Lossless Compression of 3D Hyperspectral Sounding Data Via Statistical Image Characteristics

OLGA KUBASOVA, PEKKA TOIVANEN, JARNO MIELIKAINEN
Department of Information Technology
Lappeenranta University of Technology
P.O. Box 20, 53850 Lappeenranta,
FINLAND

Abstract: - An efficient lossless compression algorithm for 3D sounding data is presented. Major phases of the algorithm are: image band reordering, prediction and coding. Proposed prediction technique allows exploiting the data redundancy in both spectral and spatial dimensions. In order to increase the efficiency of the prediction technique, image bands should be rearranged in such a way, that the bands with higher correlation are allocated together. We examined different band orderings and their influence on the final compression ratio for a given set of images. This paper emphasizes that the compression algorithm yields different compression performance (space and time ratios) depending on how image bands have been rearranged or, in other words, which metric has been applied to examine interband similarities. We experimented with a wide variety of metrics, employing each metric in the algorithm at a time. 10 real 3D sounders have been used for the experiments. The results obtained show significant impact of newly proposed metrics in lossless compression of hyperspectral data. The novel algorithm has improved JPEG-LS performance on about 10 -12% in terms of compression ratios. Computational speed of the algorithm is also very high.

Key-Words: - Sounding Hyperspectral Images, Lossless Image Compression, Prediction, Reordering

1 Introduction

In recent years considerable research has been carried out into the field of remote sensing, aiming to develop and investigate hyperspectral imaging and sounding data. Hyperspectral instruments such as Atmospheric Infrared Sounder (AIRS) [1], Cross-track Infrared Sounder (CrIS) [2], Interferometer Atmospheric Sounding Instrument (IASI) [3], Geosynchronous Imaging Fourier Transform Spectrometer (GIFTS) [4], and Hyperspectral Environmental Suite (HES) [5], and Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [6] daily generate large volumes of three-dimensional hyperspectral data. The instruments are designed to meet the Operational Weather Forecasting requirements of NOAA [7] and the research needs of the NASA [8].

Main applications of remote sensing relate to the global water and energy cycles, determination of the distribution and variations of water vapor, the climate and weather connection, improvements in the weather prediction, and trace gases. Hyperspectral sounding data provide much more precise information about atmospheric temperature, moisture, clouds, aerosols and surface properties than hyperspectral imaging data, but storage and transition of these images demand much more resources. Thus compression algorithms must be developed.

Data loss of the hyperspectral imaging data, e.g. AVIRIS can be acceptable by the human visual system. Whereas the hyperspectral sounding data requires a much higher accuracy for the useful retrieval of geophysical parameters without significant degradation, the allowable reconstructed error is visually imperceptible by the human eye. Therefore, lossless or near lossless compression algorithms for 3D sounding data are warranted. Lossless compression algorithms provide intact information about the original image, although their benefits in compression ratio and speed are not as high as they would be with lossy compression algorithms.

Previously applied lossless predictor-based compression algorithms such as JPEG-LS [9] and

CALIC [10] are not efficient enough for the sounding data compression, since these algorithms have been built to work on 2D data and cannot exploit redundancy in spectral dimension between disjoint bands. Even when their performance has been improved by involving reordering techniques, such as [11] or [12], the results are still not very high.

This paper presents an efficient lossless compression algorithm, which has produced excellent results being tested on 10 hyperspectral sounding images. Next section introduces the algorithm. Section 3 provides overview of the hyperspectral sounding data used in this study. Section 4 details obtained compression results and summarizes the paper.

2 Compression Algorithm

3D image is a specific type of imaging data with one spectral and two spatial dimensions. The 3D image *Im [M, N, S]*, where *M* is number of rows, *N* number of columns and *S* is a number of bands can be represented as a set of 2D arrays. A single entry *B [M, N, S_i]* belonging to this set represents one image band. Then two spatial dimensions of each image band can be converted into a 1D array. Hence 3D image hypercube structure has changed to a set of 1D vectors, where each vector has *Q=M* x *N* pixels and number of the vectors is *S*.

Efficient 3D compression algorithms exploit interpixel redundancy in every image dimension. The proposed algorithm extracts spectral redundancy by coding each band of the image by making use of another *prediction band*. Previously coded pixels, known as a causal set, are involved to estimate the *current pixel* value. The least squares (LS) estimation approach, which predicts the current pixel using a linear combination of its causal neighborhood, has been adapted for multispectral lossless image compression.

