

N 9 3 - 2 5 9 2 9

LOSSLESS COMPRESSION OF IMAGE DATA PRODUCTS ON THE FIFE CD-ROM SERIES

Jeffrey A. Newcomer
Hughes STX Corporation
4400 Forbes Blvd.
Lanham, Maryland 20706-9008

Donald E. Strebelt
VERSAR Incorporated
9200 Rumsey Road
Columbia, Maryland 21045

Abstract. How do you store enough of the key data sets, from a total of 120 gigabytes of data collected for a scientific experiment, on a collection of CD-ROMs, small enough to distribute to a broad scientific community? In such an application where information loss is unacceptable, lossless compression algorithms are the only choice. Although lossy compression algorithms can provide an order of magnitude improvement in compression ratios over lossless algorithms, the information that is lost is often part of the key scientific precision of the data. Therefore, lossless compression algorithms are and will continue to be extremely important in minimizing archiving storage requirements and distribution of large earth and space (ESS) data sets while preserving the essential scientific precision of the data.

Data preservation, distribution, and archiving were integral and essential elements for the satellite, aircraft, and field data collected by the FIFE (First ISLSCP¹ Field Experiment) in 1987 through 1989 over the Konza Prairie area near Manhattan, Kansas. In total, the raw and the derived data products comprise approximately 120 gigabytes of which the image data comprise over 99%. An important element of the planned final archive is a set of CD-ROMs which will contain a reduced data set felt to satisfy the primary objectives of the experiment. In order to store as much key image data as possible on the CD-ROMs and to preserve the scientific precision of the data, a lossless compression algorithm was devised. Use of the algorithm on AVHRR-LAC, Landsat TM, SPOT, NS001, and ASAS image products has resulted in an average compression ratio of 2:1 with ancillary and supporting files of information having ratios as large as 16:1.

The ordered steps of the algorithm include : 1) normalization of the columns of the data matrix, 2) normalization of the rows, and 3) viewing the remaining values in each row as layers of bits that are either runlength encoded or packed back into bytes depending on the calculated average runlength of the line. The compression package, containing functions for processing 8, 16, and 32 bit values and processing control, is written in C and is operational on a VAX computer system in the Laboratory for Terrestrial Physics at NASA GSFC. A complementary set of image restoration software was developed and can be run on a wide range of computing platforms from PCs to workstations.

We envision this intuitive compression algorithm to be useful on a broad range of spatial data sets including gridded layered modeling data sets such as terrain, spectral, and meteorological variables that will be required for coordinated earth systems field experiments during the next decade.

1. Introduction

Distributing large amounts of scientific data such as satellite and aircraft imagery, mandates some form of data compression to minimize storage and data transmission. Although lossy compression algorithms can provide compression ratios of up to 100:1, the precision that is lost

¹International Satellite Land Surface Climatology Project

is often part of the key scientific information of the data. Therefore, lossless compression algorithms are and will continue to be extremely important in dealing with the distribution of large scientific data sets and in minimizing archival storage requirements. The importance of lossless compression for earth and space science (ESS) and space physics data are addressed by Walker [1].

The First ISLSCP Field Experiment (FIFE) collected a coordinated data set useful for developing and validating models that determine surface climatology from satellite-acquired measurements. Particular attention was placed on measurements of the mass and energy balances at the boundary between the land surface and the atmosphere and on the role of the existing surface biology in controlling the land and atmosphere interactions. The experiment was also designed to explore the use of satellite observations to infer climatologically important land-surface parameters related to the land/atmosphere interactions (see [2] for further details).

The 15 x 15 km FIFE study area included the Konza Prairie Research Natural Area, an NSF Long Term Ecological Research (LTER) site near Manhattan, KS, and surrounding areas. The data collection effort consisted of long-term monitoring through satellite observations and periodic in-situ ground and meteorological measurements along with five intensive field campaigns of 10 to 20 days in length during the growing season of 1987 and late summer in 1989. Attempts were made during the IFC's to obtain coordinated (and at times simultaneous) ground, air, and satellite measurements of hundreds of variables.

Integral to this whole effort was a data system that served as a tool for designing the experiment and for organizing, manipulating, and archiving the complex data set. A dedicated and remotely accessible data system was developed at NASA's Goddard Space Flight Center (GSFC) to meet the data management requirement. Data compression has become a critical concern as the data system prepares the final CD-ROM based archive of the reduced data collection (about 10 Gbytes). To date, one prototype CD-ROM has been produced to evaluate the CD-ROM data publication facilities available at GSFC and to establish operational procedures and software for the final CD-ROM series production (see [3]).

