CGMS

CGMS-35, NOAA-WP-11
Prepared by R. Heymann
Agenda Item: G.2

Data Compression

NOAA-WP-11 discussed the compression of satellite sensor earth science data as a technology subject of importance to CGMS members. NOAA views compression of satellite sensor Earth science data as an important technology to enhance data distribution with the expected great increase in data rates in the future from new sensors' faster scanning, finer spectral, and finer spatial resolutions. Examples of such sensors would be multi-spectral imagers and FTS/Michelson Sounders.

NOAA is not the only government agency undertaking research on compression of satellite Earth science data. CNES of France and the Canadian Space Agency (CSA) are government bodies outside of the U.S. also undertaking research in this area.

NOAA has been undertaking research on compression of satellite sensor Earth science data for nearly five (5) years. Their research has been directed to "Lossless" compression. Our research shows that the data structure of satellite Earth science sensor data is unique in terms of patterns, structure, and features. Further, NOAA's research shows that the amount of "Lossless" compression achievable for Earth science sensor data, varies with sensor type; with data characteristics significantly varying among sensor types. We also find that contemporary widely used compression mathematics such as that used in the JPEG series and CCSDS compression standards perform poorly on satellite sensor Earth science data for "Lossless" compression. The common math of these standards are transforms such as Discrete Cosine Transforms (DCT) and wavelets as a data pre-processor followed by an entropy coder (e.g. Arithmetic, Huffman, Rice).The NOAA research attempts to achieve highest levels of "Lossless" compression" with satellite earth science data, Novel new approaches and mathematics are required are being developed in the NOAA research.

New classes of compression algorithms, such as those coming out of NOAA, are performing by almost a factor of two to one greater then the ability of conventional compression algorithms when applied to this problem. By this, NOAA mean where conventional algorithms might achieve a 1.5 to 1 or 2 to 1 "Lossless" reduction in data volume when applied to satellite Earth science data; new classes of algorithms, such as those emerging from NOAA research, are yielding reductions of 3 to1 for multi-spectral Imager data, near 4 to1 for dispersive hyper-spectral sounder data such as the AIRS instrument, and with preliminary results reductions of 4 to1 and 5 to 1 for FTS/Michelson interferometer data.

Data Compression

1. INTRODUCTION

NOAA sees Earth science satellite sensor data compression, through its reduction in data levels, being able to serve a number of important purposes for the Coordination on Geostationary Meteorological Satellites (CGMS) member nations. Data compression can do the following:

- Reduce costs of data distribution, be it by satellite or terrestrial land lines.
- In cases of limited RF spectrum availability, for meteorology and Earth exploration satellite data transmission use, enable a greater level of Earth science data be distributed to world nations.
- Reduce Earth science data archive costs.

Distributing maximum levels of satellite Earth science data and doing so at low cost to world nations is consistent with many policy goals. Such a data distribution is consistent with the data distribution goals of the World Meteorological Organization (WMO); and Agreements reached at the Earth Observation Summit (EOS) and by EOS signatory nations to establish the Global Earth Observation System of Systems, known as GEOSS.

2. Compression of Satellite Earth Science Sensor Data

NOAA believes its compression research program on satellite Earth science data reduction has achieved important results. It has shown compression of satellite Earth science sensor data to be a unique problem, defined by the uniqueness of the data structure itself. It has shown that new compression algorithms, introducing new applications of mathematics to this problem, can yield significantly greater data reductions then possible from adopting conventional popular standardized compression algorithms (e.g. JPEG, JPEG 2000, GZIP, EZW).

Data compression is a mathematical way to represent digital data with as few bits as possible. Compression algorithms thus reduce the volume of the data; with the data then being restored either fully or partially before use depending upon the algorithm design (Lossless, near Lossless, Lossy compression approaches). The restoration process is known as decompression. Compressed data must be decompressed before it can be used. A “Lossless” compression by accepted definition means the data can be restored exactly to its original state, without a single bit change. A “Lossy” compression means on decompression the data is restored to a usable state, but there is a loss of original data. An advantage of “Lossy” compression over “Lossless” is it is able to achieve much higher data reduction, but at a sacrifice of data quality and accuracy on decompression. With satellite Earth science data many scientists and users need data without losses or errors, and as such “Lossless” compression is needed.

