2005 WSEAS Int. Conf. on REMOTE SENSING, Venice, Italy, November 2-4, 2005 (pp110-116)

Lossless Compression of Ultraspectral Sounder Data using Matching Pursuit based Linear Prediction

Bormin Huang[1*], Alok Ahuja[1], Hung-Lung Huang[1] , and Mitchell D. Goldberg[2]
[1]Space Science and Engineering Center, University of Wisconsin-Madison, Madison, WI 53711, USA
[2]NOAA-NESDIS, Room 810, 5200 Auth Road, Camp Springs, MD 20746, USA

Abstract: - This paper investigates lossless compression of ultraspectral sounder data using Matching Pursuit-based linear prediction compression scheme. The ultraspectral sounder data features good correlations in disjoint spectral regions due to the same type of absorbing gases. The proposed scheme takes advantage of the spectral correlations for achieving higher compression gains. It consists of determining optimal predictors using the Matching Pursuit method followed by linear prediction. The prediction residuals are then rounded and entropy-coded to ensure lossless compression. Numerical results on the standard ultraspectral sounder data show that the proposed scheme outperforms JPEG2000 in terms of compression ratios.

Keywords: - Lossless compression, Ultraspectral sounder, Matching Pursuit, linear prediction, JPEG2000

1 Introduction

The evolution of contemporary and future ultraspectral infrared sounders (e.g. AIRS [1], CrIS [2], IASI [3], GIFTS [4], HES [5] etc.) has made better weather prediction and climate monitoring possible. The sounders generate an unprecedented amount of three-dimensional (3D) data that consists of two spatial and one spectral dimension. Given the large volume of 3D data generated by ultraspectral sounders each day, the use of robust data compression techniques will be beneficial for data transfer and archival. The ultraspectral sounder data is used to retrieve atmospheric temperature, moisture and trace gases profiles, surface temperature and emissivity, as well as cloud and aerosol optical properties. The physical retrieval of these geophysical parameters involves the inverse solution of the radiative transfer equation, which is a mathematically ill-posed problem, i.e. the solution is sensitive to the error and noise in the data [6]. Therefore, there is a need for lossless or near-lossless compression of ultraspectral sounder data to avoid potential retrieval degradation of geophysical parameters due to lossy compression.

Linear prediction has been successfully used for lossless compression of ultraspectral sounder data [7-9]. In this paper, we present a lossless compression scheme that uses Matching Pursuit (MP) [10] along with linear prediction for compression of ultraspectral sounder data, which is termed Optimized Orthogonal Matching Pursuit based Linear Prediction (OOMP-LP). Matching Pursuit is a method for signal approximation that has been utilized in video compression [11,12]. For ultraspectral sounder data compression, optimal spectral predictors are determined for each channel using the Optimized Orthogonal Matching Pursuit (OOMP) [13] method. Linear prediction is then applied for each channel using the predictors determined via OOMP. Thereafter, the prediction residuals are rounded and entropy-coded to ensure lossless compression. For comparison, the ultraspectral sounder data is also compressed with the state-of-the-art wavelet-based compression standard, JPEG2000 [14]. Compression results are presented for the ultraspectral sounder data by use of OOMP-LP and JPEG2000.

The rest of the paper is organized as follows. Section 2 describes the ultraspectral sounder data used in this study. Section 3 highlights the proposed compression scheme, while Section 4 presents the results of the simulations. Section 5 summarizes the paper.

2 Ultraspectral Sounder Data

The ultraspectral sounder data could be generated from either a Michelson interferometer (e.g. CrIS, IASI and GIFTS) or a grating spectrometer (e.g. AIRS). The ultraspectral sounder data set with 2107 AIRS channels was prepared at the direction of NOAA to serve as a standard test set for lossless compression studies. The data is publicly available via anonymous ftp [15]. It consists of 10 digital count granules, five daytime and five nighttime, selected from representative geographical regions of the Earth. Their locations, UTC times and local time adjustments are listed in Table 1.