2. FIFE Information System

The fundamental mission of the FIFE Information System (FIS) was to capture, preserve, organize, distribute, and archive the satellite, aircraft, and field data that were collected. During the experiment and to date the image and other large data sets have been stored and distributed on magnetic tape while the smaller data sets have been stored in an on-line data base and distributed via electronic transfer and floppy disk (see [4] for further details). In total, the raw and the derived data products now comprise approximately 120 gigabytes.

The non-image data consist of conventional meteorological and radiosonde data from all NOAA stations within 1° latitude and longitude of the study area, automated meteorological and radiation sensors reporting up to 20 variables every 5 minutes at 32 sites, and in-situ data assembled by the 30 investigator teams during the IFC's. These data sets characterize the diurnal patterns of radiation, moisture and heat flux, atmospheric properties and temperatures, moisture and wind profiles, vegetation and soil condition, and rates of photosynthesis and evaporation. Although overwhelming in diversity, these data comprise less than 1% (300 Mbytes) of the total data volume. The image data on the other hand occupy over 99% of the data volume but consist of only approximately 12 types with differing levels of processing (Table 1).

A final data archive is currently being established. An important element of the archive plan is publication of a set of some 6 CD-ROMs which will contain a reduced data set felt to satisfy the primary objectives of the experiment. The bulk of this data (90⁺%) will consist of derived (i.e.,

level-1 and level-2) image products from satellite and aircraft based instruments with spatial resolutions from 20 m to 1 km. The need for lossless compression of the image data products arose from the use of the CD-ROMs as the main FIFE data archival product. Dozier and Tilton [5] note that "Since the purpose of the archival process is to keep an accurate and complete record of data, any data compression used in an archival system must be lossless, and protect against propagation of error in the storage media." We feel lossless compression of data stored on CD-ROM media satisfies these criteria quite well.

Table 1: Volume of Selected Digital Image Data (Mbytes)²

Image Type	Processing Level			
	Level-0	Level-1	Level-2	Level-3
ASAS	-	3958	-	-
AVHRR-GAC	7406	272	-	-
AVHRR-LAC	67583	1686	812	-
Landsat TM	2157	210	2	-
NS001 (TMS)	19201	1210	24	-
PBMR	-	-	37	38
SPOT-XS	1998	462	-	-
SPOT-PAN	198	144	-	-
TIMS	<u>10775</u>	<u>-</u>	<u>18</u>	<u>-</u>
Totals:	109318	7942	893	38

3. FIS Image Data Products

To be successful, FIFE required image data products that allowed integration of multi-date, multi-sensor, and multi-resolution imagery for quantitative atmospheric and surface biophysical radiometric studies. In particular, mandatory capabilities were to accurately derive, extract, and utilize geographic location and solar and viewing conditions in conjunction with instrument calibration information for derivation of surface radiance and reflectance on a polygonal or per-pixel basis. In FIFE, these outputs were required from the pixel to the regional level as input to models of the surface energy/mass exchange processes.

Based on this set of requirements and the need to distribute the data to a large and diverse user community, a processing approach and operational processing software were developed to handle the task of processing the large volume of level-0 data into complete and user-friendly image products (see [6] for further details). Level-0 image data was defined as unmodified (but possibly reformatted) instrument values as received from the agencies supplying the data. The definitions of the level-0 and higher data products (i.e., level-1, level-2, ...) were derived from and are consistent with the EOS (Earth Observing System) formulated definitions (see [7]).

