A ZERO-ERROR OPERATIONAL VIDEO DATA COMPRESSION SYSTEM

Richard L. Kutz

I would first like to explain the significance of "Zero Error" as used in the title. The data user is usually alarmed when any operation which is unfamiliar to him is performed upon the data before it is received at the data-user processing site. In order to minimize the concern of the data user over the possible loss of data quality due to data compression, all data was reconstructed for the data user without error with respect to the data at the input point in the communication system. So this data compression communication system is zero error or lossless since it is transparent to the data user as far as data quality is concerned. You still have to contend with channel errors, as you would with any other communication system.

This data compression system was implemented as an experiment in cooperation with the National Environmental Satellite Service (NESS) of the National Oceanic and Atmospheric Administration. The data compression system operates between the NESS command and data acquisition (CDA) station at Wallops Island, Virginia, and the NESS data processing center at Suitland, Maryland.

The data compression system has been operating since February 1972, using ATS spin-scan cloud cover data. With the launch of ITOS-D in October 1972, this data compression system has become the only source of near-realtime Very High Resolution Radiometer (VHRR) image data at the Suitland NESS data processing facility. The VHRR image data are compressed and transmitted over a 50 kilobit per second wideband ground link between Wallops Island, Virginia and Suitland, Maryland.

The goal of the GSFC/NESS data compression experiment was to send data quantized to six bits at twice the rate possible when no compression is used, while maintaining zero error between the transmitted and reconstructed data. All objectives of the data compression experiment were met, and thus a capability of doubling the data throughput of the system has been achieved.

In the first figure, to the left is a CDA station receiver feeding a 100 kilobit per second data stream into the data compression equipment. The data compression hardware consists of a programmable minicomputer with special purpose input-output interfaces designed to facilitate the data compression software algorithms. In the center of Figure 1, the compressed data is transmitted at a 50 kilobit per second rate over a wideband group phone line to the data reconstruction equipment located at Suitland, Maryland. To the right of the figure, the inverse of the data compression operation is performed on the 50 kilobit per second compressed data stream from the phone line to reconstruct a 100 kilobit per second data stream output which is identical to the 100 kilobit per second data stream

SYSTEM EQUIPMENT CONFIGURATION

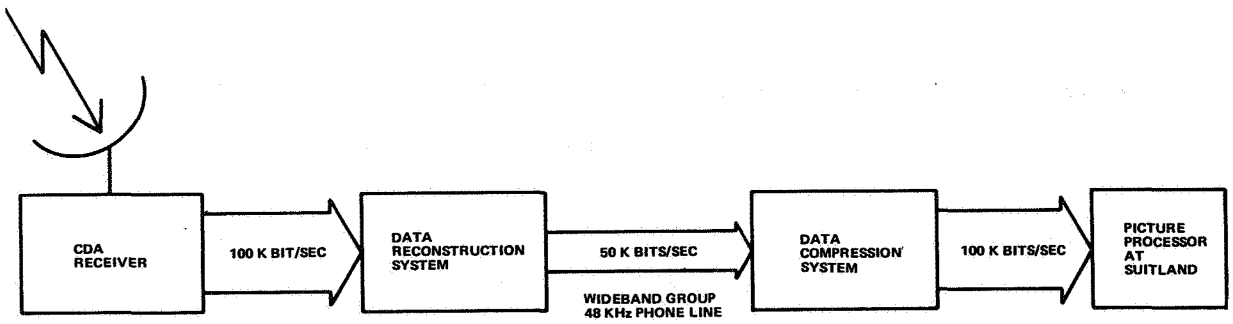


Figure 1. This data compression system provides a 100K bits/sec data transfer rate capability over a 50K bits/sec channel.

entering the data compression hardware at the Wallops CDA station. The picture processing equipment at Suitland, Maryland receives the same data quality with or without data compression, but the data rate can be doubled when data compression is used.

Before continuing, I'd like to define two terms frequently used in a discussion of data compression techniques. First, *compression ratio* is defined as the number of bits used to transmit the original or noncompressed data divided by the number of bits in the compressed data. The second term which requires definition is *entropy*: The entropy of a message is based upon its probability density function and has the units in bits. The sample entropy in bits per sample of a video image is found by summing the negative probability of a given intensity times the base-two log of the same probability over all intensities found in the empirical probability density function for the video image.