This standard ultraspectral sounder data set is obtained from NASA AIRS digital counts collected on March 2, 2004. The AIRS data includes 2378 infrared channels in the 3.74 to 15.4 µm region of the spectrum. A day's worth of AIRS data is divided into 240 granules, each of 6 minute durations. The AIRS digital count data ranges from 12-14 bits for different channels. More information regarding the AIRS instrument may be acquired from the NASA AIRS website [16]. To make the selected data more generic to other ultraspectral sounders, 271 bad channels identified in the supplied AIRS infrared channel properties file are excluded, assuming that they occur only in the AIRS sounder. Each resulting granule is saved as a binary file, arranged as 2107 channels, 135 scan lines, and 90 cross-track footprints per scan line, i.e. there are a total of 135 x 90 = 12,150 footprints per channel. Figure 1 shows the AIRS digital counts at wavenumber 800.01cm^{-1} for the 10 selected granules on March 2, 2004. In these granules, coast lines are depicted by solid curves, and multiple clouds at various altitudes are shown as different shades of colored pixels.

3 Compression Scheme

For ultraspectral sounder data, the spectral correlation is generally much stronger than the spatial correlation [17]. The proposed OOMP-LP scheme is geared towards exploiting these correlations to achieve high compression gains. It consists of the following three steps:

(I) *Matching Pursuit Predictor Determination:* The 3D ultraspectral data of size n_c x n_x x n_y is reshaped into a 2D data set of size n_c channels x n_s pixels, where $n_s = n_x \times n_y$. Thus each spatial frame with n_s pixels is considered a vector that is to be predicted. Matching Pursuit is used to determine the optimal set of predictors for any given vector. Given a signal *f*, MP represents it as a linear combination of basis signals, called *atoms*, selected from an over-complete basis set, called the *dictionary*, given by

$$\mathbf{D} = \{g_\gamma\}_{\gamma\in\Gamma}, \tag{1}$$

where $\{g_\gamma\}$ are the basis signals, $\Gamma = \{1,\ldots,N\}$ is the set of all indices, and *N* is the dictionary size. After *j* iterations of the MP method, the signal *f* is approximated as

$$f = \sum_{i=1}^{j} \alpha_i g_{\gamma_i} + R_j, \tag{2}$$

where α_i are the coefficients, g_{γ_i} are the chosen basis signals, and R_j is the current residual error. Starting with initial values of $R_0 = f$, $f_0 = 0$, and $j = 0$, the MP method consists of the following steps:

Step 1) Compute the inner products $\{\langle R_j, g_\gamma\rangle\}_{\gamma\in\Gamma}$.

Step 2) Determine γ_{j+1} such that the inner product of $g_{\gamma_{j+1}}$ with the residual R_j is maximized, i.e.

$$\left|\langle R_j, g_{\gamma_{j+1}}\rangle\right| = \sup_{g_\gamma\in\mathbf{D}}\left|\langle R_j, g_\gamma\rangle\right|.$$

Step 3) Set $f_{j+1} = f_j + \langle R_j, g_{\gamma_{j+1}}\rangle g_{\gamma_{j+1}}$, and

$$R_{j+1} = R_j - \langle R_j, g_{\gamma_{j+1}}\rangle g_{\gamma_{j+1}}.$$

Step 4) Set $j = j + 1$.

Step 5) If some convergence criterion is satisfied, STOP, else go to Step 1.

A shortcoming of the above method is that the resulting approximation after any finite iterations will be suboptimal in the following sense. Since the family of atoms $g_{\gamma_i}; i = 1,\ldots, j$ is, in general, not orthogonal, the residual R_j will not be in general orthogonal to the subspace generated by these atoms. The OOMP method [13] is an improvement over MP in that, at each iteration, it yields a residual orthogonal to the subspace of the chosen atoms, and also provides the dictionary atoms minimizing the norm of the corresponding residual error. For ultraspectral sounder data compression, each spatial frame is predicted by atoms chosen using the OOMP method. OOMP's selection criterion for atoms is implemented using the well-known Gram-Schmidt technique [18]. The selection of predictor channels for any given channel proceeds iteratively, with the dictionary for each channel consisting of all channels that have been previously predicted. Therefore, it is possible to use the same channel to predict two or more channels. For each channel, n_p predictors are determined via OOMP and a list of predictor channels $P_{i,k},\ i = 1,\ldots,n_c; k = 1,\ldots,n_p$ is formed.