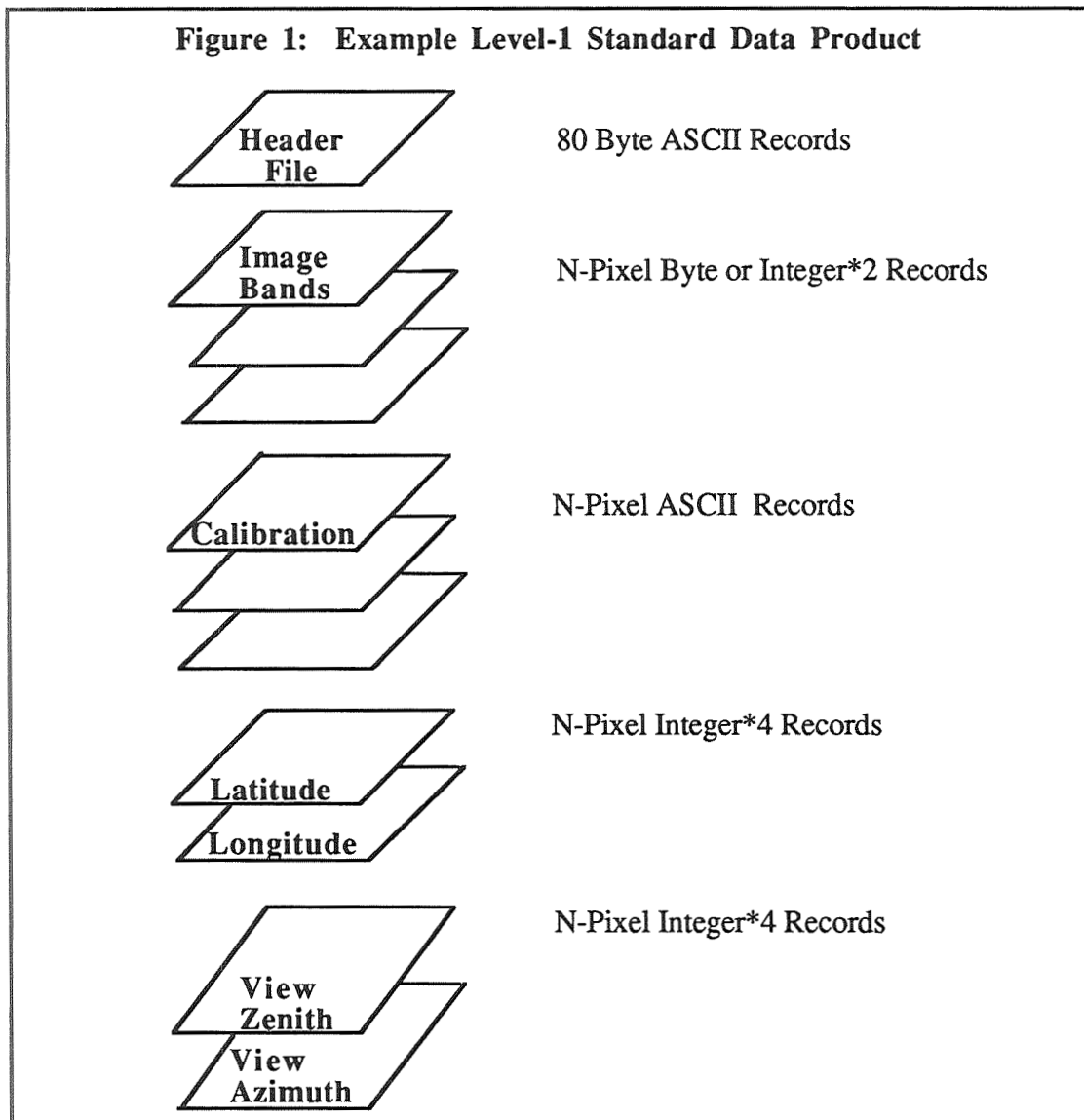
Of particular concern to the CD-ROM publication and data compression efforts was the content and format of the image products to be published; primarily the level-1 and level-2 products. An example of a generic level-1 FIFE image product is shown in Figure 1. The level-1 products were closely reviewed since they were required to contain all the information necessary to derive at-sensor radiance values and were anticipated to be (and resulted in being) the primary image-data distribution product. Each level-1 product contains a 'header' file consisting of 80 byte ASCII text records that describe the overall contents of the image product, summaries of any necessary calibration, georeferencing, viewing, and solar position information, and comments related to the processing performed. The header file is then followed by a series of files containing the spectral image data, any unpacked and reformatted calibration information, and

²See Appendix A for list of acronyms

georeferencing, view angle, and solar angle information. The georeferencing and view and solar angle files contain values for the respective variables on a per pixel basis for the selected subset of image data. In content, the FIFE level-1 image products are very similar to the level-1 MODIS (Moderate Resolution Imaging Spectrometer) products described by Salomonson [8]. All of these files except for the ASCII header file (which is simply copied) are processed by the data compression software. The ASCII header files were not compressed to allow users to easily review the summary information of a given image and to determine if the image met their particular needs before proceeding with the restoration of the data. In addition, each ASCII header file occupied less than 19K bytes of the total image product volume.

4. Image Data Compression

The need for a lossless compression procedure arose from the desire to store as much image data as possible on the planned CD-ROMs and to preserve the scientific precision of the data. From our experience, it seemed that a useful and intuitive algorithm could be developed to perform the needed compression.



4.1 Algorithm Evolution and Performance

We started with the observation that adjacent pixels in an image frequently have quite similar values (especially in a given line or column). Often too, the range of digital values in an image is only a small fraction of the total data range. Determining an overall image minimum value and subtracting it from all the pixels in the image would therefore maintain the range and distribution of values while bringing the bit representations of the pixel values into the lowest set of bits. In addition, the quantized spectral responses of adjacent pixels (in the form of digital counts or DN's) do not usually differ by more than a few counts. Essentially, the high-order bit layers represent the major structural information of the image and do not change much from one pixel to another. On the other hand, the low-order bits are quite variable between adjacent pixels and can tend to be rather random.