In order to improve the effectiveness of the prediction phase, we propose to apply optimal band reordering technique before actual prediction. Originally image bands are presented in order given by the spectral wavelengths, which is referred to the *natural band ordering*. Some image bands, located in neighboring wavelengths, might have weak correlation. Thus it is reasonable to reorder image bands in such a way, that adjacent bands have the highest correlation. Moreover, it may happen that one band has good correlation with several other image bands, or, in other words, this band could be a good predictor for those bands.

Image interband relationships should be known to perform efficient reordering. Interband similarities can be very accurate measured by one of the proposed metrics (1) – (4). These metrics take two vectors as an input and compute the real number in the range [0.0; 1.0], where 1.0 indicates that input vectors are identical, 0.0 shows that they are totally different. Each metric is presented as function $M_k(B_i, B_j)$, where B_i, B_j are the hyperspectral image bands, in a form of 1D vectors.

Measure 1

$$M_1 = \frac{B_i B_j^t}{|B_i||B_j|} = Cos(\theta) \qquad (1)$$

Measure 2

$$M_2 = \frac{|B_i| Cos(\theta) + |B_j| Cos(\theta)}{\left(|B_i|^2 + |B_j|^2 + 2|B_i||B_j| Cos(\theta)\right)^{1/2}} \qquad (2)$$

Measure 3

$$M_3 = \frac{Cos(\theta)\left(|B_i|^2 + |B_j|^2 + 2|B_i||B_j| Cos(\theta)\right)^{1/2}}{|B_i| + |B_j|} \qquad (3)$$

Measure 4

$$M_5 = \frac{\sum_{k=1}^{Q} |B_{ik} - \bar{B}_i||B_{jk} - \bar{B}_j|}{\left(\sum_{k=1}^{Q}(B_{ik} - \bar{B}_i)^2\right)^{1/2}\left(\sum_{k=1}^{Q}(B_{jk} - \bar{B}_j)^2\right)^{1/2}} \qquad (4)$$

where $\bar{B}_i = \frac{1}{N}\sum_{k=1}^{Q} B_{ik}$.

Each of these metrics can be used to estimate similarity between every two image bands. All coefficients are collected in one matrix, which is used for further reordering and prediction. The procedure of estimation interband relationships is the first step of the proposed lossless compression algorithm.

Next step of the algorithm is to reorder image bands in such a way that higher correlated bands, are allocated together. The problem of optimal image band reordering is equivalent to the problem of finding Minimum Spanning Tree (MST) in weighted directed graphs [13]. The computed matrix of interband similarity coefficients can be considered as a matrix of weight coefficients for the graph, where the graph vertices denote the image bands' numbers and the edges assign if there is a direct connection between two vertices. Prim's algorithm [14] is known to be good for constructing MST in weighted directed graphs.

Actually, we are searching for the maximum weighted tree and desirable output is a set of bands' numbers, corresponding to the path how image bands should be allocated in oreder to achieve the best compression results.

Each band must be predicted and no one twice. Although the same band might be used as a predictor for two or even more bands. No cycles are allowed in the MST, therefore decompretion is easily achivable by reversing the procedure of MST construction. Due to prohibiton to have any cycles, sometimes the predictor for the current band is not the best possible. Let us denote the band, which is used for prediction as the *predictor,* and the band, which is predicting as the *current band.*

When the optimal band ordering is achieved, and for any image band the predictor is determined, a linear prediction technique [15] is applied to exploit interpixel redundancy in the special and spectral dimensions.

The technique is based on the idea that the current pixel value can be computed using pixels form its causal neighborhood. The predicted pixel value $p'_{r,c,b}$ can be found using the following equation:

$$p'_{r,c,b} = \sum_{k=0}^{S} \sum_{j=0}^{T} \sum_{i=L_{j,k}}^{R_{j,k}} a_{r-i,c-j,b-k}\, p_{r-i,c-j,b-k} \qquad (6)$$

In (6) $a_{i,j,k}$ is a prediction coefficient of the pixel at location *(i, j, k)*, *S* is the number of bands; *T* (top) is the number of rows that take part in the causal set. $L_{j,k}$ (left) and $R_{j,k}$ (right) are the left and right delimiter as functions of rows and bands that finally specify the causal set. The causal set in case when two bands are involved is shown in Fig.1.

