NASA CASE NO. MFS-11243

PRINT FIG. Figure 6

NOTICE

The invention disclosed in this document resulted from research in aeronautical and space activities performed under programs of the National Aeronautics and Space Administration. The invention is owned by NASA and is therefore available for licensing in accordance with the NASA Patent Licensing Regulation (14 Code of Federal Regulations 1245.200).

To encourage commercial utilization of NASA-owned inventions, it is NASA policy to grant nonexclusive, royalty-free, revocable licenses to any company or individual desiring to use the invention while the patent application is pending in the U.S. Patent Office and within a specified period, presently two years, after issuance of the patent to NASA. If commercial use of the invention does not occur during this period, NASA may grant a limited exclusive, royalty-free license thereby adding an incentive to further encourage commercial development. Any company desiring to make, use, or sell this invention is encouraged to obtain a royalty-free license from NASA.

Address inquiries and all requests for licenses to Assistant General Counsel for Patent Matters, Code GP-1, National Aeronautics and Space Administration, Washington DC 20546.

AWARDS ABSTRACT

Inventor: ROBERT F. RICE

Contractor: Jet Propulsion Laboratory
Contract NAS7-100

August 23, 1971
NASA Case No. NPO-11243
~~JPL Case No. 11243~~

DATA COMPRESSION SYSTEM

The primary object of the present invention is to provide an improved data compression system, particularly adapted to compress data from a TV picture-taking system.

In the present system the information in each element, defined as a pixel, is received as a multibit word. Except for the first pixel word per scan line, which serves as a reference sample, the system (Figure 8) generates difference values in a difference converter 26. Each difference value represents the difference between a pixel word and the preceding word. A selected number of difference values represent a block. For each block a FS coder 30 generates a fundamental sequence (FS) in which each difference value is replaced by a variable length code word. Based on the length of FS and the number of difference values (or pixels) per block either FS or a coded FS, defined as Code FS or a coded complemented FS, i.e., $\overline{FS}$, defined as Code $\overline{FS}$ is selected for transmission by selector 34. When Code FS is chosen it is compared with the number of bits of the original words, used in forming the difference values for a block, which together define a Back-Up PCM. The shorter of the two compared sequences is transmitted. A control unit 22 is responsive to control signals to control the number of bits (n) per pixel word, which are used in deriving the difference values and the number of bits (k) per word which are stored in a register 28 for direct transmission.

The novelty of the invention is believed to reside in the comparison of Code FS with the Back-Up PCM and in the variation of the input word format, i.e., the number of bits (n) per pixel word, used in the compression process and the number of bits (k) which are transmitted directly.

