begins by providing a TV line from the input core storage to be processed and ends by returning the processed TV line to the output core storage. A subprogram keeps the input data buffer filled with data from the input TV data tape and transfers the output TV data from core storage to the output TV data tape. The upper loop contains a branch, either fork of which provides an initial set of weights and the starting position at which the first prediction is to be made. On the left-hand fork L TV elements are used to provide the first set of weights on the first line of the TV picture. These weights are stored for use at the start of future TV lines. The right-hand fork is entered on each cycle after the first and gets its weights from the core storage. A secondary loop located at the bottom of the diagram does the prediction, threshold-error tests, line-positioning tests, new weight calculations, and weight storage when the appropriate test is satisfied. This loop is cycled once for each TV element prediction. Each prediction replaces the TV element being predicted if the prediction error is less than or equal to T_1 . The L elements used to produce a new set of weights remain unchanged, as would be required for use in a receiver. The data going to the output tape will thus have the same form that it would have if it had been transmitted through an error-free communication channel.

The largest loop is completed when the 480th TV line has been processed. Finally, the TV element compression ratio is calculated and printed out as the program ends. A count of the number of predicted TV elements is used to calculate the element compression ratio.

An interesting case to consider is that with $M = 0$ and $L = 1$, which is the zero-order hold case referred to above. In this case, it is simpler to use the single element in the learning set as the estimate for succeeding elements, as long as the threshold T_1 is not exceeded. This procedure

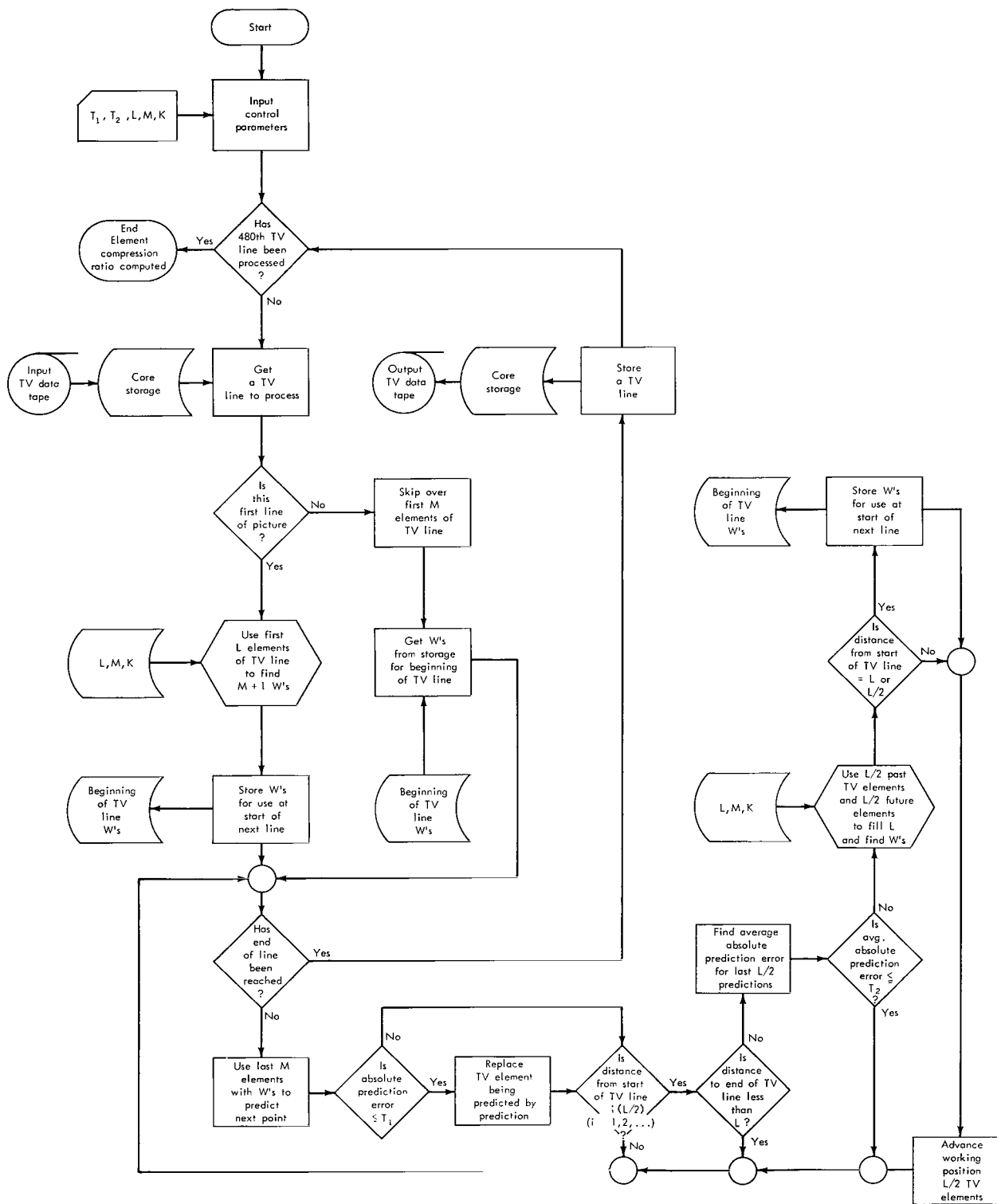


Figure 3—Flow diagram of linear TV data predictor.

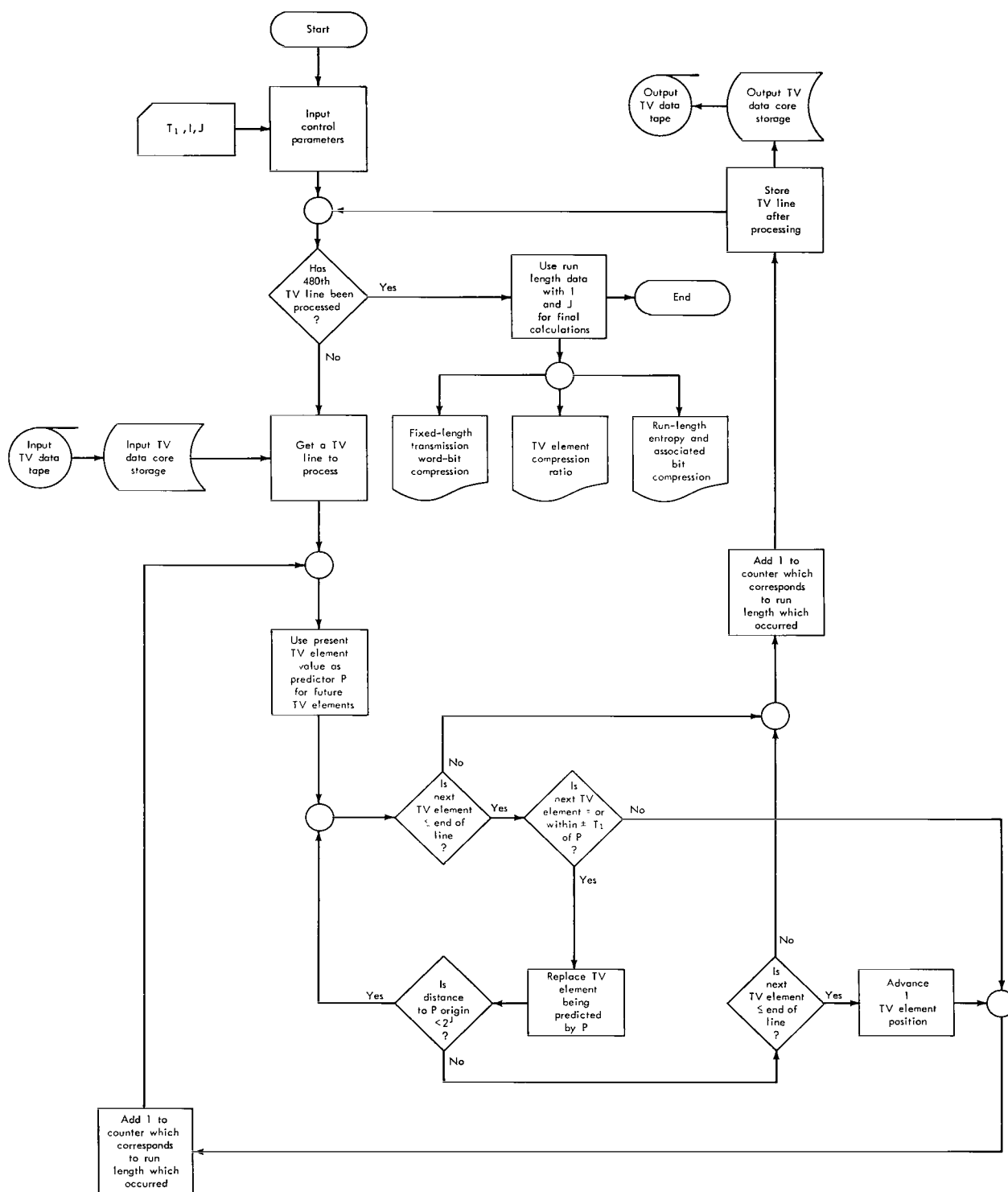


Figure 4—Flow diagram of zero-order hold predictor using first method of run-length coding.

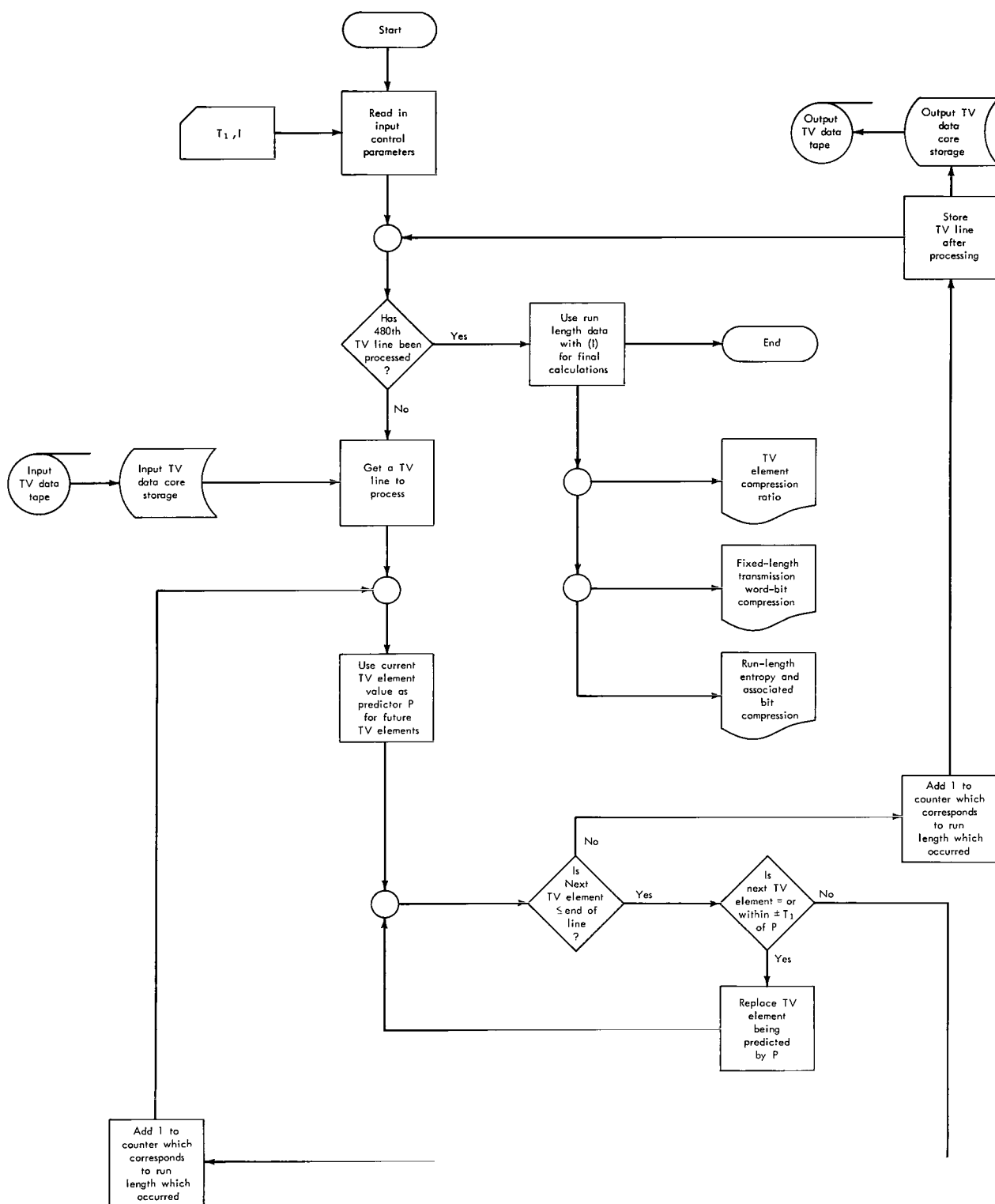


Figure 5—Flow diagram of zero-order hold predictor using second method of run-length coding.