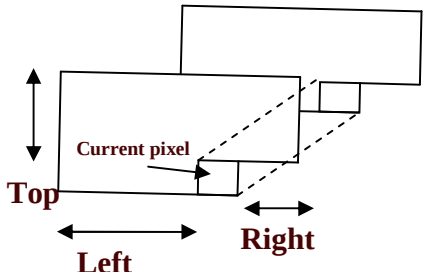

Fig 1. An example of causal set when two image bands are involved.

Examples of possible neighbourhoods are shown in Fig.2.

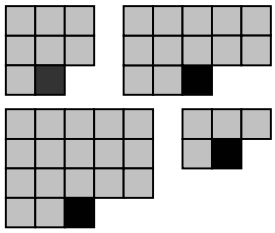

Fig 2. Examples of the sample sets used in the algorithm for prediction of the current pixel. The current pixel is marked by the black square**.**

Finally error images are computed as a difference between the original and predicted images and entropy coded.

2.1 Optimal Compression Algorithm

In order to obtain maximal possible compression ratios for a given image, regardless of the longer computational time, we have improved the prediction phase of the proposed algorithm in the following way.

We still use the linear prediction technique, but at this time, a sample set size is not fixed. Thus for each pixel prediction the best sample set has been chosen at a time.

The whole procedure is a two-pass technique. At first, each pixel is predicted using values of the pixels from its causal neighborhood, known as a sample set. In the algorithm different sample sets, shown in Fig. 2 are involved. Then residuals between the original and

predicted values are calculated and coded. Then for each special position, a sum of the absolute values of the residuals over the bands is calculated and the sample set, which gives a smaller sum, is used for the spatial position during the second pass.

Finally error images are computed as a difference between the original and predicted images. These images are coded using a range coder [16]. For the pixels that do not have a large enough causal neighborhood of the current pixel, and therefore cannot be predicted, we calculated the difference between consecutive bands, which along with other compression parameters, are compressed using general-purpose text compression program, PPMD [17].

The optimal algorithm results in very high compression ratios, but slowness of the algorithm does not make it efficient in practice. The obtained compression ratios are the maximal possible ones for a given prediction scheme.

3 Hyperspectral Sounding Data

In our experiments hyperspectral sounding images have been used. Each image has 2108 spectral bands and it could be generated from either an interferometer or a grating sounder with extremely high temporal and special resolutions (over one thousand infrared channels and with spectral widths on the order of 0.5 wave number).

Global coverage by the instruments is obtained twice daily (day and night) and the data along the orbit is processed into 6-minute granules. Each granule consists of 135 scan lines containing 90 cross-track footprints per san line; thus there are a total of 135 x 90 = 12,150 footprints per granule.

The data is available at the Distributed Active Archive Center (DAAC) located at the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) in Greenbelt, Maryland, USA. The data, used for this compression study, is available via anonymous ftp (ftp://ftp.ssec.wisc.edu /pub/bormin/HES) [18].

The algorithm has been tested on ten granules, five daytime (DT) and five nighttime (NT), which have been chosen from different geographical regions of the Earth. Their locations, UTC times, and local time adjustments are listed in Table 1.

Table 1. Ten selected AIRS granules for hyperspectral sounding data compression studies. [18]

Granule №	Time	Location
9	00:53:31 UTC -12H	Pacific Ocean, DT
16	01:35:31 UTC +2 H	Europe, NT
60	05:59:31 UTC +7 H	Asia, DT
82	08:11:31 UTC -5 H	North America, NT
120	11:59:31 UTC -10H	Antarctica, NT
126	12:35:31 UTC -0 H	Africa, DT
129	12:53:31 UTC -2 H	Arctic, DT
151	15:05:31 UTC +11 H	Australia, NT
182	18:11:31 UTC +8 H	Asia, NT
193	19:17:31 UTC -7 H	North America, DT

4 Conclusion

The majority of lossless compression algorithms is made up by efficient prediction phase, when pixel values are predicted, and error images are coded. This paper emphasizes the importance of the reordering phase, which is applied before actual prediction in order to improve its performance. The proposed lossless compression algorithm results are highly dependent on the 3D image reordering. The band ordering is determined by the statistic, which measures the interband similarities. Several metrics have been tested and analyzed in this study. These are very popular for color difference measurement in multispectral data, but most of them have never been used as measures of interband similarity.