It seemed then that runlength encoding of the high-order bits of the pixels in each line could result in significant compression. Depending on the smoothness of a given image, runlength encoding of the low-order bits may also result in good compression. So, the algorithm was developed to evaluate each series of bit values in a given bit plane across a given image line for runlength encoding. In addition, a threshold was needed to determine whether or not runlength encoding would be worthwhile for a particular line of image data and an alternate storage scheme was needed to process the lines that were not runlength encoded. The decisions on the initial threshold value and what to do with the records that were not runlength encoded resulted from the fact that the first images being considered for compression (level-1 AVHRR-LAC) contained 256 pixels in each line and the runlength counts were being stored in single byte (8 bit) values. If the average runlength for a given bit plane across the pixels in a particular line was greater than 8, then the data would require less space to be stored in a series of 8-bit runlength counts. If the average runlength count was less than or equal to 8, no storage savings could be realized, and the bits were packed into bytes and stored as a series of byte values. Actually, if the average runlength count equalled 8, both encoding formats stored the data equally well, but no storage savings resulted (see Appendix B for details of the original compressed file record structure). This type of bit plane encoding is discussed at some length by Rabbani and Jones [9].

Results of using this first version of the algorithm on a set of 97 FIFE level-1 AVHRR-LAC image products is given in Table 2. The compression ratios for the spectral bands (B1 through B5) ranged from 1.6:1 to 53:1. The values of 16:1 and 53:1 for B1 and B2 were achieved on a night image in which the images for B1 and B2 (which are visible and near infrared bands, respectively) were very smooth with only a small amount of random noise. For the remaining files, the best compression was achieved in the view azimuth (Vaz), view zenith (Vzen), solar azimuth (Saz), and solar zenith (Szen) files. An overall average compression of 2.7:1 was achieved in terms of actual storage space required. This result was promising in that it would allow us to place 270% more AVHRR-LAC imagery on the CD-ROM disc than if we had not compressed the data. The 'bit average' compression ratios in Table 2 (and subsequent tables) are smaller than the 'average' ratios in that they are weighted compression statistics calculated with the actual number of bits in the data rather than the number of bits required for storage. For example, although B1 was stored in a 16 bit value, the actual data only occupied a 10 bit range. Therefore, the bit average ratio of 2.3 for B1 equals the average ratio (3.6) multiplied by the actual bits divided by the storage bits (i.e., 10/16).

The next development in the algorithm resulted from testing the existing software on a new data set, level-1 NS001 TMS (Thematic Mapper Simulator) data. The level-1 NS001 data product consists of 8 spectral bands, 8 associated files of record-by-record housekeeping and calibration information, files of latitude and longitude, and 2 files containing one record each of view azimuth and zenith information for the given flight line. Storage of a nominal level-1 NS001 image product required 8.6 Mbytes for spectral data, 1.5 Mbytes for housekeeping information,

8.6 Mbytes for latitude and longitude information, and 5.5 Kbytes for view angle information. Table 3 shows the results from testing the initial compression scheme on the NS001 data. The highest compression from this original ('old') algorithm was achieved on the latitude, longitude, and view azimuth files. Although the overall compression ratio of 1.5:1 resulted in a space saving of 6.2 Mbytes per image product, we felt we could improve this by more closely evaluating the characteristics of the NS001 and other image data product files.

Table 2: Initial AVHRR-LAC Compression Results

	<u>B1</u>	<u>B2</u>	<u>B3</u>	<u>B4</u>	<u>B5</u>	<u>Anc</u>	<u>Lat</u>	<u>Lon</u>	<u>Vaz</u>	<u>Vzen</u>	<u>Saz</u>	<u>Szen</u>
Storage Bits	16	16	16	16	16	8	32	32	32	32	32	32
Actual Bits	10	10	10	10	10	7	25	26	17	16	17	16
Max Ratio	16	53	2.5	2.8	3.2	1.5	2.2	1.8	6.6	4.0	6.8	6.7
Min Ratio	1.8	1.8	1.6	1.8	1.8	1.4	1.8	1.6	4.8	3.8	4.5	4.6
Avg Ratio	3.6	3.4	1.9	2.3	2.3	1.5	2.1	1.7	5.7	3.8	5.6	5.6
Bit Avg Ratio	2.3	2.1	1.2	1.4	1.4	1.3	1.6	1.4	3.0	1.9	3.0	2.8
Overall Average	2.7:1											
Overall Bit Average	1.7:1											

Table 3: NS001 Compression Results

	<u>Bands 1 - 8</u>	<u>Housekeeping Files 1 - 8</u>	<u>Lat</u>	<u>Lon</u>	<u>Vaz</u>	<u>Vzen</u>
Storage Bits	8	8	32	32	32	32
Actual Bits	8	7	25	26	17	16
Old Ratio	1.1	0.7	3.4	3.4	4.4	2.4
New Ratio	1.2	1.3	5.5	6.4	1.0	1.0
Old Bit Ratio	1.1	0.7	1.8	3.6	2.2	0.4
New Bit Ratio	1.2	1.1	4.3	5.2	0.5	0.5
Overall Old Ratio	1.5:1					
Overall New Ratio	2.1:1					
Overall Old Bit Ratio	1.3:1					
Overall New Bit Ratio	1.8:1					