Compression of digital data is not a new concept. It is widely used in many elements of today’s digital era and in a way has been brought about by today’s digital era of Personal computers, internet, and global communications. It is widely used in the internet through the widely known JPEG and its successor JPEG 2000 compression standards developed for compressing two-dimensional picture imagery for internet and PC use. The JPEG compression standard utilizes the Discrete Cosine Transform (DCT) mathematics to carry out compression. Its successor standard JPEG 2000 relies on wavelet mathematics for data preprocessing followed by what is known as an entropy coding step. Compression of data volumes has become widely used in the medical field to reduce the burden of storing digital imagery records as well as readily enabling digital imagery to be transmitted for viewing, diagnosis, and consultations over long distances. Compression algorithms for reducing medical digital data volumes have traditionally relied on the mathematics of vector quantization.

NOAA satellite Earth science sensor data compression research to date has been directed at two types of widely used visible and IR sensors. The first type is the satellite multi-spectral imager such as Japan's MTSAT imager, NASA MODIS imager, and current and future GOES imagers (the present 5 channel and future 16 channel), and METEOSAT SEVIRI imager. The second type is the IR hyper-spectral sounder, being of two classes. The first class of sounder is a grating spectrometer such as the 2378 channel NASA EOS Aqua AIRS sounder. The second sounder class is the Fourier Transform Spectrometer (FTS)/Michelson Interferometer such as CNES and Alcatel developed "Infrared Atmospheric Sounding Interferometer" (IASI) sounder on the EUMETSAT METOP satellite.

To date all of our NOAA research has centered on what is categorized as "Lossless" data compression. By definition, compression is "Lossless", if on reconstruction or decompression, the original data has been exactly restored. There are no bit errors, at all. NOAA may eventually in its research explore a "near Lossless" compression, depending upon funding availability. With "near Lossless" compression some of the original data is lost on decompression. In some cases, this might be judged by the science community acceptable depending upon the impact.

We wish to highlight some key findings to the CGMS member nations coming out of our NOAA research of "Lossless" compression of satellite Earth science sensor data. We do this with the bullet list below as well as 5 abstracts from the 2006 "SPIE" meeting held in San Diego, CA from reflecting presentations and journal publications of NOAA sponsored compression research:

1. We find "Lossless" compression of Earth science satellite sensor data to be a unique problem due its being 3 dimensional data and unique nature. To achieve maximum possible data volume reductions through "Lossless" compression, new unique data compression algorithms are found required for each sensor type (Multi-spectral imager, Hyper-spectral sounder (FTS, and dispersive). For example, we find one distinct "Lossless" compression algorithm is required for FTS/Michelson sounder and another separate for the dispersive sounder.
2. Contemporary "Lossless" compression approaches embodied in well know compression standards and popular non-standardized "Lossless" compression algorithms developed principally for 2 dimensional data in web use perform poorly on satellite sensor Earth science data. By this we mean their "Lossless" compression ratios on compression are near 2 to 1. This would apply to

JPEG and JPEG 2000 compression standards, and current and past Consulting Committee for Satellite Data Standards (CCSDS) for compression. And, such widely used math developed for 2-dimensional compression such as: wavelets for preprocessing followed by entropy coding, Discrete Cosine Transforms (DCT), 3 dimensional "Embedded Zero Tree wavelet (EZW), and 3 dimensional Set Partitioning in Hierarchical Trees (SPIHT), for example.