(II) *Linear Prediction:* Linear prediction appears to be a good whitening tool for ultraspectral sounder data. It employs a set of neighboring pixels to predict the current pixel [19, 20]. As previously mentioned, the spectral correlation is generally much stronger than the spatial correlation for ultraspectral sounder data. Thus, it is natural to predict a channel as a linear combination of other channels. Interband prediction schemes that explore spectral correlations of multispectral imager data have been investigated in the past [21,22]. In the proposed OOMP-LP scheme, the predictor channel list determined via OOMP is utilized for linear prediction of each channel. The problem can be formulated as

$$\hat{\mathbf{X}}_i = \sum_{k=1}^{n_p} c_k \, \mathbf{X}_{P_{i,k}} \qquad \text{or}$$

$$\hat{\mathbf{X}}_i = \boldsymbol{X}_{\boldsymbol{m}} \mathbf{C} \,, \tag{3}$$

where $\hat{\mathbf{X}}_i$ is the vector of the current channel representing a 2D spatial frame, $\mathbf{X}_{P_{i,k}}$ is the *kth* predictor channel as determined through OOMP, $\boldsymbol{X}_{\boldsymbol{m}}$ is the matrix consisting of n_p predictor channels, and $\mathbf{C}$ is the vector of the prediction coefficients. The prediction coefficients are obtained from

$$\mathbf{C} = (\boldsymbol{X}_{\boldsymbol{m}}{}^T \boldsymbol{X}_{\boldsymbol{m}})^{\dagger} (\boldsymbol{X}_{\boldsymbol{m}}{}^T \hat{\mathbf{X}}_i) \,, \tag{4}$$

where the superscript $\dagger$ represents the pseudo-inverse that is robust against the case of the matrix being ill-conditioned [18]. The prediction error is the rounded difference between the original channel vector and its predicted counterpart.

(III) *Entropy Coding:* After the linear prediction stage, the prediction error is entropy-coded using an adaptive arithmetic coder [23].

4 Results

The ultraspectral sounder data set is compressed using both OOMP-LP and JPEG2000 schemes. Table 2 shows the compression ratios of ten tested granules obtained via OOMP-LP compression using 40 predictors for each channel. For comparison, the compression results for JPEG2000 Part I are also shown. Since the JPEG2000 Part I codec supports 2D compression only, each granule with the size of 2107 channels x 135 scan lines x 90 footprints is converted into 2D with the size of 2107 channels x 12150 samples via a horizontal zigzag scan. It is then compressed using the JPEG2000 encoder [24]. As can be seen in Table 2, our proposed scheme produces significantly higher compression ratios than JPEG2000.

5 Conclusion

Optimized Orthogonal Matching Pursuit based Linear Prediction (OOMP-LP) is developed for lossless compression of ultraspectral sounder data. The OOMP-LP method is investigated to determine optimal predictor channels for linear prediction. The average lossless compression ratio produced by OOMP-LP on the standard ultraspectral sounder data set is 27% higher than that of the state-of-the-art JPEG2000 compression standard, demonstrating the promise of using OOMP-LP for ultraspectral sounder data compression.

Acknowledgement

This work is sponsored by NOAA-NESDIS under grant NA07EC0676 and has been prepared in support of the NOAA-NESDIS GOES-R hyper-spectral sounder data compression research group led by Roger Heymann of its Office of Systems Development and Tim Schmit of its Office of Research and Applications.

References:

[1] H. H. Aumann and L. Strow, "AIRS, the first hyperspectral infrared sounder for operational weather forecasting," in *Proceedings of IEEE Aerospace Conference*, vol. 4, pp. 1683-1692, 2001.

[2] H. J. Bloom, "The Cross-track Infrared Sounder (CrIS): a sensor for operational meteorological remote sensing," in *Proceedings of the 2001 International Geoscience and Remote Sensing Symposium*, vol 3, pp. 1341-1343, 2001.

[3] T. Phulpin, F. Cayla, G. Chalon, D. Diebel, and D. Schlüssel, "IASI onboard Metop: Project status and scientific preparation,"in *Twelfth International TOVS Study Conference, Lorne, Victoria, Australia, 26 February-4 March 2002*, pp. 234-243.

[4] W. L. Smith, F. W. Harrison, D. E. Hinton, H. E. Revercomb, G. E. Bingham, R. Petersen, and J. C. Dodge, "GIFTS - the precursor geostationary satellite component of the future earth observing system," in *Proceedings of the 2002 International Geoscience and Remote Sensing Symposium*, vol. 1, pp 357-361, 2002.

[5] B. Huang, H. L. Huang, H. Chen, A. Ahuja, K. Baggett, T. J. Schmit, and R. W. Heymann, "Lossless data compression studies for NOAA Hyperspectral Environmental Suite using 3D integer wavelet transforms with 3D embedded zerotree coding," in *the Third International*

Symposium on Multispectral Image Processing and Pattern Recognition, *Proc. of SPIE*, vol. 5286, pp. 305-315, 2003.