produces the same estimate as would be obtained by multiplying W_0 by K . The process is further simplified in that T_2 is replaced by T_1 and each time T_1 is exceeded the TV element at that position is transmitted and becomes the new estimate. The first element of each TV line has been transmitted, which of course is not necessary. The zero-order hold may not produce as good an estimate over as many elements as does the case with $M = 3$; however, it has the advantage that it only requires one TV element to get a new estimator. The zero-order hold case is shown for $T_1 = 2$ and can be compared with $L = 20, M = 3$, for the same threshold.

Bit compressions were calculated for the zero-order hold case using run-length coding. Fixed-length transmission words of N bits each were assumed. Two methods of coding the run lengths were tried to determine which gave the more advantageous bit compression ratio. The run-length entropy and corresponding bit compressions were calculated for both methods:

$$\text{Entropy} = - \sum_{i=1}^n p_i \log_2 p_i$$

where n is the longest run length.

In the first method the N bits per word were derived so that the first I bits represent the TV element intensity and the last J bits represent the run length, where $I + J = N$. This method allows only run lengths of values varying from 1 to 2^J . Therefore, a new estimator will be found every 2^J elements, or more frequently if threshold T_1 is exceeded. The first method of run-length coding was tried with $I = 5$ and $J = 3, 4, 5$ for $T_1 = 2$ and $J = 4, 5, 6$ for $T_1 = 4$. TV pictures were produced but are not shown for the cases of $I = 5$ and $J = 4$ with $T_1 = 2$ and $T_1 = 4$. The bit compression ratios were not so high as with the second method of coding.

The run-length coding in the second method also used N bits per transmitted word, where $N = I + 1$. The first I bits of each word represent either a TV element intensity or run-length information, depending on the $I + 1$ bit. For the data used, I was five bits and therefore N was six bits per word. The run lengths were allowed to extend over a whole line if the data warranted it. Run lengths of 1 were indicated by transmitting the intensity of the first element. When the run length exceeded 1, the intensity of the first element was transmitted as before but was then followed by additional N -bit words to indicate the run length greater than 1. The number of elements in the run greater than 1 is first indicated by a binary number k ; k was then partitioned into enough I -bit sequences to use all the bits in k . Since k was not necessarily a multiple of I , it may happen that the last I -bit sequence is only partially filled by the last bits of k , in this case the rest of the last I -bit sequence will be zero-filled. Each of the I -bit sequences was finally transmitted as an N -bit code word.

The program using the first method of run-length coding (Figure 4) has three input control parameters which are read from a control card. The parameters are the error threshold T_1 , the number of bits I which represent each TV element intensity, and the number of bits J which represent the run length for the specified element intensity.

The major portion of the program is contained in the program's largest loop which cycles once for each TV line processed. This loop begins by providing a TV line from the input core storage to be processed and ends by returning the processed TV line to the output core storage. A subprogram keeps the input data buffer filled with data from the input TV data tape and transfers the output TV data from core storage to the output TV data tape. The second largest loop cycles once for each new predictor. Each new cycle of the secondary loop occurs when either the error tolerance T_1 is exceeded or the good prediction maximum run length 2^J is equaled. A "good" prediction is one with the associated prediction error less than or equal to T_1 . The run length is recorded by adding 1 to a counter which is indexed by the run length. The secondary loop contains the smallest loop which does the threshold testing and replaces the TV element being predicted by the predictor if T_1 is not exceeded. The data going to the output tape will thus have the same form that they would have had if they had been transmitted through an error-free communication channel.

The largest loop is completed when the 480th TV line has been processed. Finally the output information is calculated and printed out as the program ends. The output quantities are: the TV element compression ratio, the fixed-length transmission word bit compression, the run-length entropy, and the bit compression using the run-length entropy with a fixed-length element intensity word.

The zero-order hold program for the second coding method (Figure 5) uses two input control parameters which are read from a control card. These parameters are the error threshold T_1 and the number of bits I which represent each TV element.

The major portion of the program is contained in the program's largest loop which cycles once for each TV line processed. This loop begins by providing a TV line from the input core storage to be processed and ends by returning the processed TV line to the output core storage. A subprogram keeps the input data buffer filled with data from the input TV data tape and transfers the output TV data from core storage to the output TV data tape. The second largest loop cycles once for each predictor. A run length is recorded by adding 1 to a counter which is indexed by the run length. The secondary loop also contains the smallest loop which does the prediction and reconstructs the TV picture data.

The largest loop is completed when the 480th TV line has been processed. Finally, the output information is calculated and printed out as the program ends. The output quantities are: The TV element compression ratio, the fixed-length transmission word-bit compression, the run-length entropy, and the bit compression using the run-length entropy with a fixed-length element intensity word.

DISCUSSION OF RESULTS

Pictures are presented for the cases $L = 20$ and $M = 3$ elements with quantum errors $T_1 = 2.5$ and 4.5 and for the zero-order hold case $L = 1$ and $M = 0$ for $T_1 = 2$. The case for the zero memory with $L = 20$ elements was studied but showed no advantage in element compression ratio and

thus no pictures were prepared for it. The element compression ratio is considered the most meaningful value in separating the question of optimum coding from the prediction problem.

Element and bit compression ratios for Figures 6(c) to 15(c) with $L = 20$, $M = 3$, $K = 0$, $T_1 = 2.5$, and $T_2 = 3.0$ are given in Table 1. Figures 6(c) to 9(c) show essentially the same cloud features; the element compression ratios vary from 4.14 to 3.43 with an average ratio of approximately 3.8. It can be seen from a comparison of the digital originals with the compressed copies of these figures that the compressed pictures have areas with long horizontal stretches of equal brightness although the digital originals show random changes in these same areas. That this is a consequence of the chosen error limit as well as of the prediction technique is evident since the zero-order hold case with error limits the same as those used with the $L = 20$, $M = 3$, case gives poorer picture quality. Pictures prepared by the zero-order hold technique clearly illustrate this degradation, although many important structural details are preserved. For instance, Figure 4 shows a large gray area of broken cumulus clouds and a thin haze of, possibly, cirrus clouds; a small bright cloud spot can be located by measuring $3/4$ of the length of the fiducial mark from the corner of the lower left fiducial mark and drawing lines parallel to the mark. This cloud apparently projects higher than its surrounding cumuli because it is illuminated brighter than its surroundings; the same spot can be located in the digital original and the processed picture. When the satellite movement is followed on the picture sequence this cloud feature can be located with difficulty on the original analog Figures 6(a) to 9(a). It is much less clear on the digital originals, except in Figure 7(b), and also less clear on the processed pictures. Other details like the bright streaks running parallel to the field in the centrally located cloud field of Figure 8 are less pronounced in the digital originals but appear with identical definition on the processed pictures (Figures 6(c) to 9(c)).

The appearance of Figures 10 and 11 is entirely different and the pictures show no distinct patterns. Element compression ratios of 3.51 and 3.19 were achieved for Figures 10(c) and 11(c) and an average compression ratio close to the average for the preceding four frames was obtained. Few distinct features can be located except for the tearing of apparently high-rising cumulus clouds because of convection. In Figure 10 there is an elliptically shaped cloud with a darker center spot in the lower left quadrant, located by drawing lines parallel to the mark, one bisecting the horizontal mark and the other intersecting the vertical portion $3/4$ of the way up the lower left fiducial

Table 1

Element and Bit Compression Ratios for Figures 6(c) to 15(c).

Figure	6(c)	7(c)	8(c)	9(c)	10(c)	11(c)	12(c)	13(c)	14(c)	15(c)	Av.*
Element compression ratio	3.79	4.14	3.43	4.09	3.51	3.19	3.45	2.29	3.49	6.27	3.56
Bit compression ratio	2.15	2.36	2.04	2.30	2.04	1.86	2.08	1.49	2.05	3.35	2.09

*Average number of bits per picture converted to element and bit compression ratios.

mark. The longer dimension of this spot seems to be about 10 TV lines wide. It can be seen with equal clarity in both originals and in the processed picture. In the clear-to-hazy area in the lower right-hand section there are broken elongated sections where the very dark ocean is visible underneath. These sections appear equally well on all copies. An apparently high cirrus field at the center right can be surmised as such in the analog original but is difficult to recognize in the digital original. The processed picture shows the same quality as the digital original.

The next picture (Figure 12) is taken from Tiros V over the general vicinity of Bermuda and shows a distinct cloud grouping with arclike patterns in the right half of it. The fine structure and horizontal variation in the lower left third of the picture are due to bad detection of the synchronization pulse caused presumably by poor S/N ratio in the receiver. Although no record of the received field strength is available it is reasonable to assume that Tiros was low at the horizon as seen from the Wallops Island receiving site. The quality of the processed picture (Figure 12(c)) is as good as that of the previous processed pictures and the compression ratio of 3.45 is reasonable.

Figure 13 shows a decaying hurricane or cyclone off Miami and has the poorest element compression ratio. Comparison of the analog and digital originals shows a marked increase in gray background noise, presumably because of bad adjustment of the video amplifier ahead of the A/D converter. The effect of this higher level of background noise is the poorest element compression ratio produced for any of the pictures. This is much in contrast with what one expects by comparison with the preceding three pictures. The much higher element compression ratio of Figure 15(c) which shows a similar pattern and a considerably extended sky area also points to this conclusion.

Comparison of the analog and digital originals of Figure 14 shows a loss in definition of the possible storm center in the upper left third section of the photograph near the horizon. The suspicion of a storm center is partly suggested from knowledge of the succeeding frame delivered which is not included in this series. The processed picture shows no loss other than in the digital original and has approximately the same element compression ratio as do the previous pictures. The final picture is of a classic cyclone located northeast of Bermuda in May 1963. The pronounced pattern and the large sky portion offer the suggestion that better prediction should be achieved than in any of the other pictures, particularly better than in Figures 10 and 11. Comparison of the originals and the processed copy reveals no significant loss in detail even though the compression ratio was the highest of all.