Application Serial No. 177,753
Filed: September 3, 1971

NASA CASE NO NPO-11243

FIG. 1

N72-20154 25%

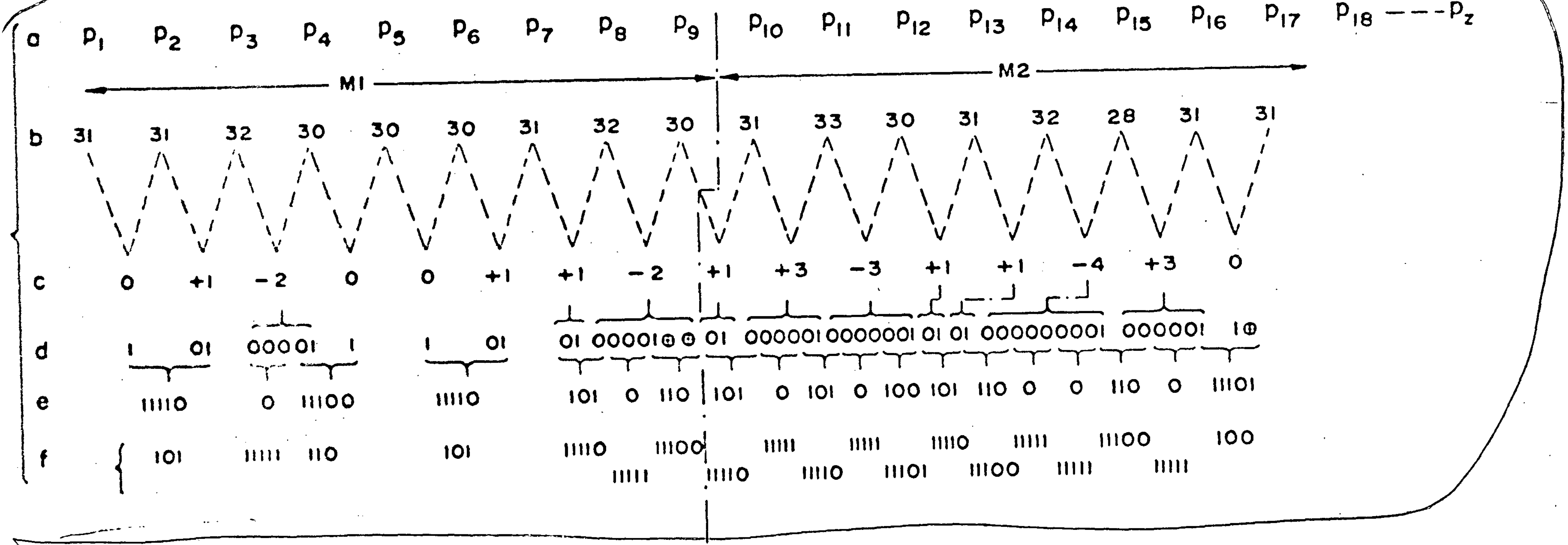

FIG. 2

DIFF. VALVE	CODE WORD
0	1
+1	01
−1	001
+2	0001
−2	00001
+3	000001
−3	0000001

FIG. 3

1J ≤ FS < 1.5J	1.5J ≤ FS < 3J	F ≥ 3J
CODE $\overline{FS}$	FS	CODE FS

ROBERT F. RICE
INVENTOR.