3. Compressing Earth science sensor data for transmission by satellite requires that additional mathematics be introduced to provide robustness against the relative high electronic noise environment of wireless communications and that for satellites this be synergistic with the "Forward Error Corrections" (FEC) coding typically used in a satellite communication system. Such FEC coding could be Turbo Product code or Low Density Parity Check (LDPC). Failing to do so, could lead to a partial or complete loss data, or striping on data decompression at the receive site. We find for example the JPEG 2000 compression standard to be susceptible to data stream errors in satellite use; a vulnerability that can lead to loss of some or all data.
4. "Lossless" compression results to date expressed in terms of compression ratio (CR) from NOAA research on satellite Earth science sensor data, with novel new algorithms developed around sensor data characteristics are:

    a. New working Lossless compression algorithms (two have been developed by two separate research teams) for **AIRS** grating/dispersive spectrometer **sounder** sensor yielded CRs of 3.7 and 4.3. At these CRs neither algorithm had error protection embedded within needed for satellite use.
        i. With robust error protection embedded the CR of 3.7 is reduced to 3.1 to 1.
    b. The following table identifies the various results obtained with contemporary compression algorithms applied to AIRS data in the needed 3 dimensional manner

Contemporary compression Algorithm	CR "Lossless"
JPEG 2000 standard applied 3 dimensionally	**2.37 to 1**
JPEG LS	**2.46 to 1**
CCSDS with Rice compression	**1.97 to 1**
GZIP	**1.89 to 1**
SPIHT	**1.97 to 1**
EZW	**1.95 to 1**

These above tabularized compression research results (CRs) are published in a journal paper by the SPIE society, titled "Wavelet Lossless Compression of Ultraspectral Sounder Data". The authors are: Joan Serra-Sagristà and Fernando García-Vílchez, Department of Information and Communications Engineering ETSE, Universitat Autònoma Barcelona, SPAIN
Julià Minguillón and David Megías, Estudis d'Informàtica i Multimedia, Universitat Oberta de Catalunya, SPAIN
Bormin Huang and Alok Ahuja, CIMSS/SSEC, University of Wisconsin, Madison,USA

The paper is available from both SPIE and Amazon within a published book of edited papers or singularly as a paper from SPIE: "Satellite Data Compression, Communications And Archiving" (Proceedings of SPIE, August 2005 annual meeting, San Diego , CA), by Bormin Huang (Editor), Roger W. Heymann (Editor), Charles C. Wang (Editor

c. A new NOAA working Lossless compression algorithm for FTS/Michelson interferometer sounder sensor yields a CR of ~5.0 to 1, for the aircraft flown NOAA NPOESS NASTI sounder. A revised algorithm for a 2nd FTS sounder, has a preliminary CR of 4 to 1.
d. Multispectral Imager compression:
    i. Meteosat-8 SEVERI using 11 or its 12 channels – NOAA research program estimate of the maximum and possible achievable "Lossless" compression of the SEVIRI imager with data from 11 or its 12 channels is CR 3.1 to1. The 11 SEVIRI channels were selected to reflect the next generation GOES-R series ABI imager.
e. NOAA obtained a "Lossless" CR of 2.9 to 1 directly from a newly developed "Lossless" compression algorithm tailored to the Japan Meteorological Agency (JMA) MTSAT Imager and its HiRID stream of data at 660Kbps

5. Undocumented but reported Lossless compression of NASA "USES" compression algorithm based on the "RICE" entropy coder applied to the future GOES-R series ABI data is a CR 1.7 to 1.

6. Below are 5 abstracts to 5 journal papers published by the "SPIE" society for its San Diego, CA meeting of August 13-17, 2006

Below 5 selected abstracts associated with 5 published papers identifying 2006 results of NOAA's funded satellite Earth science sensor data compression research are provided. The results of the NOAA funded research reflected in the abstracts were presented at the 2006 August 13-17, 2006 SPIE annual meeting, in its OEI106 conference. An edited book of all papers at the OEI106 conference, including the journal for the 5 below abstracts can be obtained through both SPIE and Amazon.com in the following published book: " Satellite Data Compression, Communications and Archiving", by Roger W. Heymann (Author), Charles C. Wang (Author), Timothy J. Schmit (Author).