The sample entropy taken after some operation has been performed on the data tells the minimum number of bits per sample necessary to send the modified representation of the data. For example, an ATS spin-scan image quantized to six bits per sample has an entropy of 4.7 bits per sample before any other operations are performed on the data. Our goal was to achieve a two-to-one compression ratio, which implies that on the average, three bits per sample are used to encode each sample. This objective cannot be realized without employing a reversable operation on the data, which makes use of the inherent sample-to-sample dependence present in most sampled data streams. The operation selected was one which produces the difference between successive samples, because it is both easy to implement and results in a sample entropy of 2.1 bits per sample. It is now possible to construct an encoding procedure with a per scan line average of less than three bits per sample, even when synchronization and noise protection bits are added.

The algorithm used to encode the data employs the Shannon-Fano procedure. The Shannon-Fano encoding procedure shows one how to construct an encoded data stream with nearly as few bits per sample as the sample entropy for the same data. The encoding procedure produces a variable number of bits for each sample encoded. Some samples are represented by one bit, and some by as many as ten bits. Since word synchronization is maintained by correctly decoding the previous word, one must consider the effect of channel errors on word synchronization.

Two types of synchronization words are used in each image scan line. Each scan line starts with a 32-bit line synchronization word, which establishes both line and word synchronization. The 4096 data samples in each scan line are next encoded into 16 blocks with each block representing 256 data samples. At the end of each block is a 16-bit synchronization word which ensures that word synchronization is maintained and that word synchronization can be acquired within 256 data samples. Each block of 256 encoded data samples starts with a sample from the input data. The remaining 255 samples in each block consist of the encoded successive differences between input data samples. Due to the variable length coding, the number of bits between two successive synchronization words varies from 265 to 2560. Although the data compression encoding procedure selected leads to a data format where word synchronization depends on a low probability of channel error, experience has shown that the data compression communication system has a more robust synchronization characteristic than an image transmission system in which there is only one line synchronization word on a scan line and in which word synchronization is obtained from a fixed word length data format. The variable block length and hence variable line length data format produces a variable number of bits per image scan line; the number of bits depends on the redundancy in the scene at the scan line location.

In Figure 2, the image scan line number runs from 0 to 2000 on the abscissa, and the ordinate indicates the percentage of the original uncompressed line length remaining to be transmitted after data compression. The image data to which Figure 2 corresponds is an