For comparison the analysis was repeated using five-bit data and permitting a quantum error of ± 4.5 quantum steps. The element and bit compression ratios given in Table 2 were achieved with $L = 20$, $M = 3$; $K = 0$, $T_1 = 4.5$, and $T_2 = 5.0$. It should be noted that with such a large error tolerance (9 out of 32 possible levels) only a coarse intensity gradation is achieved. While any one of the 32 possible amplitudes can affect the reproduced picture, dividing the total range by the nine quantum error levels yields only three and a half center levels which significantly degrade the gray scale of the original picture. The most disturbing feature of this picture series is the constant amplitude over extended line segments. This streakiness is a direct consequence of the coarse error tolerance. Large quantum errors can thus be eliminated from further consideration.

Figures 6(e) to 15(e) were produced with $T_1 = 2$ for $L = 1$ and $M = 0$, the zero-order hold case, and with the following compression and bit ratios given in Table 3. As can be seen by comparison

Table 2

Element and Bit Compression Ratios for Figures 6(d) to 15(d).

Figure	6(d)	7(d)	8(d)	9(d)	10(d)	11(d)	12(d)	13(d)	14(d)	15(d)	Av.*
Element compression ratio	7.29	11.10	6.92	10.34	10.17	6.18	5.40	5.39	7.47	11.90	7.59
Bit compression ratio	3.80	5.26	3.42	4.98	4.86	3.25	2.87	2.85	3.81	5.50	3.84

*Average number of bits per picture converted to element and bit compression ratios.

Table 3

Element Compression and Bit Ratios for Figures 6(e) to 15(e).

Figure	6(e)	7(e)	8(e)	9(e)	10(e)	11(e)	12(e)	13(e)	14(e)	15(e)	Av.*
Element compression ratio	5.42	4.88	4.55	4.93	3.71	3.60	4.43	2.29	4.15	9.50	4.22
Bit compression ratio	2.94	2.64	2.51	2.67	2.16	2.11	2.48	1.54	2.39	4.67	2.43

*Average number of bits per picture converted to element and bit compression ratios.

the compression ratios are all larger than in the $L = 20$, $M = 3$ case for $T_1 = 2.5$ but it appears that a significant amount of detail is lost. Figures 6(e) to 9(e) show particularly serious losses. The bright spot near the lower left fiducial mark in Figure 7(e) is difficult to recognize and much detail in the central cloud area is not visible. The only way to account for this is by assuming that the signal is rarely predicted correctly but stays within the error limit for long lengths. The difficulties in assigning the gray scale to the SC-4020 microfilm recorder are partly to blame for the bright overexposure of large areas but cannot account for the general streakiness of large sections; this is a result of the lack of accurate prediction. This is especially apparent in Figure 10(e) where the small elliptic cloud does not show a central darker area and the deep black patterns at the lower right-hand section between the fiducial mark angle and cross are not visible.

The explanation as to why the element compression ratio is higher for the zero-order hold, that is, with $M = 0$, than for $M = 3$ comes from the fact that only a short learning period is required with $M = 0$. In particular the learning period used herein was one element with $M = 0$. This means that the predictor is up-dated every time a TV element is not predicted so that changes in the data are followed relatively quickly compared with the $M = 3$ case. Data from the ten picture average showed 53,000 predictors used with $M = 0$ compared with only 800 predictors for $M = 3$. These figures indicate that the predictor for $M = 0$ is changed 66 times for each predictor used with $M = 3$.

The economy of learning with $M = 0$ allows frequent up-dating of the predictor so that even though the zero-order hold neither predicts as well nor maintains its quality for as long as the $M = 3$ case a higher compression ratio is still achieved with it. A second method also given by Balakrishnan and referred to as Method II in Reference 1 has been programed and has yielded results with better element compression ratios than has the zero-order hold.

CONCLUSIONS

Compression ratios achieved offer sufficient promise to continue the investigation and refine the method. A comparison of the results obtained using $M = 0$ with those obtained using $M = 3$ suggests that the two cases be combined and the constant term in the $M = 3$ case be replaced by the zero-order hold. The learning error may be more conveniently controlled and a better solution therefore found if the steepest descents solution is used in solving for the weights as mentioned in Reference 1. At the time of this writing preliminary results are available using Method II of Reference 1 which show even better compression ratios than those reported herein. Further optimization of the parameters chosen remains to be accomplished and the corresponding effect on picture quality determined.

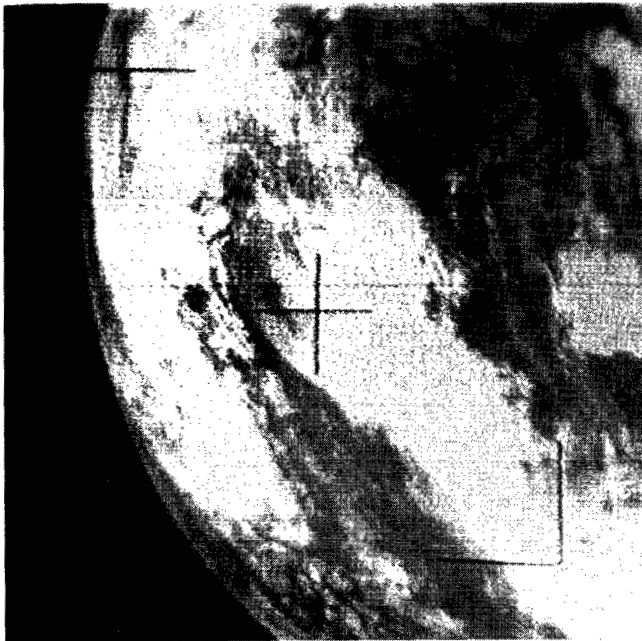
ACKNOWLEDGMENTS

The authors are indebted to Herbert J. Gillus who programed the IBM 7094 and to Joseph A. Sciulli who obtained the ten TV pictures and checked the manuscript.

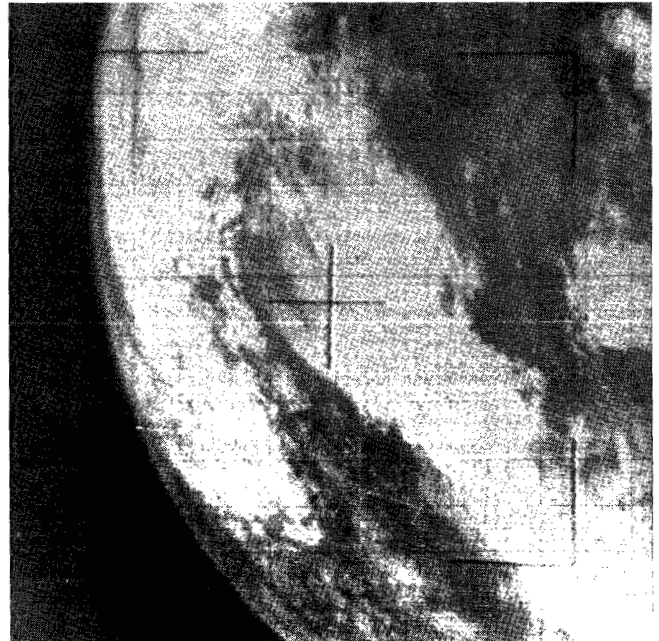
(Manuscript received August 31, 1965)

REFERENCES

1. Balakrishnan, A. B., "An Adaptive Non-Linear Data Predictor," Paper No. 6-5, Proc. 1962 National Telemetering Conf. (May 23-25, 1962, Washington, D. C.), The Institute of Radio Engineers, New York N.Y., 1962.
2. Hestenes, M. R., and Stiefel, E., "Method of Conjugate Gradients for Solving Linear Equations," *J. Res. Nat. Bur. Standards* 49(6):409-436, December 1952.

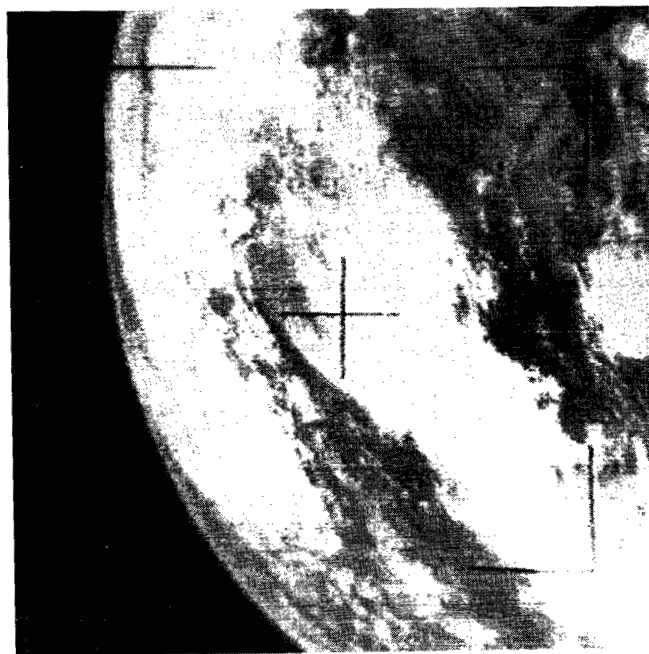


(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

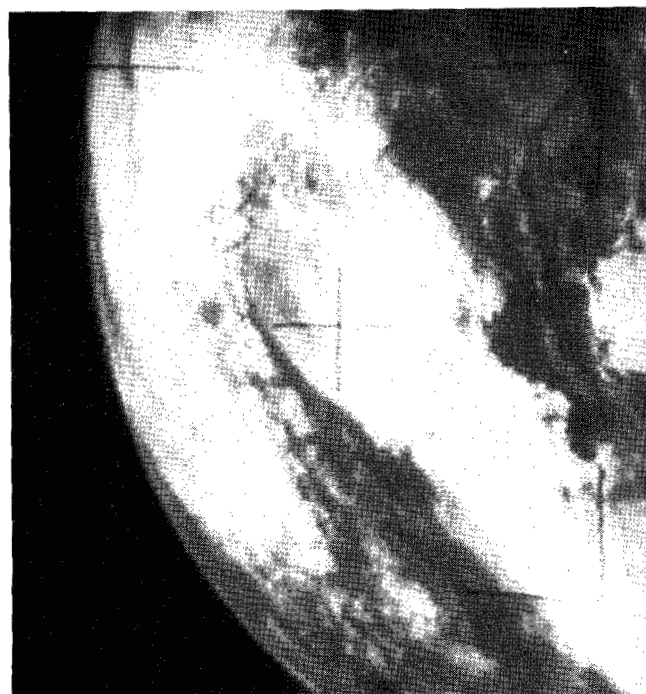
Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 3.79; bit compression for fixed-length code of 6 bits per data word, 2.15.



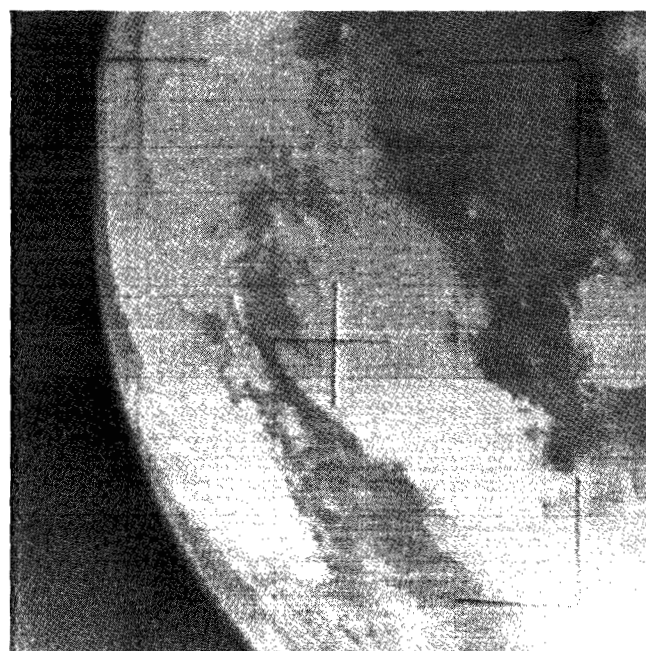
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 5.42; bit compression for fixed-length code of 6 bits per data word, 2.94.