[6] B. Huang, W. L. Smith, H.-L. Huang, H. M. Woolf, "Comparison of linear forms of the radiative transfer equation with analytic Jacobians," *Applied Optics*, vol. 41, no. 21, pp. 4209-4219, July 2002.

[7] B. Huang, A. Ahuja, H.-L. Huang, T. J. Schmit, and R. W. Heymann, "Predictive partitioned vector quantization for hyperspectral sounder data compression," in *SPIE Annual Meeting, 2-6 August 2004, Denver, Proceedings of SPIE*, vol. 5548, pp. 70-77, 2004.

[8] B. Huang, A. Ahuja, H.-L. Huang, T. J. Schmit, and R. W. Heymann, "Fast precomputed VQ with optimal bit allocation for lossless compression of ultraspectral sounder data," *Proceedings of the 2005 IEEE Data Compression Conference*, pp. 408-417, 2005.

[9] B. Huang, A. Ahuja, H.-L. Huang, *Lossless Compression of Ultraspectral Sounder Data*, Hyperspectral Data Compression, edited by G. Motta and J. Storer, Springer-Verlag, 2005.

[10] S. G. Mallat, and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, Dec. 1993.

[11] R. Neff, A. Zakhor, and M. Vetterli, "Very low bit rate video coding using matching pursuits," in *Proceedings of SPIE*, vol. 2308, pp. 47-60, 1994.

[12] R. Neff, and A. Zakhor, "Matching pursuit video coding. I. Dictionary approximation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 13-26, 2002.

[13] L. Rebello-Neira, and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Proc. Letters*, vol. 9, no. 4, April 2002.

[14] ISO/IEC 15444-1: "Information technology - JPEG2000 image coding system-part 1: Core coding system", 2000.

[15] ftp://ftp.ssec.wisc.edu/pub/bormin/Count.

[16] http://www-airs.jpl.nasa.gov.

[17] B. Huang, A. Ahuja, H.-L. Huang, T. J. Schmit, and R. W. Heymann, "Lossless compression of 3D hyperspectral sounding data using context-based adaptive lossless image codec with Bias-Adjusted Reordering," *Opt. Eng.*, vol. 43, no. 9, pp. 2071-2079, 2004.

[18] G. H. Golub and C.F. Van Loan, *Matrix Computations*, John Hopkins University Press, 1996.

[19] X. Wu and K. Barthel, "Piecewise 2D autoregression for predictive image coding," in *Proc. Int. Conf. Image Processing*, vol. 3, pp. 901-905, Oct. 1998

[20] X. Li and M. Orchard, "Edge-directed prediction for lossless compression of natural images," *IEEE Trans. Image Processing*, vol. 10, pp. 813-817, Jun. 2001.

[21] X. Wu and N. Memon, "Context-based lossless interband compression – extending CALIC," *IEEE Trans. Image Processing*, vol. 9, pp. 994-1001, Jun. 2000.

[22] J. Mielikainen, P. Toivanen, and A. Kaarna, "Linear prediction in lossless compression of hyperspectral images," *Opt. Eng.*, vol. 42, no. 4, pp. 1013-1017, April 2003.

[23] I. H. Witten, R. M. Neal, and J. C. Cleary, "Arithmetic Coding for Data Compression," *Comm. ACM*, vol. 30, no. 6, pp. 520-541, June 1987.

[24] http://www.kakadusoftware.com/.

Granule 9	00:53:31 UTC	-12 H	(Pacific Ocean, Daytime)
Granule 16	01:35:31 UTC	+2 H	(Europe, Nighttime)
Granule 60	05:59:31 UTC	+7 H	(Asia, Daytime)
Granule 82	08:11:31 UTC	-5 H	(North America, Nighttime)
Granule 120	11:59:31 UTC	-10 H	(Antarctica, Nighttime)
Granule 126	12:35:31 UTC	-0 H	(Africa, Daytime)
Granule 129	12:53:31 UTC	-2 H	(Arctic, Daytime)
Granule 151	15:05:31 UTC	+11 H	(Australia, Nighttime)
Granule 182	18:11:31 UTC	+8 H	(Asia, Nighttime)
Granule 193	19:17:31 UTC	-7 H	(North America, Daytime)

Table 1. Ten selected AIRS granules for ultraspectral sounder data compression studies.