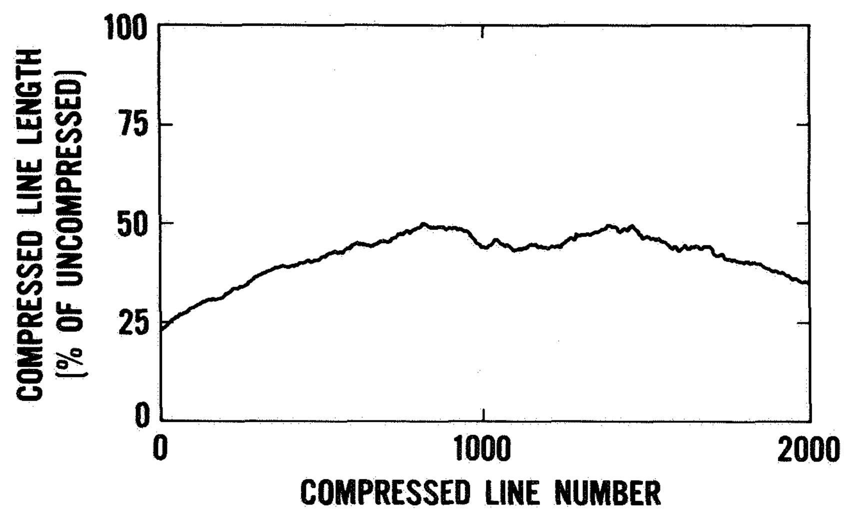


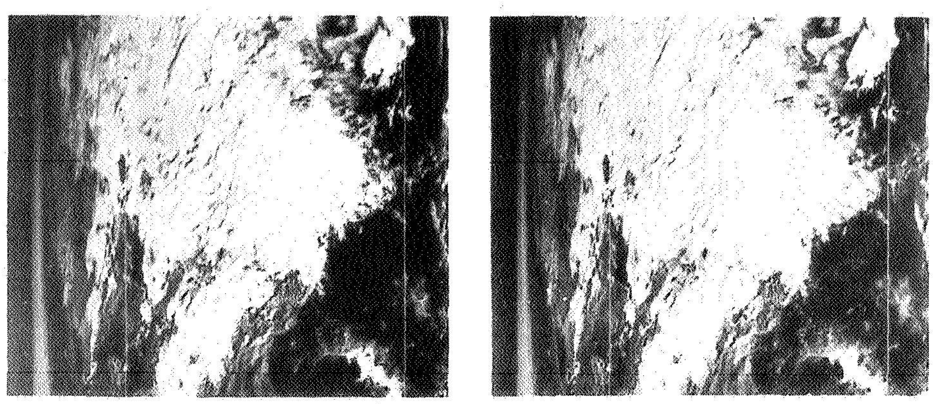
Figure 2. This graph shows the percent of data remaining after data compression on a line-by-line basis. The compression ratio varies as a function of local scene content, but on the average only half of the original number of bits need to be transmitted.

ATS spin-scan cloud cover image of the full earth disc enclosed by a black space-view surround. Figure 2 shows that more data must be transmitted in the middle of the picture than at either end, since the intersection of the scan line with the earth disc has its maximum chord length near line number 1000. However, not all of the data compression in each line is due to the constant amplitude space view, since space view accounts for only 10 percent of the uncompressed data near line 1000 and yet the compression ratio exceeds two-to-one in this area of Figure 2. The area above the curve in Figure 2 represents the increased capacity of the data compression communication system as a function of the ATS image scan line number.

Figure 3 shows a cloud cover image taken from the ITOS-D Very High Resolution Radiometer visible channel. To the left of the figure, the image is shown without data compression; to the right of the figure, the same image is shown after data compression, transmission from Wallops Island to Suitland, and reconstruction. A comparison of the data represented by these two images showed that no error was introduced into the compressed image data. In the rare cases where a channel error effect is discernible in a reconstructed compressed image, the effect shows up as a streak less than 1/16th of the image scan line length.

The results of the data compression work can be readily applied to missions where communications bandwidth or time are primary considerations. Data storage efficiency can also be increased for use onboard a spacecraft or for use with data archival systems on the ground. Data compression can also be applied to save land line costs in any spacecraft program involving a polar orbit where the Alaska CDA stations are required to send large quantities of data over great distances, as for example in the cases of ITOS, Nimbus, and ERTS. Since a 48 kilohertz wideband group from Maryland to Alaska is currently

TRANSMISSION LINK - WALLOPS TO SUITLAND



ORIGINAL IMAGE (NO DATA COMPRESSION) | **IMAGE WITH DATA COMPRESSION**

Figure 3. The original uncompressed image is the same as the compressed, transmitted, and reconstructed VHRR ITOS-D image.

rented at a cost of about $700,000 per year, a doubling of the utilization of one of these channels would be very cost effective. The Department of the Interior is currently considering an earth observation spacecraft system requiring three CDA stations. One of the three CDA stations is necessary partially because insufficient time exists to dump the spacecraft data over the other two CDA stations. In their case, data compression could double the spacecraft data dump capacity and thus reduce the number of CDA data dump stations required from three to two, a potential savings of millions of dollars.

MR. MATHEWS:

If you tolerate small errors, couldn't you make the compression ratio very much larger?

MR. KUTZ:

Yes, I'm sure that we could get better results by allowing a small error tolerance. One reason for avoiding any error in the data up to this point in time was to help the data user gain confidence in data compression. Also, some of the image data are used to perform quantative measurements. For example, the infrared VHRR data can be used to measure sea surface temperatures. Where there is a requirement for error-free data on the part of the user, we have shown that data compression can satisfy that requirement.

MR. MATHEWS:

We may have to encourage the data users to review their requirements for zero-error data to see if the zero-error constraint can't be relaxed in some cases.