Figure 6—Pictures from Tiros III, orbit 4, frame 2, camera 2, direct transmission from satellite.

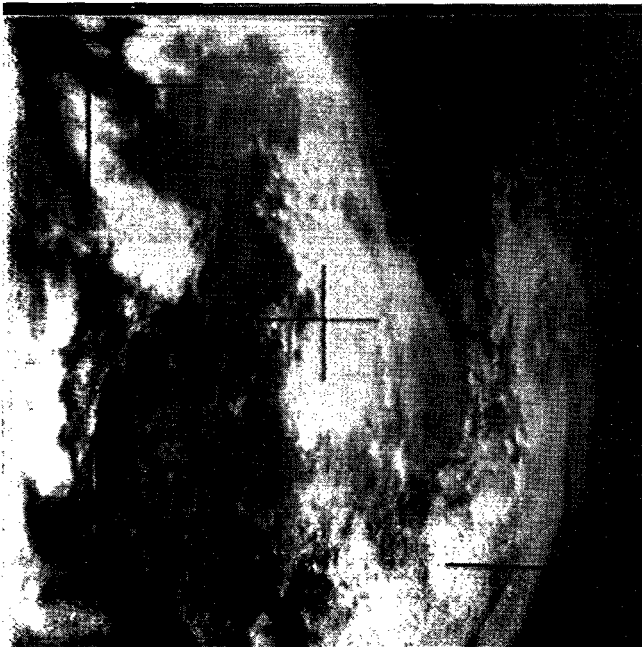


(a) Analog original.



(d) Linear predictor with 4 quantum steps error.

Processed copy; $t_1 = 4.5$ $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 7.29; bit compression for fixed-length code of 6 bits per data word, 3.80.



(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 4.14; bit compression for fixed-length code of 6 bits per data word, 2.36.



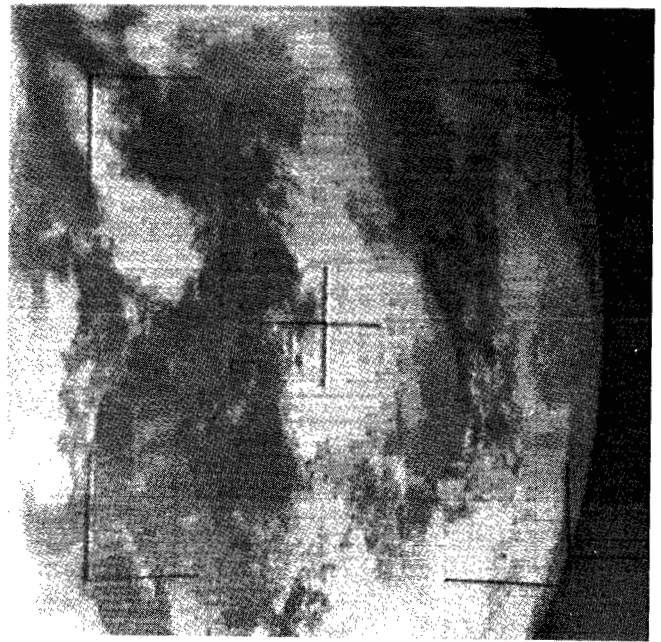
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 4.88; bit compression for fixed-length code of 6 bits per data word, 2.64.

Figure 7—Pictures from Tiros III, orbit 4, frame 3; camera 2; direct transmission from satellite.



(a) Analog original.

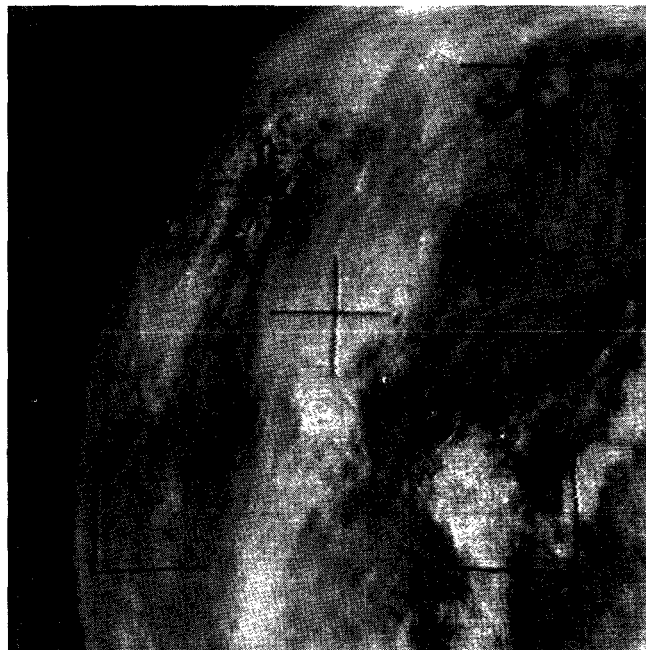


(d) Linear predictor with 4 quantum steps error.

Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 11.10; bit compression for fixed-length code of 6 bits per data word, 5.26.



(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

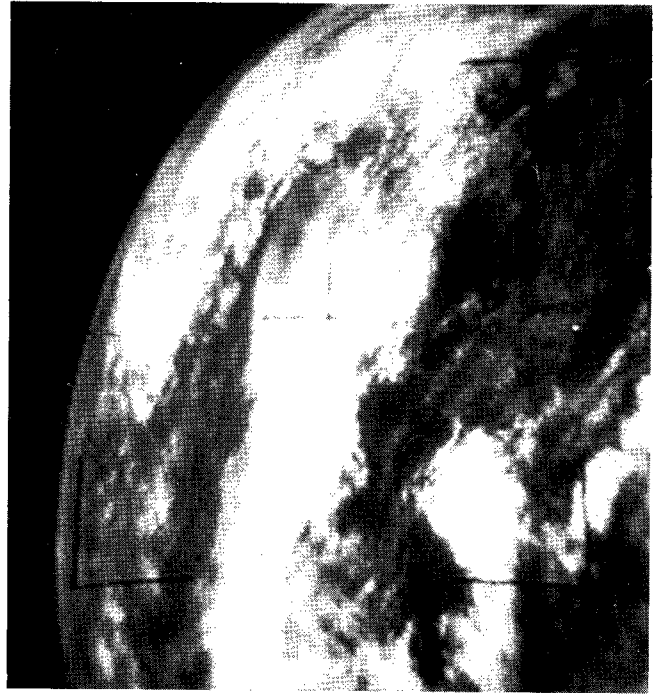
Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 3.43; bit compression for fixed-length code of 6 bits per data word, 2.04.



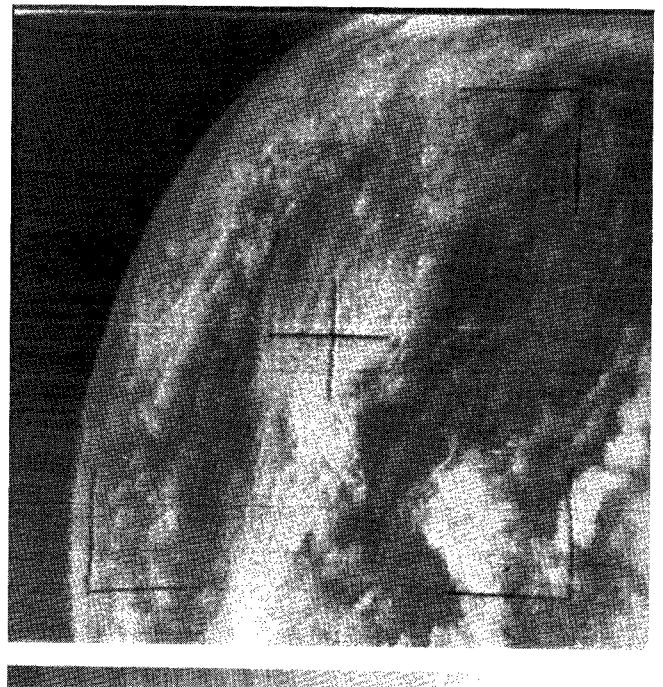
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 4.55; bit compression for fixed-length code of 6 bits per data word, 2.51.

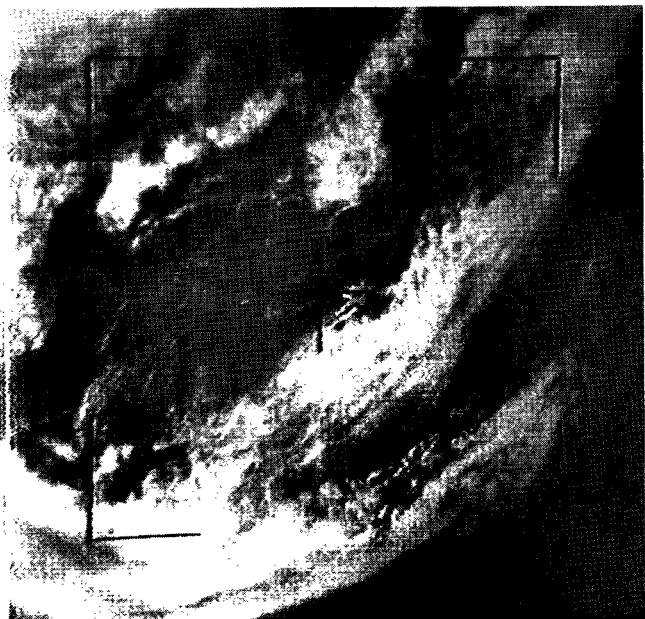
Figure 8—Pictures from Tiros III, orbit 4, frame 4, camera 2, direct transmission from satellite.



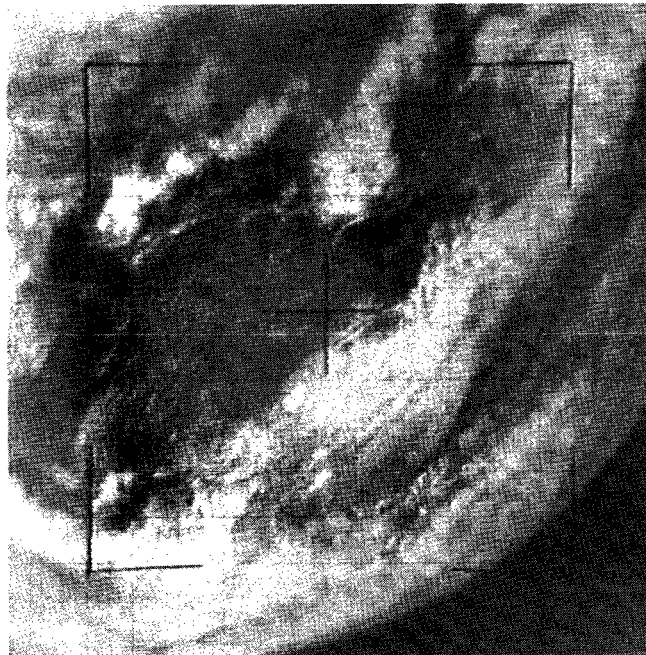
(a) Analog original.



