P1.14

COMPARISON OF ARITHMETIC CODING AND PREFIX CODING WITH THE CCSDS LOSSLESS COMPRESSION RECOMMENDATION FOR SATELLITE DATA

Bormin Huang*, Alok Ahuja, and Hung-Lung Huang
CIMSS, University of Wisconsin-Madison
*bormin@ssec.wisc.edu

Timothy J. Schmit and Roger W. Heymann
NOAA, National Environmental Satellite, Data, and Information Service

1. INTRODUCTION

Multispectral and hyperspectral imagery is used in a wide variety of remote sensing applications such as global environmental monitoring, mapping, charting and land use planning (Vaughn et al. 1995). The evolution of sensor technology has dramatically improved the quality of multispectral imagery through increase in spectral, spatial and radiometric resolutions of earth observations. This is accompanied with a substantial increase in data volumes. Compression techniques are thus beneficial for data transmission and storage.

In a lossless compression scheme, there is perfect reconstruction of the original data after decompression. The Consultative Committee for Space Data Systems (CCSDS) adopted its first lossless data compression recommendation (CCSDS 1997), that consists of two parts – a preprocessor, and an adaptive entropy coder based on the Rice algorithm. In this paper, we provide a comparison of two newer entropy coding schemes, prefix coding (Moffat et al. 1995) and arithmetic coding (Said 2004), with the CCSDS-Rice entropy coder. Both prefix coding and arithmetic coding require minimal side information to be sent to the decoder, and adapt extremely well to the source statistics of the input images.

The paper is arranged as follows. Section 2 describes the multispectral and hyperspectral image data used for comparison. Section 3 highlights the different entropy coding schemes while Section 4 provides the results of the performance comparison. Section 5 summarizes the paper.

2. DATA

The multispectral image data set from the NASA Moderate Resolution Spectroradiometer (MODIS) (King et al. 1992), and the hyperspectral image data set from the NASA JPL Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Vane 1987) are used in this study. The MODIS data set was from the website http://modis.gsfc.nasa.gov, whereas the AVIRIS data set from http://aviris.jpl.nasa.gov.

The MODIS instrument aboard the Terra satellite consists of 36 spectral bands at spatial resolutions of 250m, 500m, and 1km. Two 12-bit MODIS radiance data of size 900 by 1500 pixels are acquired on Sept. 20, 2001. We used 17 bands (20-36) at 1km resolution. Figure 1 shows band 32 (12.02 µm) of the two data sets.

The AVIRIS optical sensor has 224 contiguous spectral channels from 400 to 2500nm at 10nm intervals. These spectra have a 20m spatial resolution from a NASA ER-2 aircraft flying at 20 km altitude. Figure 2 depicts two AVIRIS scenes, Cuprite and Jasper Ridge, which are used in this study.

3. COMPRESSION SCHEMES

The object of this study was to compare two newer entropy coding schemes, namely prefix coding and arithmetic coding, with the Rice entropy coding scheme. Therefore, for a fair comparison the same preprocessor was used before the entropy coding part for all schemes. We used the preprocessor specified in the CCSDS recommendation whose function is to decorrelate the image pixels and reformat them into non-negative integers. The output from the pre-processor is then fed to each entropy coder.

An entropy coder assigns *codewords* to the provided *symbols* based on their probabilities of occurrence. Shorter codewords are assigned to symbols that occur with a higher probability and vica versa. The three different entropy coding schemes used in this study are highlighted below.

3.1 *Rice Coding*

This method generates variable-length codewords utilizing Rice's adaptive coding technique (Rice et al. 1971). Different coding options are tried concurrently for the input symbols, and the option that yields the shortest codeword for that set of symbols, is chosen for transmission. A detailed description of the algorithm is provided in the CCSDS recommendation (CCSDS 1997).

3.2 *Prefix Coding*

Prefix coding assigns variable-length codewords to symbols such that no codeword is a proper prefix of any other codeword. Huffman codes (Huffman 1952) are an example of prefix codes. The prefix codes used in this study are minimum-redundancy codes (Moffat et al. 1995). These codes allow extremely fast encoding and decoding. Unlike Huffman coding, the codeword table does not need to be sent to the decoder. Only the symbols table and the length of their corresponding codewords are to be sent to the decoder. Approaches to construction of memory efficient prefix codes with fast encoding and decoding times have been suggested in literature (Moffat et al. 1997).

3.3 *Arithmetic Coding*

Arithmetic coding is a technique utilizing the concept of interval subdivision, where successive input symbols are encoded as intervals on the range [0,1) based on their probability of occurrence. It is superior to Huffman coding in the sense that it can assign a fractional number of bits for the codewords of the symbols, whereas in Huffman coding an integral number of bits have to be assigned to a codeword of a symbol. An adaptive arithmetic coder is used in these tests.