In the spectral bands the response of the pixels in each column of the NS001 imagery is differentially affected by limb brightening caused by the large across-track scan angle of the instrument. Strong column similarities existed also in the latitude or longitude file (depending on the flight direction) and in the housekeeping files which are series of column formatted ASCII records containing the timing, calibration, and instrument-status information. Corresponding column-element similarities exist also in data collected by linear-array (push-broom) scanners. Based on this, it was decided to take advantage of these column similarities by determining the minimum value of each data column, storing these values as part of the compressed file structure, and subtracting the minimum column value from each element in the column. This resulted in a more 'column normalized' data matrix and the subtraction of the overall image minimum was no longer needed.

The values in each record of the georeferencing, view-angle, and solar-angle files for the image products were also quite similar depending on the orientation of the image. Although not applicable to the NS001 data, other remote sensing instruments (like Landsat TM) gather 'n' lines

of data in each across-track scanning pass with a set of n detectors. Since each detector has a somewhat different response, consecutive lines of data over the same target can contain somewhat different values. It was assumed that these line-to-line differences would be somewhat 'independent' of the column similarities and would still exist in the 'column normalized' data matrix. Determining, storing, and subtracting the row minimum values was used to remove these row similarities and to bring the data down into a 'smallest' number of bits for subsequent encoding.

The results of applying this refined algorithm on the NS001 data are given as the 'new' values in Table 3. Although only incremental improvements were realized in compressing the spectral-image files, and the view-azimuth and view-zenith files remained at their original volumes, an overall improvement in compression of 20% was realized for the whole image product. The decrease in the compression of the view angle files is due to the fact that each file contains only one record of information and the additional 'overhead' storage required to store the column minimum values which was not present in the old version. Since these files comprise less than 0.03% of the data volume, we felt this was a small price to pay for maintaining the generality of the compression overall. If the files contained a record for each image line as did the level-1 AVHRR-LAC data, we believe they would compress at least as well as the view angle files in the AVHRR-LAC data. As is evident from the values in Table 3, the real savings came from the housekeeping and latitude and longitude file compression which improved by 66% and 12%, respectively. The small improvement in the spectral data compression and the large improvement in compressing the latitude, longitude, and housekeeping files (which comprised a significant part of the data product volume) resulted in the 20% improvement. Overall, a 2.1:1 compression was realized in relation to actual data storage space requirements.

After testing the improved algorithm on the NS001 data, we went back to check it on the level-1 AVHRR-LAC data. Table 4 contains the results of applying the updated algorithm to the 707 level-1 AVHRR-LAC data products which were processed for CD-ROM publication. The largest improvements in the compression again came from the view angle, solar angle, and latitude and longitude files. Overall an additional 8% improvement in storage space was realized. Although this seems small, it resulted in saving an additional 40 Mbytes (about 6.7% of the total CD-ROM volume) in storing the level-1 AVHRR-LAC data. This allowed all the level-1 satellite monitoring data (i.e., AVHRR-LAC, Landsat TM, and SPOT) from 1987 through 1989 to be archived on a one rather than two CD-ROMs.

Table 4: Updated AVHRR-LAC Compression Results

	<u>B1</u>	<u>B2</u>	<u>B3</u>	<u>B4</u>	<u>B5</u>	<u>Anc</u>	<u>Lat</u>	<u>Lon</u>	<u>Vaz</u>	<u>Vzen</u>	<u>Saz</u>	<u>Szen</u>
Storage Bits	16	16	16	16	16	8	32	32	32	32	32	32
Actual Bits	10	10	10	10	10	7	25	26	17	16	17	16
Max Ratio	64	85	2.7	3.1	3.1	1.5	4.4	8.4	43	102	22	24
Min Ratio	1.7	1.7	1.4	1.3	1.3	1.5	2.3	2.0	6.8	7.6	6.0	4.6
Avg Ratio	3.6	3.4	1.9	2.3	2.3	1.5	3.1	2.4	8.6	33	8.5	18
Bit Avg Ratio	2.3	2.1	1.2	1.4	1.4	1.3	2.4	2.0	4.6	16	4.5	9.1
Overall Average	3.5:1											
Overall Bit Average	2.2:1											

Tables 5, 6, and 7 contain the results of applying the algorithm to Landsat Thematic Mapper (TM), SPOT multispectral (XS), and SPOT panchromatic (PAN) 'browse' image products. Note that the actual level-1 image products could not be distributed without infringing on EOSAT or CNES copyrights. Therefore, the data were spatially degraded by averaging the original pixels

within a 2 by 2 moving window. The resultant degraded Landsat TM, SPOT-XS, and SPOT-PAN images were termed as 'browse' products with spatial resolutions of 60, 40, and 20 meters, respectively. Compared to the original level-1 imagery, these would be considered to be lossy data products. The largest compression ratios were achieved in the latitude files, with the longitude files also compressing significantly. The spectral bands themselves were compressed an average of 1.4:1. This is lower than the AVHRR-LAC results but higher than the NS001 spectral data compression. The increase over the NS001 data which is of similar spatial resolution and spectral content is likely due to the averaging done in degrading the original data to derive the browse products.