(d) Linear predictor with 4 quantum steps error.
 Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 6.92; bit compression for fixed-length code of 6 bits per data word, 3.42.

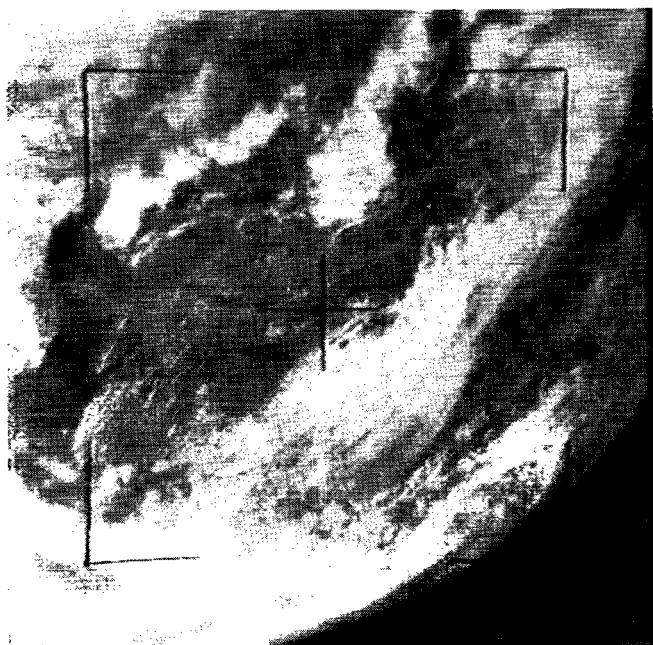


(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

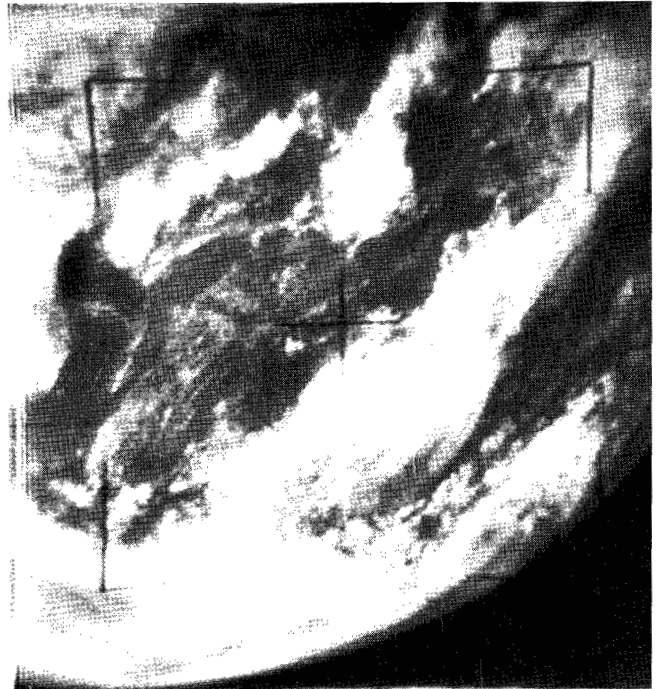
Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 4.09; bit compression for fixed-length code of 6 bits per data word, 2.30.



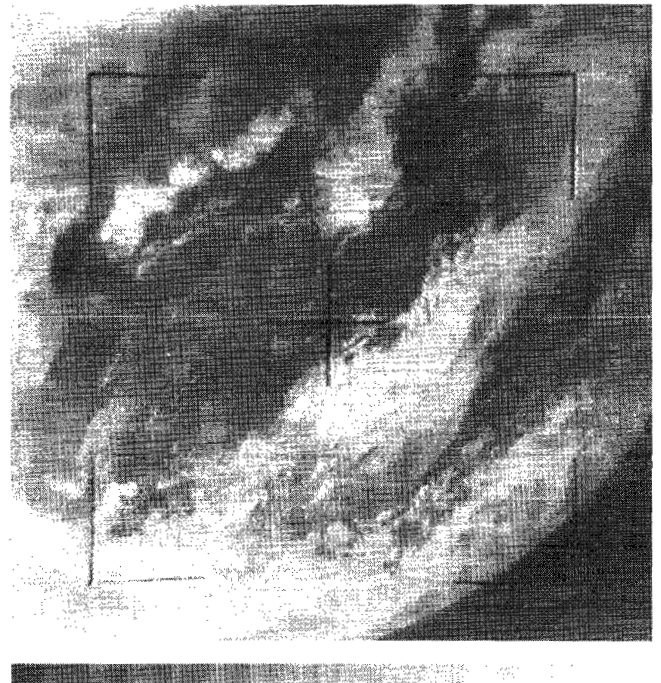
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 4.93; bit compression for fixed-length code of 6 bits per data word, 2.67.

Figure 9—Pictures from Tiros III, orbit 4, frame 5, camera 2, direct transmission from satellite.

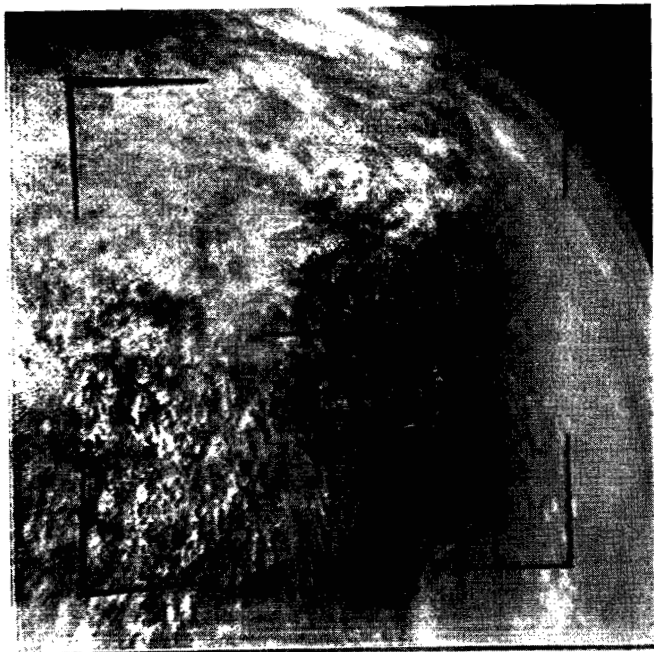


(a) Analog original.

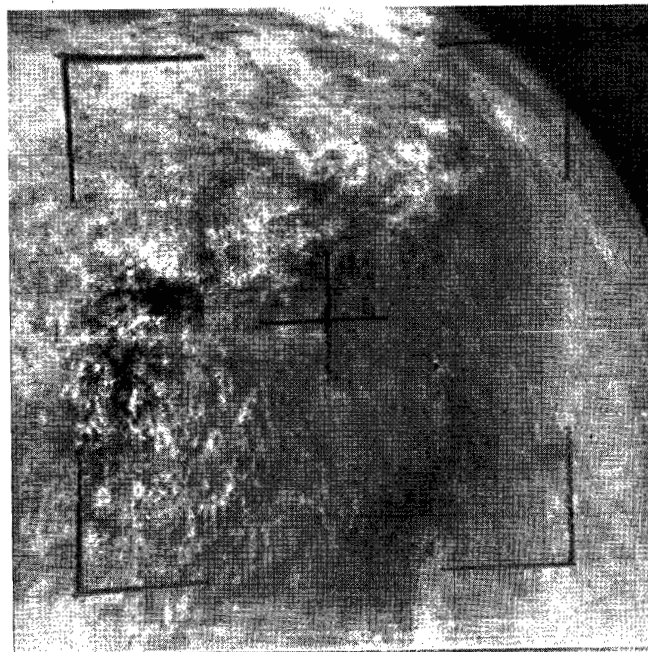


(d) Linear predictor with 4 quantum steps error.

Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 10.34; bit compression for fixed-length code of 6 bits per data word, 4.98.



(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

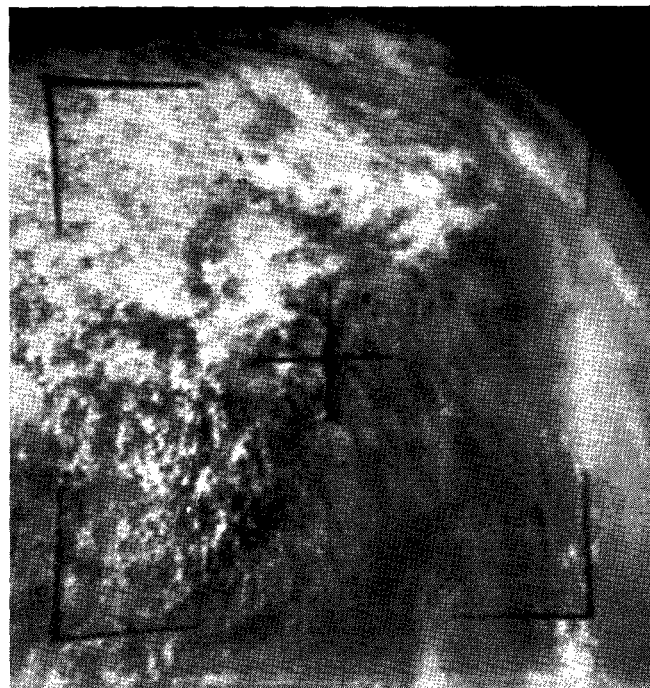
Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 3.51; bit compression for fixed-length code of 6 bits per data word, 2.04.



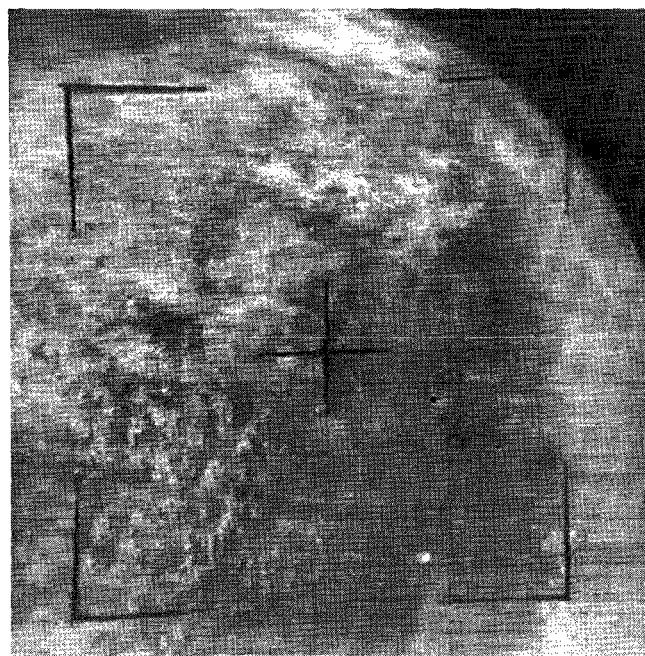
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 3.71; bit compression for fixed-length code of 6 bits per data word, 2.16.

Figure 10—Pictures from Tiros III, orbit 102, frame 1, camera 1, taped before transmission from satellite.