"A lossless compression algorithm for hyperspectral data"

I. Gladkova [1]**, M. Grossberg** [1]
1 CCNY, NOAA/CREST, 138th Street and Convent Avenue, New York, NY 10031
SPIE Meeting August 13-17, 2006. Meeting Journal paper **6300-1**

ABSTRACT
In this paper, which is an expository account of a lossless compression techniques that have been developed over the course of a sequence of papers and talks, we have sought to identify and bring out the key features of our approach to the efficient compression of hyperspectral satellite data. In particular we provide the motivation for using our approach, which combines the advantages of a clustering with linear

modeling. We will also present a number of visualizations which help clarify why our approach is particularly effective on this dataset. At each stage, our algorithm achieves an efficient grouping of the data points around a relatively small number of lines in a very large dimensional data space. The parametrization of these lines is very efficient, which leads to efficient descriptions of data points. Our method, which we are continuing to refine and tune, has to date yielded compression ratios that compare favorably with what is currently achievable by other approaches. A data sample consisting of an entire day's worth of global AQUA-EOS AIRS Level 1A counts (mean 12.9 bit-depth) data was used to evaluate the compression algorithm. The algorithm was able to achieve a lossless compression ratio on the order of 3.7 to 1.

"Adaptive VQ-based Linear Prediction for Lossless Compression of Ultraspectral Sounder Data"

Bormin Huang 1, Alok Ahuja 1, and Mitchell D. Goldberg 2
1 CIMSS, University of Wisconsin-Madison
2 NOAA/NESDIS, Satellite Applications and Research
SPIE Meeting August 13-17, 2006. Meeting Journal paper **6300-2**

ABSTRACT
Contemporary and future ultraspectral sounders represent a significant technical advancement for environmental and meteorological prediction and monitoring. Given their large volume of spectral observations, the use of robust data compression techniques will be beneficial to data transmission and storage. In this paper, we propose a novel Adaptive Vector Quantization (VQ)-based Linear Prediction (AVQLP) method for ultraspectral data compression. The method is compared with several state-of-the-art methods such as CALIC, JPEG-LS and JPEG2000. The compression experiments show that our AVQLP method is the first to surpass the 4 to 1 lossless compression barrier for a selected set of AIRS ultraspectral sounder test data.

"Preliminary Lossless Compression Results with Michelson Interferometer Data"

Timothy J. Schmit
NOAA/NESDIS, STAR (Satellite Applications and Research),
Advanced Satellite Products Branch (ASPB)
Bormin Huang, Y. Sriraja and Hung-Lung Huang
CIMSS, University of Wisconsin-Madison
SPIE Meeting August 13-17, 2006. Meeting Journal paper **6300-4**

ABSTRACT
The next-generation GOES-R (Geostationary Operational Environmental Satellite) HES (Hyperspectral Environmental Suite) Sounder will be either a grating or interferometer design. The HES will be able to provide hourly atmospheric soundings with spatial resolutions of 5 ~ 10 km with higher accuracy than the current geostationary sounder. A number of GOES-R products will be made from the HES data, this information will help both in forecasting and numerical model initializations. Extensive research has been done with lossless data compression with data from a

grating-type ultraspectral instrument. NAST-I aircraft data is chosen for testing data from interferometers until IASI (Infrared Atmospheric Sounding Interferometer) and CrIS (Cross-track Infrared Sounder) are available. Preliminary work at CIMSS with lossless data compression of Michelson Interferometer data achieves compression ratios (CR) above 5.

"Priority-Based Error Correction Using Turbo Codes for Compressed AIRS Data"

I. Gladkova,[1] **M. Grossberg,**[1]
E. Grayver,[2] **D. Olsen,**[2]
N. Nalli,[3] **W.Wolf,**[3] **L. Zhou,**[3]
M. Goldberg[4]
1 CCNY, NOAA/CREST, 138th Street and Convent Avenue, New York, NY 10031
2 The Aerospace Corporation, El Segundo, CA 90245
3 QSS Group, Inc. NOAA/NESDIS, E/RA1 5211 Auth Road, Camp Springs, MD 20746
4 ORA NOAA/NESDIS, E/RA1 5200 Auth Road, Camp Springs, MD 20746
SPIE Meeting August 13-17, 2006. Meeting Journal paper **6300-8**