BY

ATTORNEYS

N72-20154

N72-20154

FIG. 5

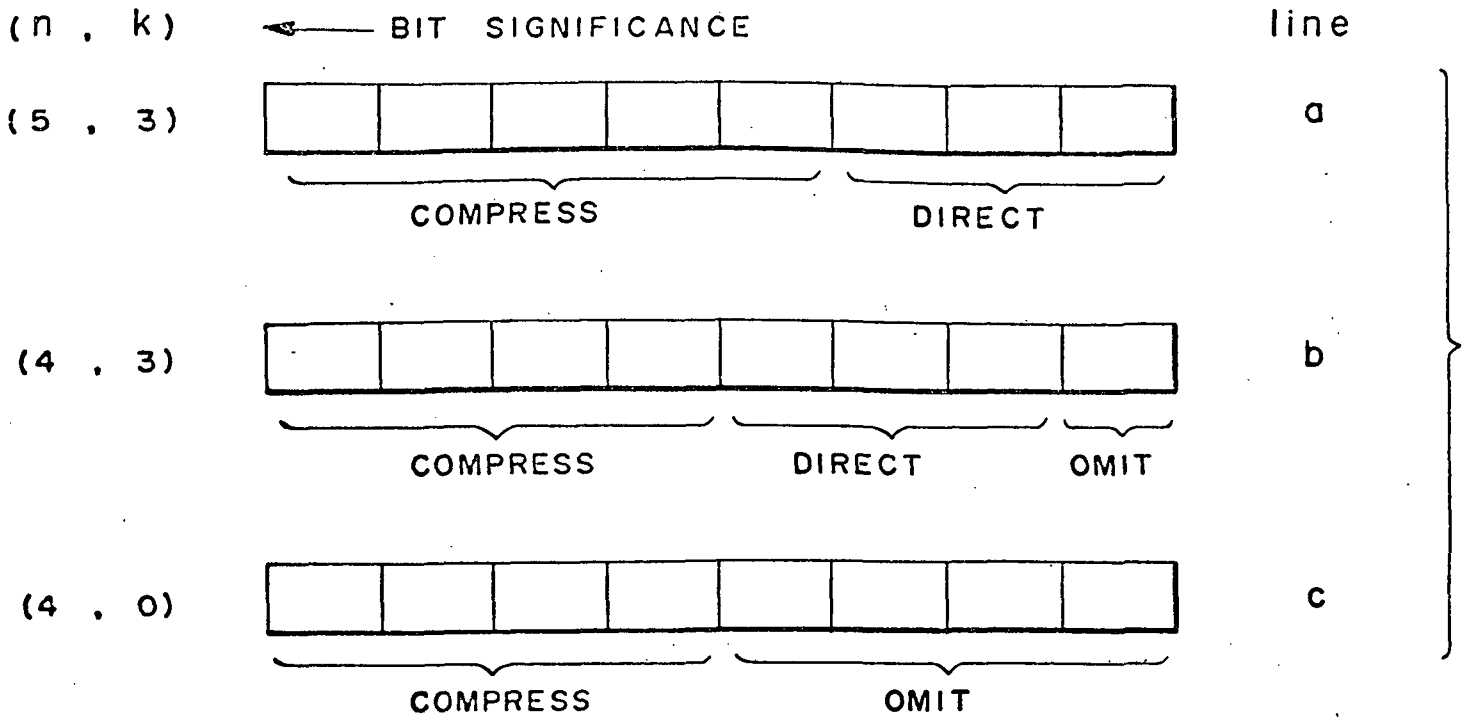

FIG. 6

N72-20154 40%

CODE	BLOCK FORMAT
BACK-UP-PCM	ID (2 bits) (1)(2)---(J) (n×J bits) (1)(2)----(J) (k×J bits)
FS	ID (2 bits) FUND. SEQ. (VARIABLE) (1)(2)----(J) (k×J bits)
CODE FS	ID (2 bits) CODE FS (VARIABLE) (1)(2)----(J) (k×J bits)
CODE $\overline{FS}$	ID (2 bits) CODE $\overline{FS}$ (VARIABLE) (1)(2)----(J) (k×J bits)

3 TUPLES OF FS	CODE WORD
0 0 0	0
0 0 1	1 0 0
0 1 0	1 0 1
1 0 0	1 1 0
0 1 1	1 1 1 0 0
1 1 0	1 1 1 0 1
1 0 1	1 1 1 1 0
1 1 1	1 1 1 1 1

FIG. 4

ROBERT F. RICE
INVENTOR.

BY

ATTORNEYS

NASA CASE NO. NPO-11193

ORIGIN OF INVENTION

The invention described herein was made in the performance of work under a NASA contract and is subject to the provisions of Section 305 of the National Aeronautics and Space Act of 1958, Public Law 85-568 (72 Stat. 435; 42 U.S.C. 2457).

BACKGROUND OF THE INVENTION

1. Field of the Invention:

The present invention generally relates to a data handling system and more particularly to a novel information-preserving data compression system.

2. Description of the Prior Art:

There are many applications in which it is necessary to compress information-representing data and thereafter use the compressed data in a manner which enables the retrieval of substantially all the original information. One such application is in space exploration, wherein scientific information, such as TV information, often has to be compressed for communication reasons. The compressed data, once received on Earth, is operated upon to reconstruct and retrieve all the original TV information. The major desired features of a data compression system includes efficiency of conversion, flexibility of operation to accommodate varying input conditions and the ability to completely reconstruct the original input information or data from the compressed data.