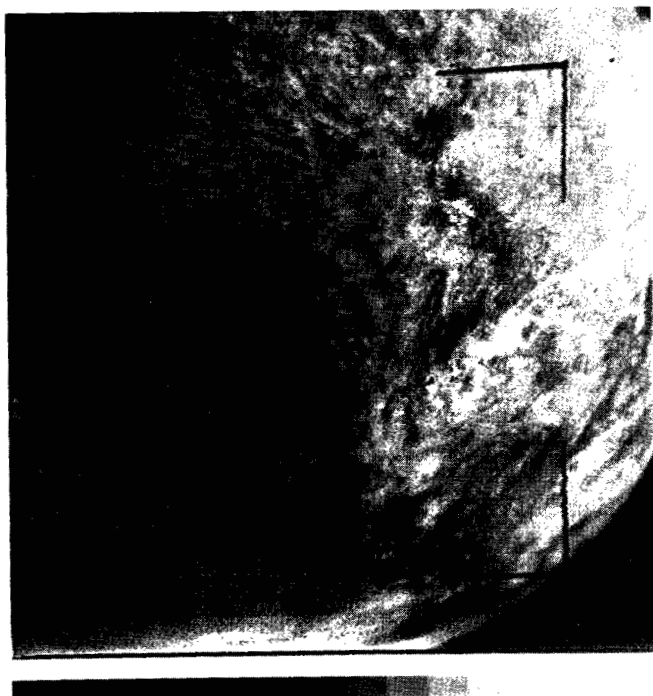


(a) Analog original.

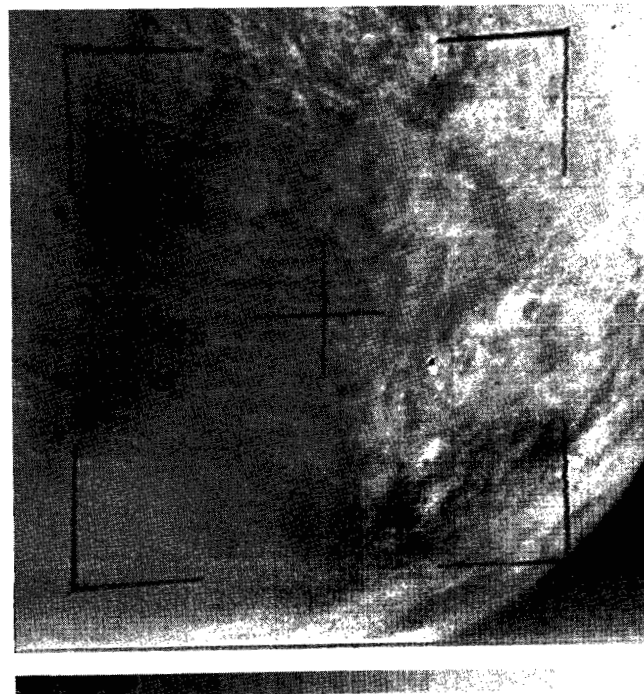


(d) Linear predictor with 4 quantum steps error.

Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 10.17; bit compression for fixed-length code of 6 bits per data word, 4.86.



(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

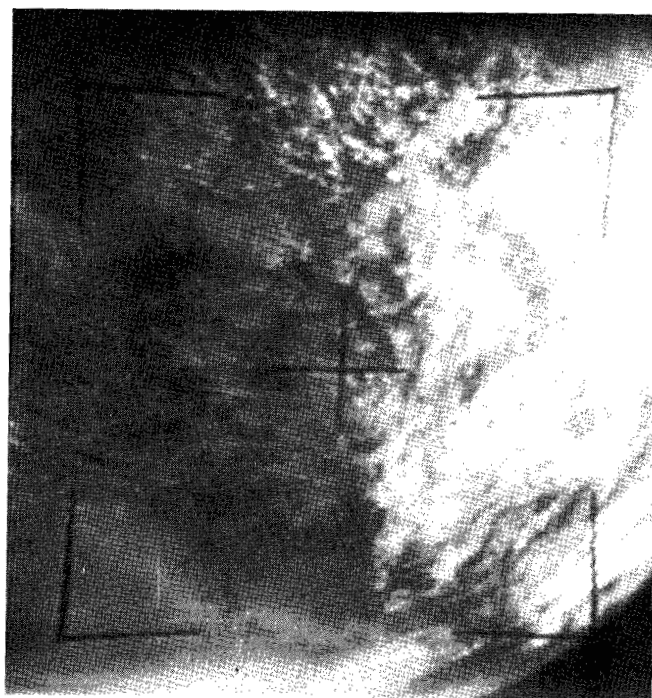
Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 3.19; bit compression for fixed-length code of 6 bits per data word, 1.86.



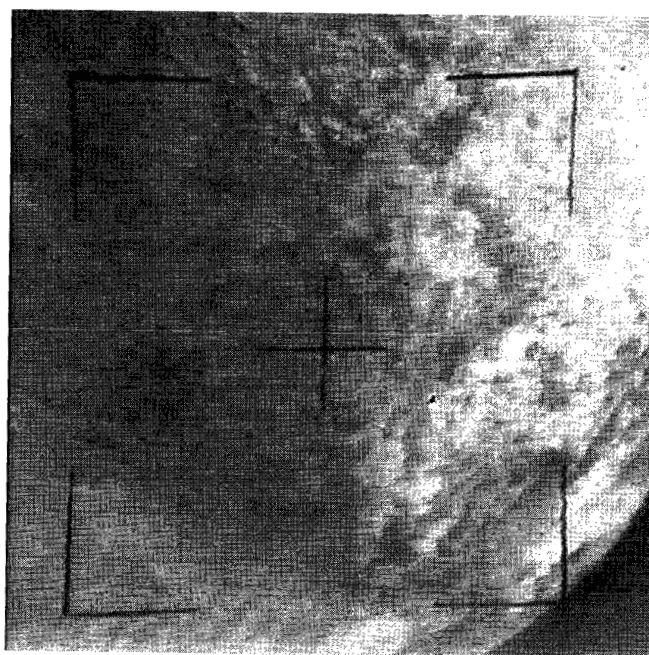
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 3.60; bit compression for fixed-length code of 6 bits per data word, 2.11.

Figure 11—Pictures from Tiros III, orbit 102, frame 2, camera 1, taped before transmission from satellite.



(a) Analog original.

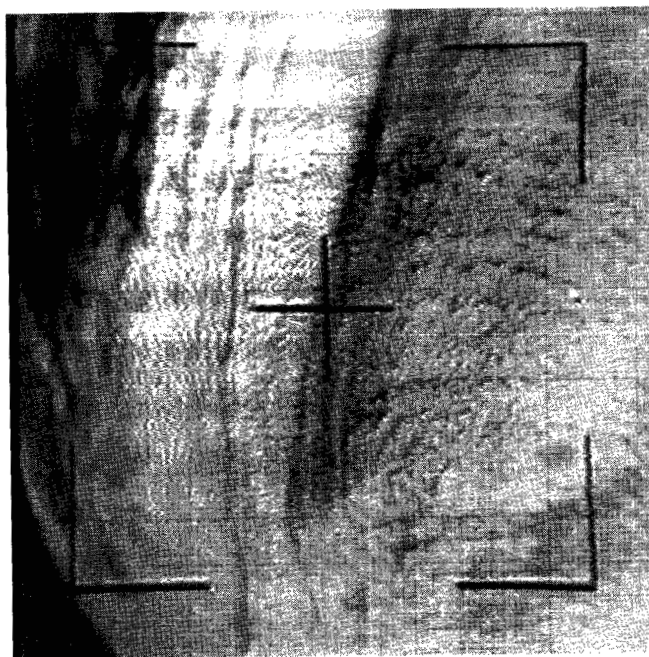


(d) Linear predictor with 4 quantum steps error.

Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 6.18; bit compression for fixed-length code of 6 bits per data word, 3.25.

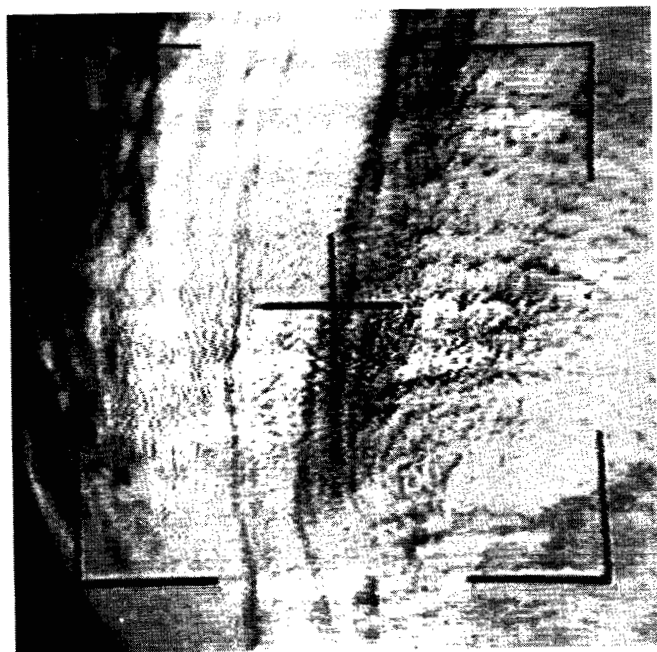


(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 3.45; bit compression for fixed-length code of 6 bits per data word, 2.08.



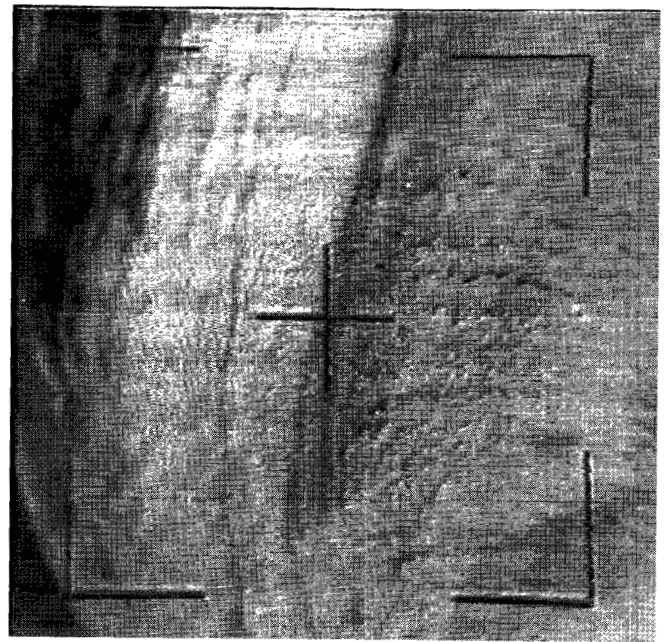
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 4.43; bit compression for fixed-length code of 6 bits per data word, 2.48.

Figure 12—Pictures from Tiros V, orbit 3143, frame 7, camera 1, direct transmission from satellite.

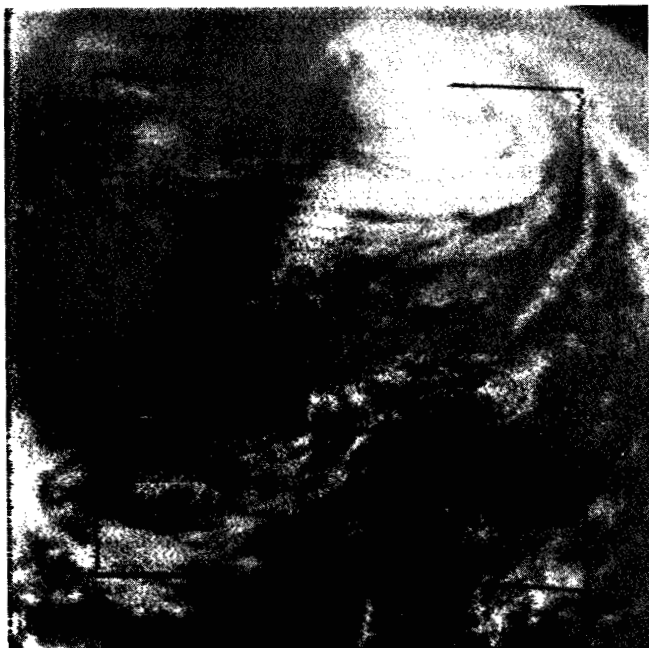


(a) Analog original.