4. COMPARISON RESULTS

All tests were performed on a linux machine with an AMD Opteron processor running at 1804Mhz. The last 100 channels (125-224) of the two AVIRIS scenes, and 17 bands of the MODIS granules were compressed using different entropy coders after preprocessing. For the Rice algorithm, the Szip compression software was used (http://hdf.ncsa.uiuc.edu/doc_resource/SZIP). The arithmetic codec was based on an implementtation by Amir Said (http://www.cipr.rpi.edu/research/SPIHT/EW_Code/FastAC.zip) with extension to 16-bit data sources.

Figure 3 shows the average compression ratios obtained on 100 channels (125-224) of the AVIRIS scenes, Cuprite and Jasper Ridge. It is seen that Rice coding gives the lowest compression ratios. Furthermore, arithmetic coding outperforms Rice coding by 20% and 19% on average for Cuprite and Jasper Ridge respectively. The disparity in compression ratios can be explained by the fact that arithmetic coding is better able to adapt to the incoming source statistics. The compression times for arithmetic coding and Rice coding are quite compatible as shown in Table 1.

The average compression ratios for the MODIS data are depicted in Figure 4. Arithmetic coding gives superior compression ratios when compared to prefix coding and Rice coding. Moreover, the compression times for arithmetic coding and Rice coding are also comparable as shown in Table 2.

5. SUMMARY

Presented is a comparison of two newer entropy coding schemes, namely prefix coding and arithmetic coding with the CCSDS-Rice coding. Experiments show that arithmetic coding and prefix coding provide superior compression performance than Rice coding, while arithmetic coding is comparable to Rice coding in terms of execution speed.

Acknowledgement

This research is supported by National Oceanic and Atmospheric Administration's National Environmental Satellite, Data, and

Information Service under grant NA07EC0676. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. Government position, policy, or decision.

References

Consultative Committee for Space Data Systems, CCSDS, 1997: 120.0-B-1, Lossless data compression: recommendation for space data systems standards (blue book).

Huffman, D.A., 1952: A method for construction of minimum redundancy codes, *Proc. IRE*, **40**, 1098-1101.

King, M.D., Kauffman Y.J., Menzel W.P. and Tanre D., 1992: Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS), *IEEE Trans. Geo. Remote Sens.*, **30**, 2-27.

Moffat, A. and Katajainen J., 1995: In-place calculation of minimum-redundancy codes, *Proc. Workshop Algo. and Data Struc.*, 393-402.

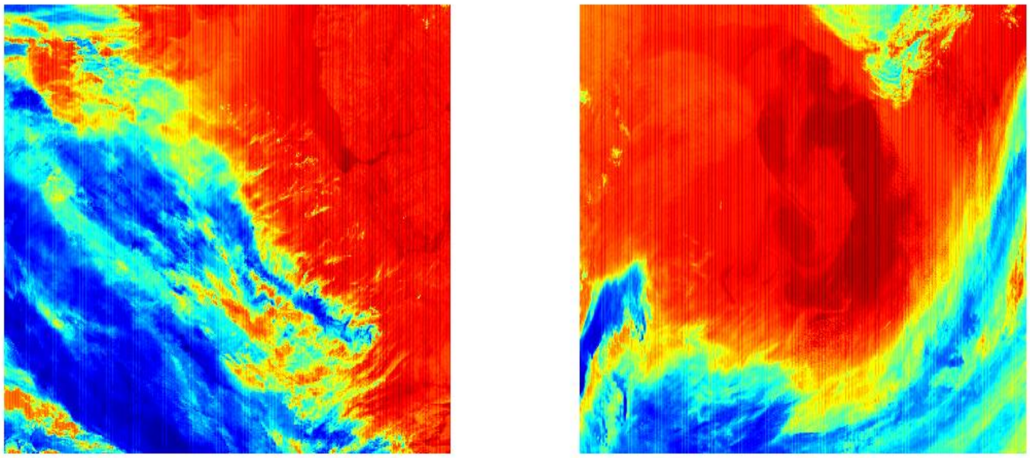
Moffat, A. and Turpin A., 1997: On the implementation of minimum redundancy prefix codes, *IEEE Trans. Comm.*, **45**, 1200-1207.

Rice, R.F. and Plaunt J.R., 1971: Adaptive variable length coding for efficient compression of spacecraft television data, *IEEE Trans. Comm. Tech.*, V.COM-19, 1, 889-897.

Said, A., 2004: Introduction to Arithmetic coding theory and practice, *Hewlett-Packard Laboratories Report*, HPL-2004-76.