The following table reflects compression ratio of the images, which have been compressed by the proposed algorithm, when each of the statistic has been employed at a time. The compression ratio is measured as the size of the original file divided by the size of the compressed file.

Table 2. Bit rates for the 10 tested granules compressed by one of the proposed techniques.

Granule №	*Natural band ordering*	*The alg. with M1*	*The alg. with M2*	*The alg. with M3*
9	2.080	2.1800	2.180	2.180
16	2.000	2.1110	2.111	2.111
60	1.940	2.0670	2.066	2.067
82	2.008	2.1030	2.103	2.104
120	1.937	2.0280	2.028	2.028
126	2.035	2.1530	2.153	2.153
129	1.986	2.0680	2.068	2.068
151	2.081	2.1710	2.171	2.171
182	1.906	2.0170	2.018	2.018
193	2.035	2.1460	2.146	2.146

Granule №	*The alg. with M4*	*Opt. Comp.*	*JPEG-LS*
9	2.1790	2.205	1.990
16	2.1110	2.128	1.942
60	2.0640	2.072	1.895
82	2.1040	2.123	1.950
120	2.0280	2.039	1.900
126	2.1570	2.180	1.939
129	2.0690	2.082	1.934
151	2.1700	2.198	2.015
182	2.0150	2.031	1.887
193	2.1470	2.167	1.944

The first column in Table 2 reflects the image name. The rest of the Table shows compression ratios of the images, compressed by the proposed algorithm, when each of the metrics has been employed at a time. Only the 2nd column reflects compression ratios of the images, compressed by the same algorithm, but when no reordering technique has been applied. In other words, images have had natural band ordering. The 3rd -6th columns reflect compression ratios of the images produced by the algorithm when each of the proposed metrics M1,…, M4 has been involved at a time. Each metric implies different band reordering, thus compression results are highly depend on the metric being used. In the next column compression ratio of the images compressed by the optimal algorithm are shown. The results are high, but the algorithm is extremely time-consuming. The last column shows results produced by JPEG-LS algorithm for the same images.

The time required for compression of the images with the natural band ordering, the optimal band ordering and the orderings implied by M1 and M4 metrics are reported in Table 3. Compression computational time is much higher than the decompression time, since enormous amount of time is spent on achieving the optimal band reordering. The more complicated is the way of the image band reordering, the longer is the computational time. Thus, the fastest way of compression is to ignore reordering phase, or to compress an image with the natural band reordering. But since the purpose of compression in achieving a small file size as well as the fast computational speed, omitting the reordering phase could cause poor results in terms of compression ratios. The smallest file size has been achieved with the optimal band reordering, although the algorithm with this technique requires the longest compression time. Thus aiming to achieve a compromise between a fast computational speed and a high compression results, we have proposed in this paper a set of metrics, which can be used on the reordering phase. The most efficient of them are M1 and M4. Analyzing compression results obtained with these metrics and shown in Table 2, we can conclude that they are very close to the results, obtained by the algorithm, when the optimal band reordering has been involved. Nevertheless computational speed is much higher than the speed allowable with the optimal band reordering.

Table 3. Compression and decompression time in seconds for the 10 tested granules compressed by one of the proposed techniques.

Gr. №	*Natural band ordering*	*The alg. with M1*	*The alg. with M4*	*Opt. Comp. algorithm*
	Comp/ Decomp.	*Comp/ Decomp.*	*Comp/ Decomp.*	*Comp/ Decomp.*
9	25.5/8.1	107.4/7.7	111.2/7.8	20082.06/8.1
16	26.5/9.6	107.8/8.1	111.3/8.1	19600.67/8.5
60	26.6/9.8	108.0/8/2	111.6/8.2	26329.65/9.1
82	26.8/9.9	108.0/8.2	111.6/8.1	18375.97/8.8
120	26.6/9.8	108.1/8.6	111.7/8.5	18939.36/8.9
126	26.2/9.2	107.7/7.8	111.2/7.8	25424.86/8.1
129	25.8/9.3	107.8/8.2	111.3/8.2	17729.78/8.5
151	25.3/8.8	107.4/7.9	111.1/7.8	17161.61/8.1
182	28.6/13.5	108.7/8.7	112.3/8.9	23565.38/10
193	26.1/9.3	107.7/7.9	111.2/7.9	20587.77/8.3

One of the proposed reordering techniques can be chosen to accompany the compression algorithm, based on the application requirements. Concluding the discussion above, it should be mentioned that the fastest compression speed is achievable with the algorithm with natural band ordering and the best compression savings are provided by the optimal algorithm. The best results in terms of both time and compression ratios have been obtained, when M1 and M4 statistics have been involved in the reordering phase of the algorithm.