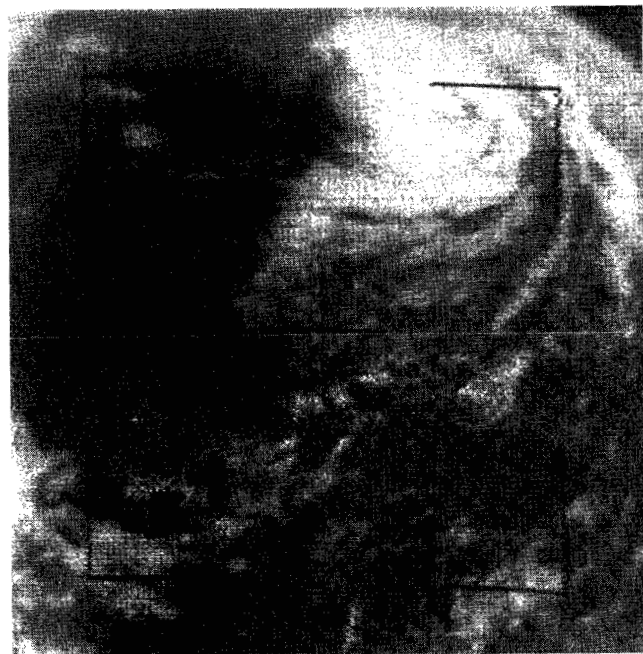


(d) Linear predictor with 4 quantum steps error.

Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 5.40; bit compression for fixed-length code of 6 bits per data word, 2.87.

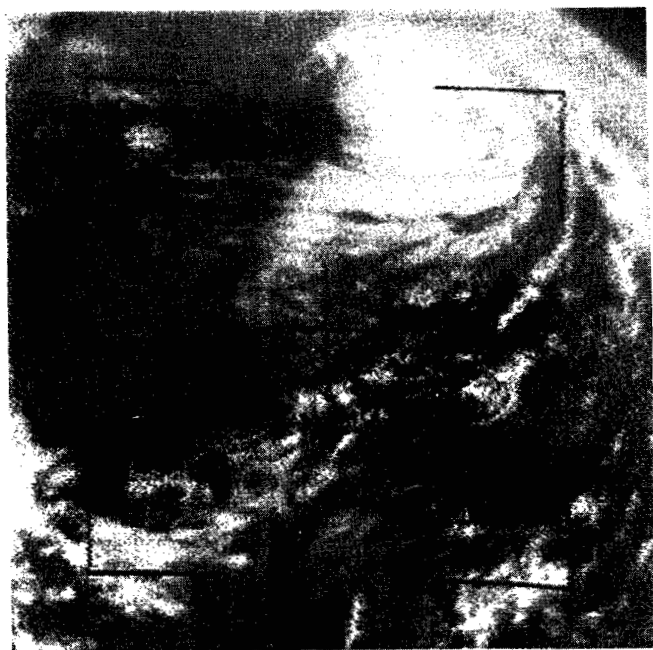


(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

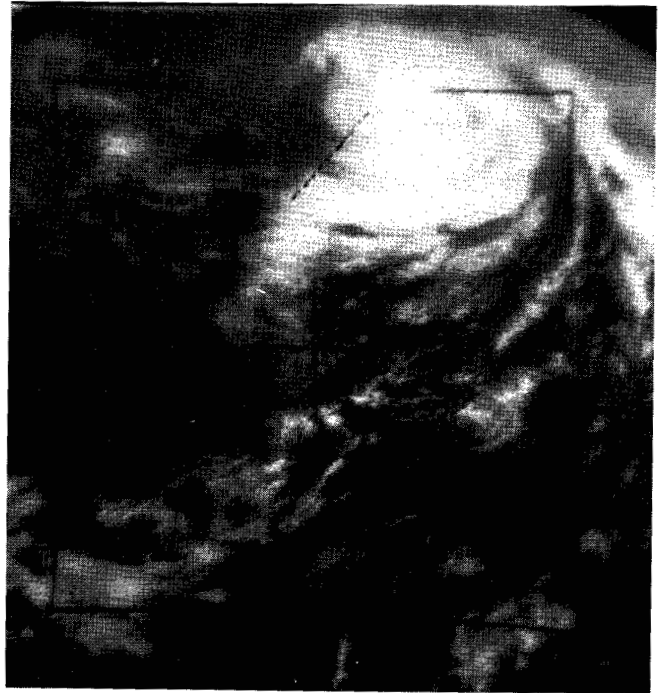
Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 2.29; bit compression for fixed-length code of 6 bits per data word, 1.49.



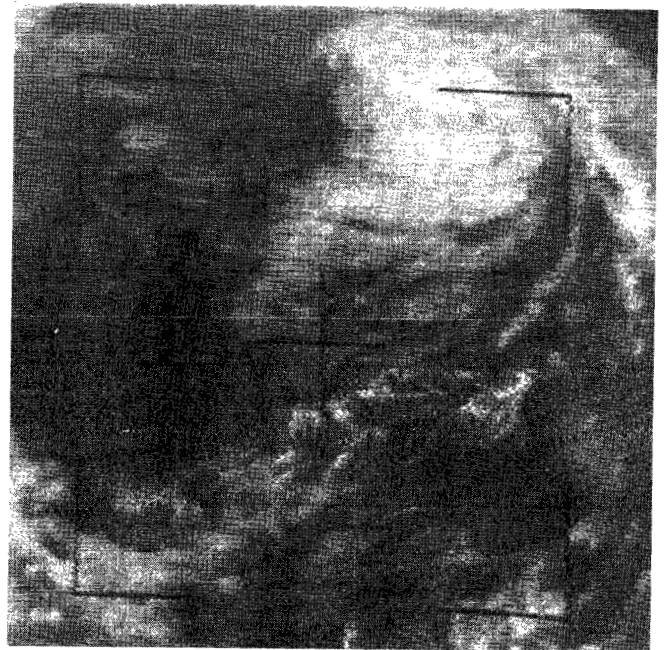
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 2.29; bit compression for fixed-length code of 6 bits per data word, 1.54.

Figure 13—Pictures from Tiros VI, orbit 1100, frame 15, camera 1, direct transmission from satellite.

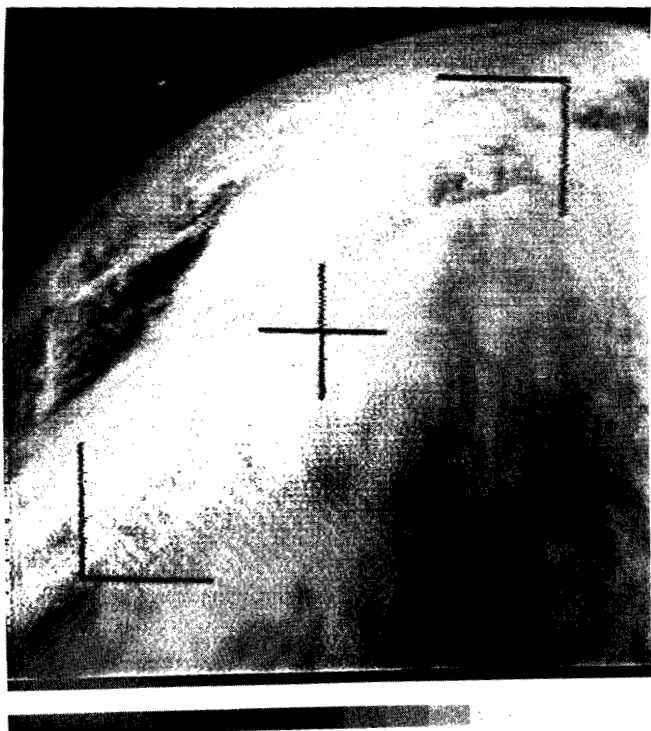


(a) Analog original

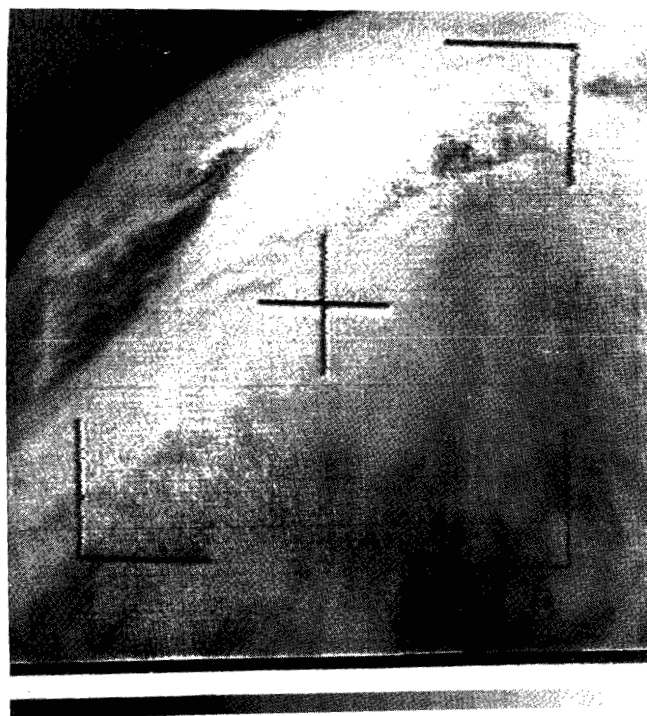


(d) Linear predictor with 4 quantum steps error.

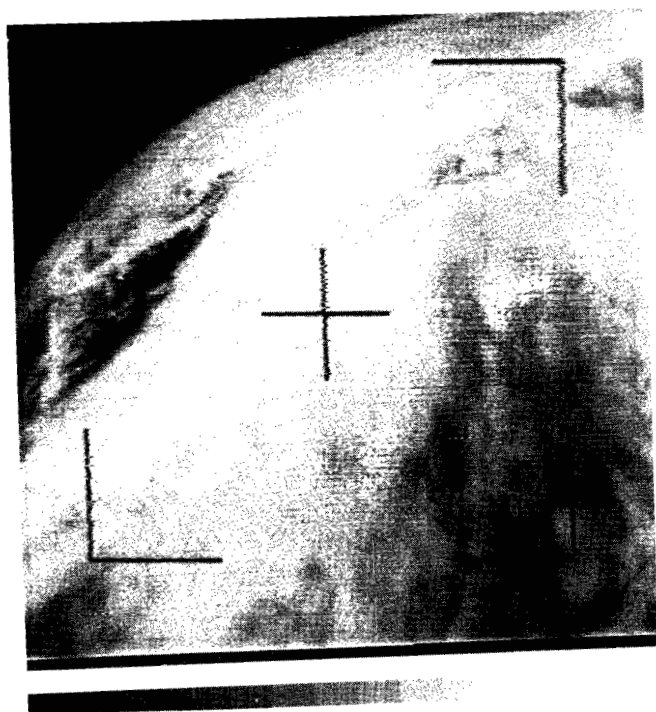
Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 30$; $M = 3$; $K = 0$; element compression, 5.39; bit compression for fixed-length code of 6 bits per data word, 2.85.