Table 5: Landsat TM Compression Results

	<u>B1</u>	<u>B2</u>	<u>B3</u>	<u>B4</u>	<u>B5</u>	<u>B6</u>	<u>B7</u>	<u>Lat</u>	<u>Lon</u>
Storage Bits	8	8	8	8	8	8	8	32	32
Actual Bits	8	8	8	8	8	8	8	25	26
Max Ratio	1.5	1.8	1.5	1.3	1.1	1.7	1.3	16	18
Min Ratio	1.3	1.5	1.2	1.1	1.0	1.2	1.1	15	5.9
Avg Ratio	1.4	1.6	1.3	1.1	1.0	1.4	1.2	16	7.3
Bit Avg Ratio	1.4	1.6	1.3	1.1	1.0	1.4	1.2	13	5.9
Overall Avg	2.4:1								
Overall Bit Avg	2.1:1								

Table 6: SPOT-XS Compression Results

	<u>B1</u>	<u>B2</u>	<u>B3</u>	<u>Lat</u>	<u>Lon</u>
Storage Bits	8	8	8	32	32
Actual Bits	8	8	8	25	26
Max Ratio	1.9	1.9	1.5	21	19
Min Ratio	1.0	1.0	1.1	5.2	3.8
Avg Ratio	1.5	1.5	1.2	15	8.6
Bit Avg Ratio	1.5	1.5	1.2	7.0	7.0
Overall Avg Ratio	3.8:1				
Overall Bit Ratio	2.8:1				

Table 7: SPOT-PAN Compression Results

	<u>B1</u>	<u>Lat</u>	<u>Lon</u>
Storage Bits	8	32	32
Actual Bits	8	25	26
Max Ratio	1.5	20	20
Min Ratio	1.3	17	6.5
Avg Ratio	1.4	18	9.7
Bit Avg Ratio	1.4	14	7.9
Overall Avg	6.7:1		
Overall Bit Avg	5.5:1		

Preliminary tests of the algorithm on level-1 ASAS and level-2 AVHRR-LAC image products, have resulted in average compression ratios of 2.5:1 and 2.2:1, respectively. The level-1 ASAS image product consists of seven spectral data files (one for each of seven look angles) containing 29 spectral bands and an ASCII header file. The spectral image bands are sequentially extracted from the look angle files and compressed. The level-2 AVHRR-LAC product contains an ASCII header file, two files containing cloud and urban/water feature masks, two files of reflectance values, one file of surface temperature values, one file of vegetation index values, and five files of original spectral data. Each of the files has been resampled to a spatial resolution of approximately 300 meters from the original 1 km resolution. The better compression of the ASAS data over that of NS001 is due to the increased similarity of pixels from its narrow spectral bands and the linear arrays of detectors that gather data. Both of these features would lend themselves to enhanced compression based on the steps in the algorithm. The level-2 AVHRR-LAC results are overall comparable to the compression results for the spectral bands shown previously in Table 4.

4.2 Algorithm Implementation

The ordered steps of the improved operational algorithm include : 1) normalization of the data matrix columns by subtracting the respective column minimum value from each element in the column, 2) normalization of the rows by subtracting the respective row minimum value from each value in the row, and 3) viewing the remaining values in each row as layers of bits that are either runlength encoded or packed back into bytes depending on the calculated average runlength of the line. A description of the improved compressed file format is provided in Appendix C.

The compression package, which consists of a base set of compression functions for handling 8, 16, and 32-bit values and processing control functions, is written in C and currently runs on a VAX computer in the Laboratory for Terrestrial Physics at NASA GSFC. Adherence to modularity of the sourcecode functions has resulted in the ability to implement compression of a new image-data product in a day or less. A complementary set of image-restoration software, also written in C, was developed, and can be run on several computing platforms including IBM compatible PCs, Macintosh (Plus, SE, and II), Sun and SPARC workstations, Personal Iris, and HP 9000 RISC workstations. This system independence was required to allow use of the data from the distributed CD-ROMs by as large a group of science users as possible. Although total system independence is not achieved, implementation of the restoration software on any of the mentioned machines simply requires editing a single line of the C source code file to set the type of system architecture (i.e., Intel or Motorola) being used before compiling and linking the package.