The use of pulse-code-modulation (PCM) in converting video signals, received from a TV camera on a space vehicle, which records terrain of a planet being explored, is well known. As the TV camera scans the terrain each picture element, hereafter referred to as a pixel is converted in a multibit word, e.g., an 8-bit word. Since the number of pixels per picture is large it is desirable to provide a system which can compress the TV data, i.e., the 8-bit words to reduce the amount of data which has to be transmitted to Earth per picture and thereby enable the transmission of a greater number of pictures with the same amount of transmitted data. However, in order not to lose valuable information it is a requisite that the original picture information can be reconstructed from the compressed data. Since the nature of the scanned terrain is variable it is desirable that the compression system be able to adapt itself to the degree of variation of the terrain, so as to maximize the retrieval information while maximizing the compression rate. Although data-compression systems have been proposed in the past, a need still exists for an improved system, which has all the aforementioned properties and which is simple and highly reliable.

OBJECTS AND SUMMARY OF THE INVENTION

It is a primary object of the present invention to provide a new improved data compression system.

Another object is to provide a data compression system with which the original information can be completely reconstructed.

A further object of the invention is to provide a new data compression system which adapts itself to the variations in the data being compressed in order to optimize the compression while preserving all information for subsequent reconstruction.

These and other objects of the invention are achieved in a data-compression system which employs variable length coding. For explanatory purposes the system will be described in connection with a TV pulse-code-modulated (PCM) system in which the video of each pixel, in the form of a multibit word, is supplied as an input to the system. The TV system is assumed to scan the terrain in a sequence of scan lines in a manner well known in the art. Except for the first pixel input in each line, the compression system produces a difference value between any pixel input and the preceding pixel input. These difference values are coded to provide a binary fundamental sequence (FS) for a given number of pixels, representing a block. Since the differences between pixels are not constant, being smallest when the pictured terrain is relatively uniform, the coding is chosen to provide a variable length FS which increases in length with terrain variations.

After the FS is generated its bit length is analyzed to determine whether it, a sequence representing a coded FS (hereafter referred to as Code FS), or a sequence representing a coded complement of FS (hereafter referred to as Code $\overline{FS}$), should be chosen. If Code FS is chosen its length is compared with the number of bits of the original input, i.e., the PCM code for all the pixels in the block. The shorter of the two compared sequences is the one selected for actual transmission.

The system further includes means whereby the number of bits per pixel word which are compressed can be changed subject to appropriate commands. One or more bits per pixel word may be omitted and the rest compressed by the system. Also some bits, such as the least significant bit or bits, may be omitted and of the remaining bits (in each pixel word) some may be transmitted directly while the rest are compressed to generate the FS, Code FS and Code $\overline{FS}$.

The novel features of the invention are set forth with particularity in the appended claims. The invention will best be understood from the following description when read in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a multiline diagram useful in explaining certain aspects of the present invention;

Figure 2 is a coding table for forming a Fundamental Sequence;

Figure 3 is a table summarizing a decision criterion employed in the present invention;

Figure 4 is another coding table;

Figure 5 is a diagram of various word formats;

Figure 6 is a chart of block formats for different codes;

Figure 7 is a line format; and

Figure 8 is a simplified block diagram of circuitry capable of executing the teachings of the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The basic principles of the present invention may be described in connection with a specific example. Therein it is assumed that for each line scan of a TV camera, the data compression system receives a succession of multibit words. Each word represents the video from a different pixel. Thus the word may be referred hereafter as the pixel word. It is assumed that each word includes 8 bits and that the number of pixel words per line is 1+8M, where 8 represents the number of pixels per block and the M the number of blocks per line. The first pixel word in a line represents a reference word which is transmitted without compression. Each subsequent pixel word is compared with the previous word and a difference value is derived by the compression system.