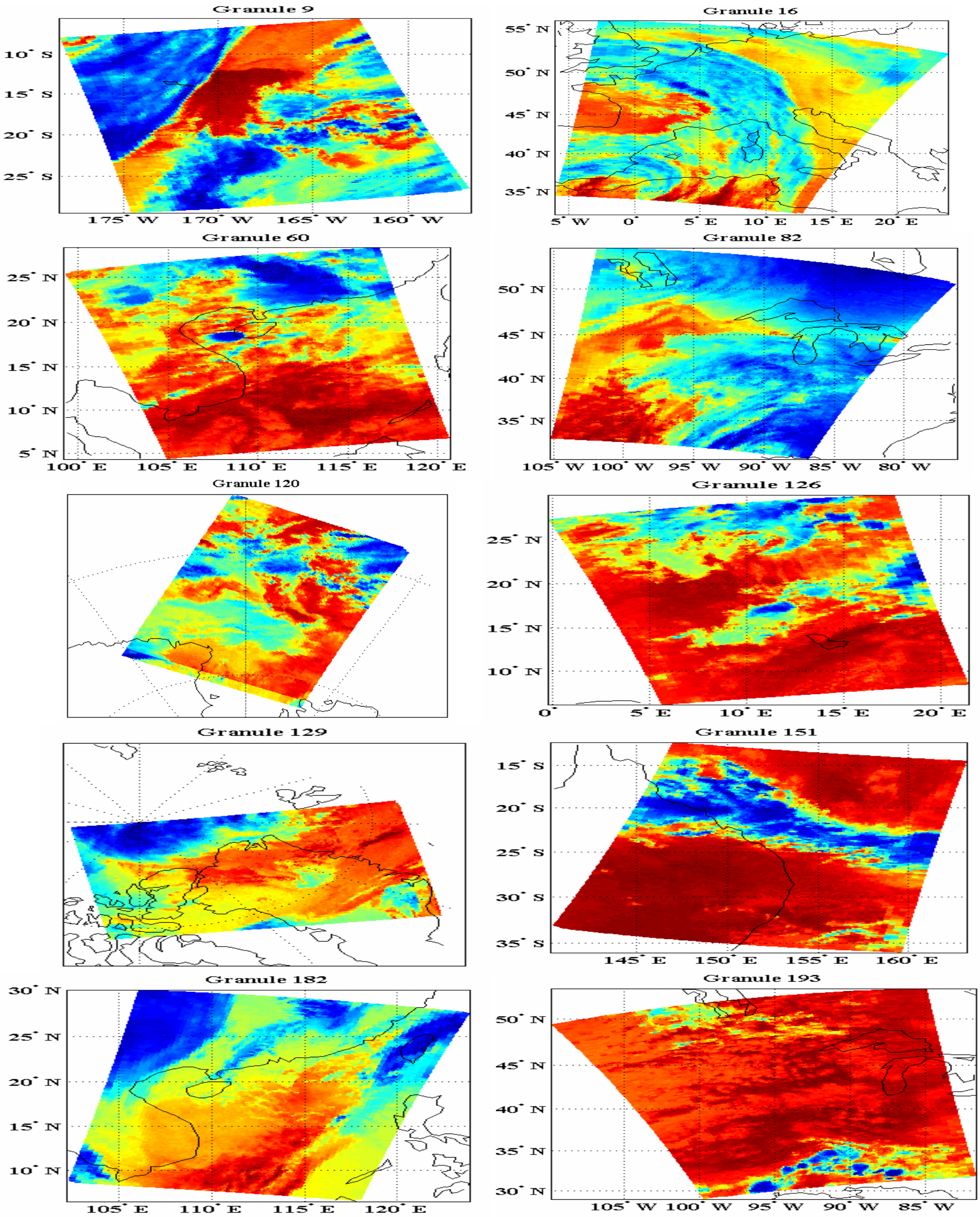

Fig. 1. AIRS digital counts at wavenumber 800.01cm^{-1} for the ten selected granules on March 2, 2004.

Granule	JPEG2000	OOMP-LP
9	2.38	3.21
16	2.44	3.30
60	2.29	2.78
82	2.52	3.11
120	2.40	3.26
126	2.29	2.81
129	2.52	2.81
151	2.33	2.79
182	2.25	2.74
193	2.30	3.33
Average	**2.37**	**3.01**

Table 2. Compression ratios of JPEG2000 and OOMP-LP for the ten tested AIRS granules.