An average time of 12 minutes is currently required to compress a FIS level-1 AVHRR-LAC image product consisting of 13 files from 2.4 Mbytes to an average of 0.7 Mbytes on a VAX 11/780 system. A more modern system such as the VAX 6000 planned for use in the Boreal Ecosystem Atmosphere Study (BOREAS) (see [10]) could reduce this time to 1 minute or less. Restoration of the image product requires an average of 10 minutes on the loaded VAX 11/780 system; an average of 8 minutes on a 286 based PC; and 8 seconds on a Hewlett-Packard Series 9000 Model 720 RISC workstation.

5. Conclusions

The overall compression ratios realized by using the described algorithm on the FIFE image-data products are consistent with the rule-of-thumb limit of 2:1 for lossless image compression algorithms. The algorithm will definitely serve its purpose for the CD-ROM production effort in decreasing the amount of storage required for distributing the image data and has shed some

light on its potential use and enhancement for future efforts. For each image product, an overall ratio of 2:1 or greater was achieved from the compression, which allows production of only half as many CD-ROMs as would have otherwise been required. Although the original intent was to compress the spectral data, the best compression ratios were achieved in the georeferencing and view and solar-angle files. The compression of the latitude and longitude files within each of the AVHRR-LAC and NS001 products was quite similar, but there was a notable difference in the compression of the latitude and longitude files in the Landsat TM and two SPOT image products. In these three, the compression ratio of the latitude files were generally twice as large as the compression for the longitude files. Review of these files showed that the least-squares equations used in calculating the latitude coordinates were predominantly linear in nature while the equations derived to calculate the longitude coordinates were generally quadratic. For the one or two instances where the longitude coordinate equation was linear, the longitude coordinate file compressed comparably with the latitude coordinate file. This nonlinearity is the result of the small second-order effect caused by the convergence of the lines of longitude over the scenes, which is not present in the latitude values. This effect will be more important in northern latitudes, such as the proposed BOREAS site.

For current purposes, the algorithm is 'frozen' and operational in producing the FIFE CD-ROM image data sets, however, we are exploring options for improving on the current results. One possible improvement in the compression would be to map the remaining values in each line (after column and row minimum removal) into a Gray code (as discussed by Rabbani and Jones [9]) before proceeding with the runlength encoding. In theory, this should improve (increase) the length of bit runs for runlength encoding purposes. These improvements would benefit the upcoming BOREAS project, in which CD-ROM is being considered as a primary data-distribution mechanism for the satellite image data. It would also benefit distribution of the data using available communications networks.

The algorithm's performance is quite broad (as are most compression algorithms) depending on the data layer being processed, but its generality has been a significant point in handling the FIFE data sets. Based on our experience with the algorithm, we envision this 'intuitive' compression algorithm to be useful on a broad range of spatial data sets including gridded layered modeling data sets which could contain terrain, spectral, and meteorological variables. Such data sets will be required for coordinated earth systems field experiments executed during the next decade.

Acknowledgements

The work reported here was supported by the FIFE project through money supplied by the Land Processes Branch, NASA Headquarters, under contract to Hughes STX Corporation and VERSAR Incorporated. The authors thank Andy Newton (University College London, England) for his feedback on the initial algorithm implementation which made us further examine portability of the image restoration software and resulted in a much more useable package.

References

- [1] Walker, R.J., "Scientific Requirements for Space Science Data Systems", *Proceedings of the Space and Earth Science Data Compression Workshop*, NASA Conference Publication 3130, pp. 37 - 50, April 11, 1991.
- [2] Sellers, P.J., F.G. Hall, G. Asrar, D.E. Strebel, and R.E. Murphy, "The First ISLSCP Field Experiment (FIFE)", *Bulletin of the American Meteorological Society*, vol. 69, no. 1, pp.22-27, 1988.
- [3] Strebel, D.E., D. Landis, J.A. Newcomer, J. DeCampo, and B.W. Meeson, "Collected Data of the First ISLSCP Field Experiment, Prototype Volume", Published on CD-ROM by NASA, 1991.
- [4] Strebel, D.E., J.A. Newcomer, J.P. Ormsby, F.G. Hall, and P.J. Sellers, "The FIFE Information System", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 703-710, July 1990.
- [5] Dozier, J. and J.C. Tilton, "Data Compression for Data Archival, Browse, or Quick-Look", *Proceedings of the Space and Earth Science Data Compression Workshop*, NASA Conference Publication 3130, pp. 61 - 65, April 11, 1991.
- [6] Newcomer, J.A., S.J. Goetz, D.E. Strebel, and F.G. Hall, "Image Processing Software for Providing Radiometric Inputs to Land Surface Climatology Models", *Proceedings of IGARSS'89*, vol. 3, pp. 1779-1782, 1989.
- [7] NASA, "Report of the Eos Data Panel", Volume Ila, NASA Technical Memorandum 87777, 1986.
- [8] Salomonson, V.V., "The Moderate Resolution Imaging Spectrometer: An Eos Facility Instrument Candidate for Applications of Data Compression Methods", *Proceedings of the Space and Earth Science Data Compression Workshop*, NASA Conference Publication 3130, pp. 13 - 23, April 11, 1991.
- [9] Rabbani, M. and P. Jones, *Digital Image Compression Techniques*, SPIE Optical Engineering Press, Volume TT7, 221 pages, 1991
- [10] Sellers, P., J. Cihlar, M. Apps, B. Goodison, F. Hall, B. Harriss, D. Leckie, E. Ledrew, P. Matson, and S. Running, *Charting the Boreal Forest's Role in Global Change*, Eos, Vol. 72, No. 4, pp. 33, 35, and 40, January 22, 1991.