In Figure 1, line b, the numbers from left to right represent quantization levels of a sequence of pixel words. The pixel words are designated in line a as p_1 through p_z, where p_1 is the first reference word. The difference between pixel words p_2-p_9 and their preceding words are used to form block M1, while the differences between pixel words p_{10}-p_{17} and those preceding them are used to form block M2, etc. In line c, the sequences of the difference values for blocks M1 and M2 are shown. Each difference value is coded to provide a fundamental sequence (FS). The code is chosen so that the most probable difference values are coded to produce the smallest number of bits in the FS. A code table for coding the difference values of the specific example is shown in Figure 2. As seen a zero (0) difference value is coded by a 1, while +1 and -1 difference values are coded by 01 and 001, respectively. Larger differences are coded by longer code words. As seen the code is chosen so that each code word ends with a binary 1 preceded by X zeros where X=0, 1, 2, 3, etc.

In Figure 1, line d, the FS's for M1, M2, using such coding are diagrammed. Once the FS is generated for a block a decision is made whether the FS should be selected for transmission or whether a coded FS (referred to as Code FS), or the coded complement of FS, (referred to as Code $\overline{FS}$) should be selected. The decision is made based on the bit length of FS as a function of the number of pixels per block, which is eight in the present example.

The decision criterion in terms of FS length is charted in Figure 3. It is seen that if $1J \leq FS < 1.5J$, i.e., as long as the length of FS is less than 1.5 times the number of pixels per block, designated J and equal or greater than 1J, Code $\overline{FS}$ is chosen. On the other hand if $1.5J \leq FS < 3J$, FS is chosen and when $FS \geq 3J$, Code FS is chosen.

Basically Code FS is generated by coding each n-tuple of FS with a code word. Figure 4 is a coding table wherein the left-hand column lists various 3-tuples of FS and the right-hand column their corresponding code words. Code $\overline{FS}$ is generated by complementing FS and coding its 3-tuples with the proper code words. In Figure 1, lines e and f, Code FS and Code $\overline{FS}$ for blocks M1 and M2 are diagrammed. It should be stressed that when the FS is coded by 3 tuples, unless it is divisible by 3, dummy zeros represented by ⊕ are added thereto. In the FS of M1 two dummy zeros are added and in the FS of M2 only one is needed.

Based on the decision criterion, summarized in Figure 3, for block M1 since FS is 19 bits long and since the number of pixels/block is 8, 1.5<FS<3J. Therefore FS is selected for transmission and Code FS or Code $\overline{FS}$ need not be generated. However, for block M2, FS is 35 bits long and consequently FS>3J. Therefore, Code FS is generated and is selected for transmission. It should be pointed out that the original PCM code consisting of the 8-bit words forming a block is 8J bits long. Thus whenever FS or Code $\overline{FS}$ is chosen it is apparent that the chosen code is shorter than the PCM code. However, this may not be the case when the Code FS is chosen since the latter is chosen whenever FS>3J and therefore the resulting Code FS may be longer than the original PCM code. To prevent a code longer than the original PCM code from being transmitted in accordance with the present invention, whenever Code FS is chosen and generated before being transmitted, its bit length is compared with the bit length of the PCM code, which in the present example is 64 bits long since each pixel word consists of 8 bits and there are 8 pixels per block. Only if the Code FS is shorter than the PCM code, hereafter also referred to as the Back-Up PCM, is the Code FS transmitted. Otherwise the Back-Up PCM is transmitted.

Summarizing the foregoing description in accordance with the teachings of the present invention, except for the first pixel word in each scan line, which is transmitted directly, for each successive pixel word a difference value is produced. Except for the first pixel word in a line, the rest are divided into equal length blocks. The difference values for each block are coded to provide a block FS code. Based on the length of the block FS code, and the number of pixels per block, a decision is made whether the FS code is selected for transmission or whether a coded FS, defined as Code FS, or a coded FS complement defined as Code $\overline{FS}$ should be selected. When the Code FS is selected and generated its length is compared with the length of the Back-Up PCM. Only when the former is shorter than the latter is the selected Code FS transmitted. Otherwise the Back-Up PCM is transmitted.