ABSTRACT
Errors due to wireless transmission can have an arbitrarily large impact on a compressed file. A single bit error appearing in the compressed file can propagate during a decompression procedure and destroy the entire granule. Such a loss is unacceptable since this data is critical for a range of applications, including weather prediction and emergency response planning. The impact of a bit error in the compressed granule is very sensitive to the error's location in the file. There is a natural hierarchy of compressed data in terms of impact on the final retrieval products. For the considered compression scheme, errors in some parts of the data yield no noticeable degradation in the final products. We formulate a priority scheme for the compressed data and present an error correction approach based on minimizing impact on the retrieval products. Forward error correction codes (e.g., turbo, LDPC) allow the tradeoff between error correction strength and file inflation (bandwidth expansion). We propose segmenting the compressed data based on its priority and applying different-strength FEC codes to different segments. In this paper we demonstrate that this approach can achieve negligible product degradation while maintaining an overall 3-to-1 compression ratio on the final file. We apply this to AIRS sounder data to demonstrate viability for the sounder on the next-generation GOES-R platform.

"An analysis of optimal compression for the advanced baseline imager-based on entropy and noise estimation" _

M. Grossberg,[1] **S. Gottipati,**[1] **I. Gladkova,**[1]
M. Goldberg, [2] **L. Roytman.**[1]
1 CCNY, NOAA/CREST, 138th Street and Convent Avenue, New York, NY 10031
2 ORA NOAA/NESDIS, E/RA1 5200 Auth Road, Camp Springs, MD 20746
SPIE Meeting August 13-17, 2006. Meeting Journal paper **6300-22**

ABSTRACT

As new instruments are developed, it is becoming clear that our ability to generate data is rapidly outstripping our ability to transmit this data. The Advanced Baseline Imager (ABI), that is currently being developed as the future imager on the Geostationary Environmental Satellite (GOES-R) series, will offer more spectral bands, higher spatial resolution, and faster imaging than the current GOES imager. As a result of the instrument development, enormous amounts of data must be transmitted from the platform to the ground, redistributed globally through band-limited channels, as well as archived. This makes efficient compression critical. According to Shannon's Noiseless Coding Theorem, an a upper bound on the compression ratio can be computed by
estimating the entropy of the data. Since the data is essentially a stream, we must determine a partition of the data into samples that capture the important correlations. We use a spatial window partition so that as the window size is increased the estimated entropy stabilizes. As part of our analysis we show that we can estimate the entropy despite the high dimensionality of the data. We achieve this by using nearest neighbor based estimates. We complement these a posteriori estimates with a priori estimates based on an analysis of sensor noise. Using this noise analysis we propose an upper bound on the compression achievable. We apply our analysis to an ABI proxy in order estimate bounds for compression on the upcoming GOES-R imager.

Selected Compression results extracted from the 6300-22 paper are:

- "We also analyzed the limits of compression based on computing the entropy of the Meteosat-8 imager."
- "We also obtained a lower bound of 2.6-to-1 on the optimal lossless compression using a linear model applied to this data."
- "With this novel tool we have estimated that 10x10 windows of 11 channel Meteosat-8 data have a theoretical compression bound of 3.1 to 1. Although this number cannot be directly applied to the ABI, the similarity of a significant portion of the data sets to 11 channel Meteosat-8 should be helpful in accessing the risk associated with ABI designs dependent on significantly higher lossless compression."

.

A more recent reflection of space agency research results on satellite sensor earth science data compression will be available in the October 2007 time frame. SPIE will hold its annual meeting of some 3000 talks from August 26-30, 2007 in San Diego, CA. SPIE will publish an edited book of papers from its Conference 6683 "Satellite Data Compression, Communications, and Archiving111". The conference chairs and book editors are: R.W. Heymann NOAA, NESDIS; B. Huang, CIMSS, Univ of Wisconsin/Madison; I Gladkova, CREST, City College, City Univ of NY (CUNY). That conference 6683 book will hold information on satellite Earth science sensor compression research going on in France at CNES of France; in Canada at the Canadian Space Agency (CSA), in the USA at NOAA-NESDIS, in Argentina at the Univ of Buenos Aires, and in Spain at the Univ of Barcelona. That edited book of papers will be available from SPIE at www.spie.org and likely from Amazon.