Appendix A FIFE Image Product Acronyms

ASAS	-	Advanced Solid-State Array Spectroradiometer
AVHRR-LAC	-	Advanced Very High Resolution Radiometer - Local Area Coverage
AVHRR-GAC	-	AVHRR - Global Area Coverage
Landsat TM	-	Landsat Thematic Mapper
NS001 (TMS)	-	NS001 Thematic Mapper Simulator
PBMR	-	Push-Broom Microwave Radiometer
SPOT-XS	-	Système Probatoire d'Observation de la Terre (Multispectral mode)
SPOT-PAN	-	SPOT (Panchromatic mode)
TIMS	-	Thermal Imaging Multispectral Scanner

Appendix B Original Compressed File Format

[Note that two-byte fields are in VAX byte order (i.e., low-order byte first)]

Record 1 (Header record)

Bytes 1 - 4	:	Overall image minimum value
Byte 5	:	Number of bits being encoded for each image line (NBits)
Byte 6	:	Total number of bits in the file values (8, 16, or 32)
Byte 7	:	Number of 256 element segments across a line of the image (NSeg)
Bytes 8 - 9	:	Number of records (lines) in the original file (NLines)

Records 2 - n (Encoded data records) (where $n = NBits * NSeg * NLines + 1$)

Runlength Encoded Record

Byte 1	:	Record type (0 = runlength encoded)
Byte 2	:	First runlength bit value (equal 0 or 1)
Byte 3	:	Number of runlength counts that follow (m)
Byte 4	:	Runlength count - 1 for first bit value (record byte 2)
Byte 5	:	Runlength count - 1 for opposite bit value

.

.

Byte m+3 : Runlength count - 1 for last bit series

Bit Encoded Records

Byte 1	:	Record type (1 = bit encoded)
Byte 2	:	Contains packed bits for segment values 1 - 8
Byte 3	:	Contains packed bits for segment values 9 - 16

.

.

Byte 32 : Contains packed bits for segment values 249 - 256

Appendix C Improved Compressed File Format

[Note that two and four-byte fields are in VAX byte order (i.e., low-order byte first)]

Record 1 (Header record)

Byte 1 : Total number of bits in the file values (8, 16, or 32) (Total_Bits)
 Bytes 2 - 3 : Number of lines/records in the original file (NLines)
 Bytes 4 - 5 : Number of values per line (NVals)

Record 2 (Column minimum record)

Bytes 1 - n : Column minimum values (1, 2, or 4 bytes each depending on Total_Bits)
 (NVals in number)

Records 3 - z (Encoded sets of compressed data information)

Each set of information used to reconstruct an output data file record contains a Row_Minimum value, an NBits value, and a set of encoded records in the following manner:

Row_Minimum : Stored as an 8, 16, or 32 bit value depending on the respective value of Total_Bits in the header record.
 NBits : Stored as an 8 bit value. Indicates the number of bits that were left for encoding in this line after column and row minimum subtractions.

Series of NBits encoded data records as Runlength encoded or Bit encoded records

Runlength Encoded Record

Byte 1 : Record type (0 = runlength encoded)
 Byte 2 : First runlength bit value (equal 0 or 1)
 Byte 3 - 4 : Number of runlength counts that follow (Nr)
 Byte 5 - 6 : Runlength count - 1 for first bit value (record byte 2)
 Byte 5 : Runlength count - 1 for opposite bit value

Byte m - m+1 : Runlength count for last bit series, where $m = 2 * Nr + 3$

Bit Encoded Records

Byte 1 : Record type (1 = bit encoded)
 Byte 2 : Contains packed bits for segment values 1 - 8
 Byte 3 : Contains packed bits for segment values 9 - 16

Byte J+1 : Contains packed bits for last values in the record where
 $J = NVals/8$ if NVals is a multiple of 8 or
 $J = NVals/8 + 1$ if NVals is not a multiple of 8