Herebefore it has been assumed that the compression system responds to all the bits of each pixel word. It has been discovered that under some conditions substantially all the useful TV information can be reconstructed by compressing less than complete pixel words. It has been discovered that under some conditions substantially all the useful TV information can be reconstructed by compressing less than complete pixel words. Briefly based on actual performance it has been discovered that sometimes the least significant bit or bits of each pixel word change due to noise or other conditions rather than due to terrain information. In such a case it is desirable to omit one or more of the least significant bits of each pixel word and operate on the rest of the bits. It has also been discovered that under some conditions it is desirable to supply of each pixel word only some of the most significant bits and of the remaining bits transmit all or some of them directly without compression.

This aspect of the invention may best be explained in connection with Figure 5 wherein in lines a, b and c different formats of an 8-bit pixel word are shown. Therein n defines the number of bits per word which are compressed and k defines the number of bits per word which are transmitted directly. In line a it is assumed that of each pixel word 5 bits are compressed and 3 bits are sent directly. None of the bits are omitted. In line b it is assumed that 4 bits are compressed, 3 bits are sent directly while 1 bit is omitted altogether. In line c it is assumed that 4 bits are compressed and 4 bits are omitted. None is transmitted directly. In such applications the Back-Up PCM includes the n bits of each pixel word in the block.

While Figure 5 is a diagram of different pixel word formats any one of which may be supplied to the system, Figure 6 is a diagram of transmitted block formats for the 4 possible codes. The first 2 bits of each block format are used for identification (ID), i.e., to identify the particular code used for the block. These two bits are followed by a variable length sequence. In the Back-Up PCM the block includes n x J bits where n is the number of bits of each pixel word used in the compression process. These are followed by k x J bits which represent the k bits which are selected from each pixel word in the block for direct transmission. In the Code $\overline{FS}$ mode the variable length Code $\overline{FS}$ is transmitted followed by the k bits of each of the J pixels in the block which are chosen for direct transmission. Similarly the FS or Code FS is transmitted followed by the k bits of each of the pixels in the block which are chosen for direct transmission.

While the particular code which is transmitted for each block may vary from block to block, the word format which is operated upon is generally the same for a complete picture. If desired it may be changed from line to line. However, for any one line the word format is the same. Figure 7, to which reference is now made, is the transmitted format for each line. It starts with 4 bits, designated SP, defining the word format for the line. These are followed by the (n+k) most significant bits of the first pixel word in the line, which serves as the reference sample. These bits are followed successively by the block formats of the blocks in the line which are designated L(1) through L(M). Thereafter an End of Line ID is included to designate the line in the picture.

It should be appreciated that various combinations of elements and circuits may be employed to practice the teachings herebefore described. Thus the invention is not intended to be limited to a specific circuit arrangement but rather to any means capable of performing the various functions, herebefore described. However, for purposes of completeness a simple block diagram of a generalized circuit arrangement is shown in Figure 8 to which reference is now made. Therein numeral 20 designates a pixel word input unit which is assumed to receive a multibit word per pixel. Unit 20 is under the control of a control unit 22. The latter receives an external word format command and in response thereto supplies a word format control signal to unit 20. In response to this signal the input unit 20 automatically routes the proper number of bits (n) for compression, the proper number of bits (k) which are to be transmitted directly and the number of bits which are to be omitted.

Based on the external word format command, unit 22, which also receives a line start signal supplies the 4 bit SP to an output unit 24. It also activates unit 20 to supply the bits of the first pixel word in a line which are chosen for compression and direct transmission, i.e., bits n + k, as the reference sample. The n bits of each pixel word which are to be compressed are supplied to a difference converter 26 which provides the difference value between successive pixel words. The k bits per pixel word which are to be transmitted directly are stored in a register 28.