Moreover the comparison of the results obtained with the proposed algorithm and JPEG-LS has been made. Since JPEG-LS has been developed to compress 2D images, it is not sufficient in working on 3D data. The results obtained clearly show the advantages of the novel technique.

References:

[1] Aumann H.H. and Strow L., AIRS, The First Hyperspectral Infrared Sounder for Operational Weather Forecasting, *in Proc. of IEEE Aerospace Conference*, 2001, pp. 1683-1692.

[2] Bloom H.J., The Cross-track Infrared Sounder (CrIS): a Sensor For Operational Meteorological Remote Sensing, *in Proc. of the International Geosciences and Remote Sensing Symposium*, 2001, pp. 1341-1343.

[3] Phulpin T., Cayla F., Chalon G., Diebel D., and Schlüssel D., IASI Onboard Metop, *in Proc. of the 12th International TOVS Study Conference*, 2002.

[4] Smith W.L., Harrison F.W., Hinton D.E., Revercomb H.E., Bingham G.E., Petersen R., and Dodge J.C., GIFTS - The Precursor Geostationary Satellite Component of the Future Earth Observing System, *in Proc. of the International Geoscience and Remote Sensing Symposium*, 2002, pp. 357-361.

[5] Huang B., Huang H.L., Chen H., Ahuja A., Baggett K., Smith T.J., Heymann R.W., Data Compression Studies for NOAA Hyperspectral Environmental Suite (HES) using 3D Integer Wavelet Transforms with 3D Set Partitioning in Hierarchical Trees, *in Proc. of the SPIE International Symposium on Remote Sensing*, 2003.

[6] Jet Population Laboratory, Available: http://aviris.jpl.nasa.gov/ .

[7] National Ocean and Atmospheric Administration, Available: http://www.noaa.gov/.

[8] National Aeronautics and Space Administration (NASA), Available: http://www-airs.jpl.nasa.gov .

[9] Weinberger M., Seroussi G, Sapiro G., The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS, *Hewlett-Packard Laboratories Technical Report No. HPL-98-193R1*, 1998-1999.

[10] Wu X. and Memon N., Context-based, adaptive, lossless image coding, in *IEEE Trans. Commun.*, vol. 45, pp. 437–444, Apr. 1997.

[11] Huang B., Alok A, and Hung-Lung H., Improvements to Predictor-based Methods in Lossless Compression of 3D Hyperspectral Sounding Data via Higher Moment Statistics, *WSEAS Transactions on Electronics*, Issue 2, Vol 1, 2004

[12] Kubasova O., Toivanen P., Lossless Compression Methods for Hyperspectral Images, *in Proc. of 17th International Conference on Pattern Recognition*, 2004 (to be published)

[13] Purdom W., Brown J., *The analysis of algorithms*, CBS College publishing, USA 1985.

[14] Baase S., *Computer Algorithms: Introduction to Design and Analysis* , Addison-Wesley, 1988.

[15] Mielikainen J., Toivanen P., and Kaarna A., Linear Prediction in Lossless Compression of Hyperspectral Images, *Optical Engineering Journal*, Vol. 4 No 4, 2003, pp. 1013-1017.

[16] G. N. Martin, Range encoding: an algorithm for removing redundancy from a digitalized image, *in Proc. of Data Compression Conference*, 1979.

[17] D. Shkarin, PPM: One step to practicality, *in Proc. of Video and Data Compression Conference, Snowbird, Utah*, pp. 202-211, 2002.

[18] Huang B., Hung-Lung H., Alok A., and Hao C., Lossless Data Compression for Infrared Hyperspectral Sounders – An Overview, *in Proc. of 20th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, 2004.