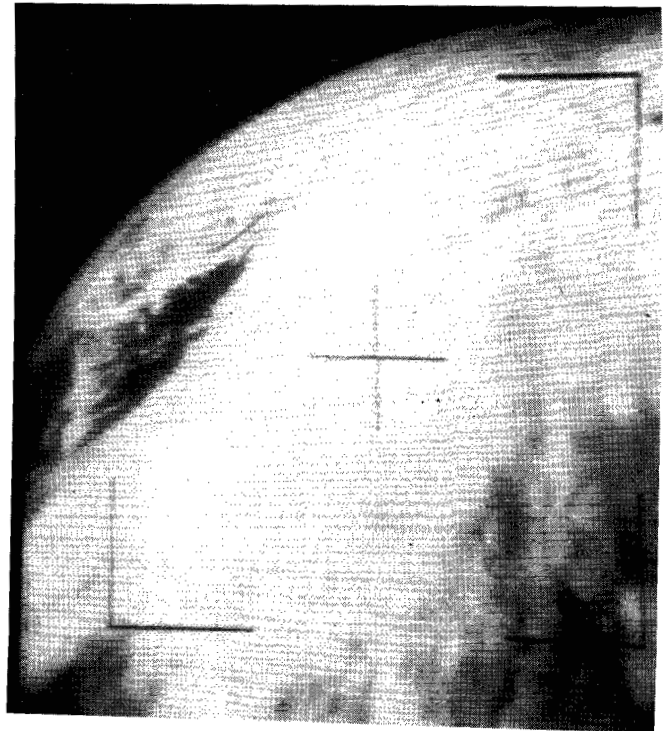
(b) Digital original.



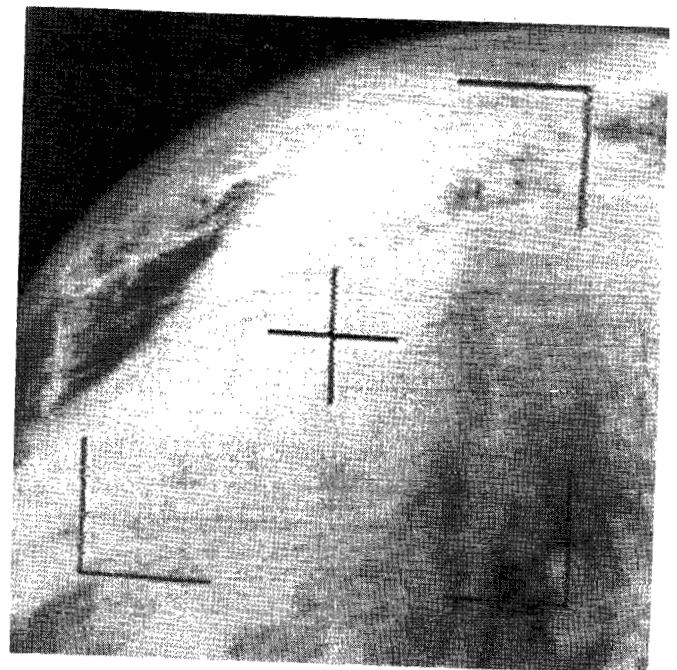
(c) Linear predictor with 2 quantum steps error.
Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 3.49; bit compression for fixed-length code of 6 bits per data word, 2.05



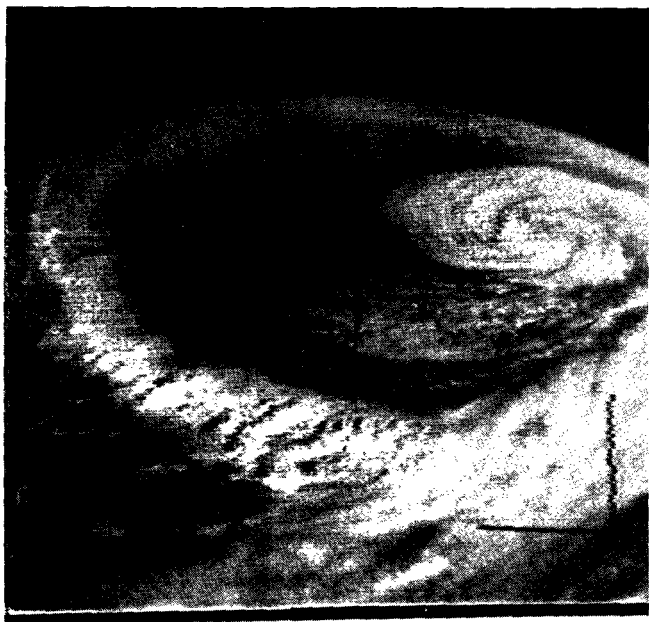
(e) Zero order hold with 2 quantum steps error.
Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 4.15; bit compression for fixed-length code of 6 bits per data word, 2.39.



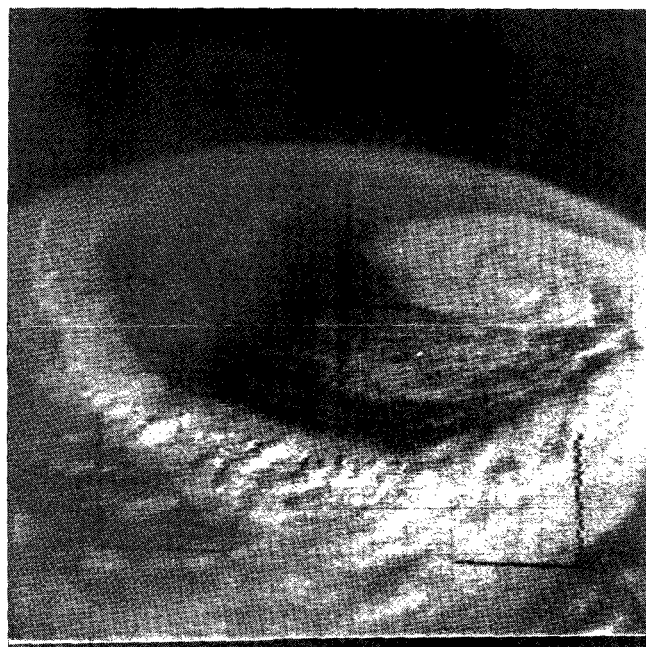
(a) Analog original.



(d) Linear predictor with 4 quantum steps error.
 Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 7.47; bit compression for fixed-length code of 6 bits per data word, 3.81.

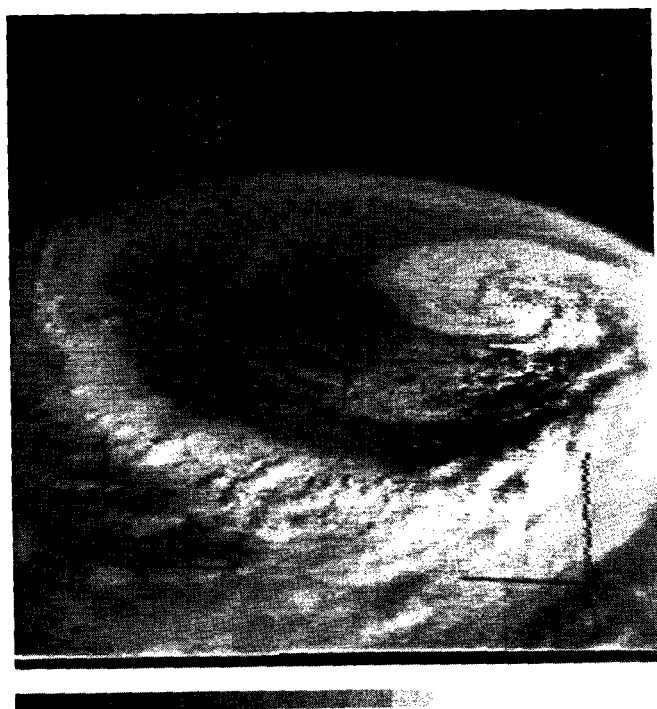


(b) Digital original.



(c) Linear predictor with 2 quantum steps error.

Processed copy; $t_1 = 2.5$; $t_2 = 3.0$; $L = 30$; $M = 3$; $K = 0$; element compression, 6.27; bit compression for fixed-length code of 6 bits per data word, 3.35.



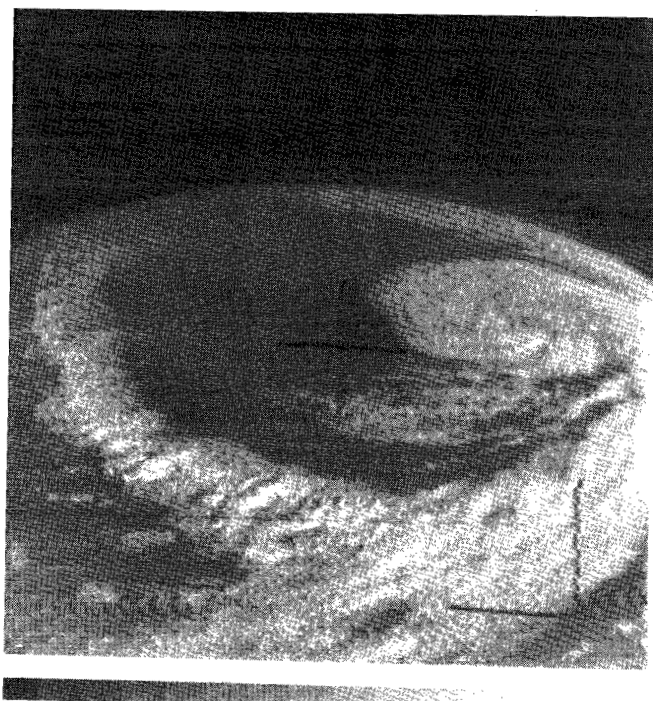
(e) Zero order hold with 2 quantum steps error.

Processed copy; $t_1 = 2.0$; $L = 1$; $M = 0$; element compression, 9.50; bit compression for fixed-length code of 6 bits per data word, 4.67.

Figure 15—Pictures from Tiros VI, orbit 3692, frame 31, camera 1, taped before transmission from satellite.



(a) Analog original.



(d) Linear predictor with 4 quantum steps error.
 Processed copy; $t_1 = 4.5$; $t_2 = 5.0$; $L = 20$; $M = 3$; $K = 0$; element compression, 11.90; bit compression for fixed-length code of 6 bits per data word, 5.50.

Appendix A

Symbols

B	bandwidth (cps)	P_r	probability of exceedence of threshold
C	symmetric positive cost function	p_i	probability numbers
D	determinant	$R = S^T S$	
f	frequency	S	memory matrix
$G = S^T P$		S^T	transpose of S
I	TV element intensity or run-length bits information	T	duration of waveform
J	run length bits	T_1	preset threshold of difference between prediction and data element, the quantum error
K	constant associated with each prediction	T_2	periodic average prediction error threshold
k	number of elements in run greater than one	t	time parameter
L	learning period, the number of picture elements from which weighing function is derived; also, length	t_0	value of t in interval $0 < t < T - \Delta$
M	memory, the number of picture elements preceding data point to be predicted used to make each prediction	W	weight vector containing set of M + 1 weights
N	number of bits per transmitted word	$x(t)$	continuous or discrete parameter source
n	number of sample	$x^*(t_0 + \Delta)$	predicted value of data
O	prediction operator	x_n	data samples
P	column vector of TV elements	Y_n	N-column vector
		α_j	optimal filter weights
		Δ	differential of time
		ϵ	prediction error
		$\rho(t)$	positive weight function
		σ	dummy variable of time