The difference values from converter 26 are coded into the FS by a FS coder 30 whose output is the FS. The bits to be compressed representing the original PCM data or the Back-Up PCM are held in a register 32. The length of FS is then analyzed by a code selector 34. Basically it determines on the basis of the length of FS and J the number of pixels per block, which of FS, Code FS and Code $\overline{FS}$ should be chosen. If FS is chosen selector 34 merely recites the FS from coder 30 to output unit 24. If Code $\overline{FS}$ is chosen selector 34 activates an

8 word coder 36 to respond to the output of an inverter 38 which complements FS to provide Code $\overline{FS}$. Coder 36 is designated as an 8 WORD CODER since when 3-tuples are coded the maximum number of different 3-tuples is 8, as shown in Figure 4. Code $\overline{FS}$ is directly routed by code selector 34 to unit 24. If Code FS is selected the code selector 34 activates the coder 36 to respond to FS directly in order to provide Code FS. However, once it is produced its length is compared in the code selector with the Back-Up PCM in register 32. Only when Code FS is shorter than the Back-Up PCM is the former routed to unit 24. Otherwise Back-Up PCM is sent to unit 24. After the compressed data for a block is received by unit 24 the k x J bits in register 28 which represents the k bits of the pixels chosen for direct transmission are supplied to unit 24, before the pixel words of the next block are operated upon. It is the output of unit 24 which is sent to a transmission system for transmission to Earth.

It should be appreciated that the circuitry of Figure 8 may be replaced by a computer properly programmed to perform the various operations herebefore described. Such a computer would be supplied with the pixel words a line start signal and a word format command, the latter defining the number of bits per pixel word which should be compressed, the number of bits to be transmitted directly and any bits to be omitted. The computer would then perform the various functions and would load its output unit with a sequence of bits in the line format, shown in Figure 6.

Herebefore it was assumed that the word format for each line is controlled by an external word format command which was assumed to be supplied to unit 22. Neglecting omitted bits, which in accordance with the present invention are assumed to be controlled by an external command, it has been discovered that improved performance may be obtained by controlling the word format for each line as a function of the number of compressed bits, produced per pixel in a preceding line.

Again neglecting omitted bits, for an 8-bit pixel word the possible word formats are 8,0; 7,1; 6,2; 5,3; and 4,4, where the first number in each pair is n, representing the number of compressed bits, and the second number is k, representing the bits which are to be transmitted directly. Defining the word format or mode for any line as (n,k) and defining L_n as the number of compressed bits per pixel, it has been discovered that the preferred word format for the next line is one which depends on the magnitude of L_n as shown in the following Table:

Present Word Format=(n,k)	Next Line Word Format
$L_n<3$	(n+1, k-1)
$3\leq L_n<4$	(n,k)
$L_n\geq 4$	(n-1, k+1)

It is thus seen that when L_n is less than 3, the word format is changed for the next line by increasing the number of compressed bits by 1 to n+1 and by decreasing the number of directly transmitted bits by 1 to k-1. On the other hand if $L_n\geq 4$ on the next line the number of bits per pixel word which are compressed is decreased by one and the number of directly transmitted bits is increased by 1. Otherwise, the word format for the next line is not altered from that used for the preceding line.

In practice these teachings can easily be implemented by counting the number of compressed data bits which are routed to unit 24 for each line. At the end of the line, this number is divided by the number of pixels per line (excluding the first) to derive L_n. Once L_n is known, a simple decision is made which of the possible three word formats, i.e., (n,k); (n+1, k-1); or (n-1, k+1) should be used for the next line where n,k represents the word format used for the preceding line. These functions are assumed to be performed by the word format selector 40 shown in Figure 8. It should be apparent that various known design techniques may be used in the implementation of selector 40, designed to derive L_n and on the basis of its magnitude send a 2-bit signal to control unit 22 to select one of the three possible word formats with respect to the word format of the preceding line.

Although particular embodiments of the invention have been described and illustrated herein, it is recognized that modifications and variations may readily occur to those skilled in the art and consequently it is intended that the claims be interpreted to cover such modifications and equivalents.