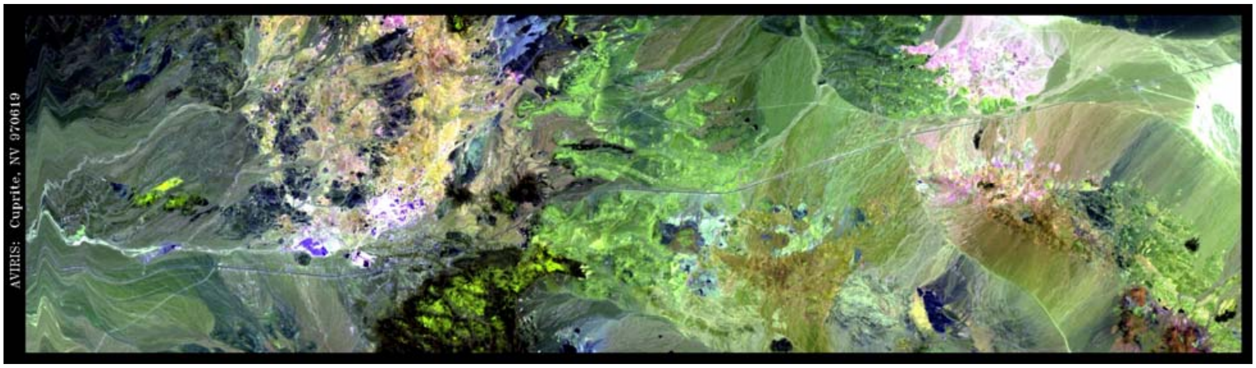
Vane, G., 1987: Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), *Jet Propulsion Laboratory Publ.*, 87-38.

Vaughn, V.D. and Wilkinson T.S., 1995: System considerations for multispectral image compression designs, *IEEE Signal Proc. Mag.*, **12**, 19-31.

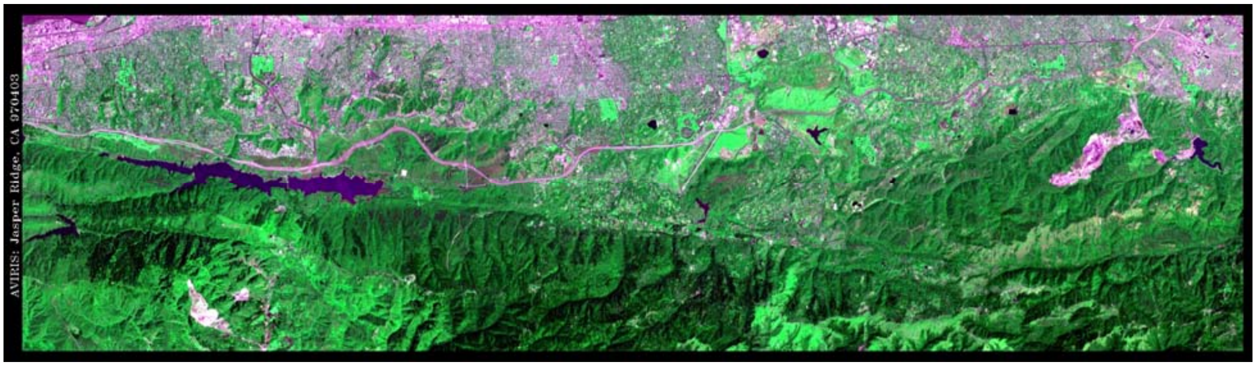


(a) (b)

Figure 1. Two MODIS radiance data at band 32 (12.02 µm) with 1km spatial resolution on Sept. 20, 2001.



(a)



(b)

Figure 2. Color composite image of AVIRIS scenes. (a) Cuprite 1997 (b) Jasper Ridge 1997.

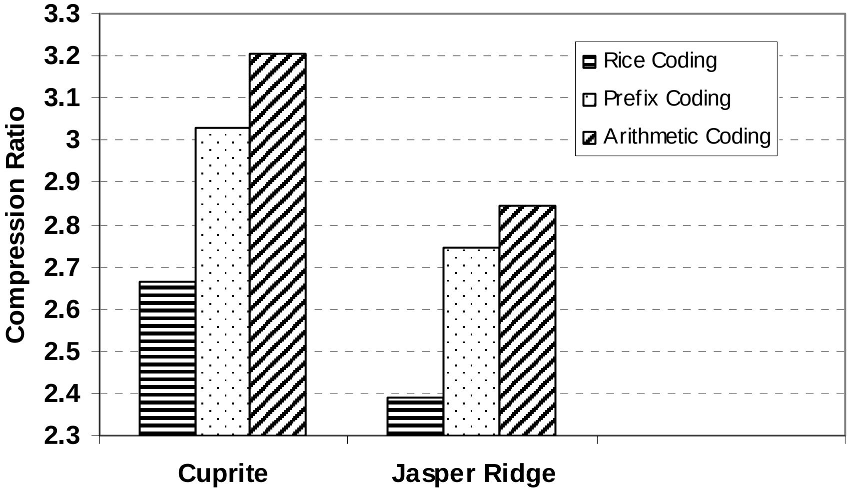


Figure 3. Average compression ratios for 100 channels (125-224) of AVIRIS scenes: Cuprite and Jasper Ridge.

	Rice Coding	**Arithmetic Coding**
Cuprite	**4.1707**	**5.2079**
Jasper Ridge	**4.8201**	**5.7706**

Table 1. Compression time in seconds for last 100 channels of AVIRIS scenes.

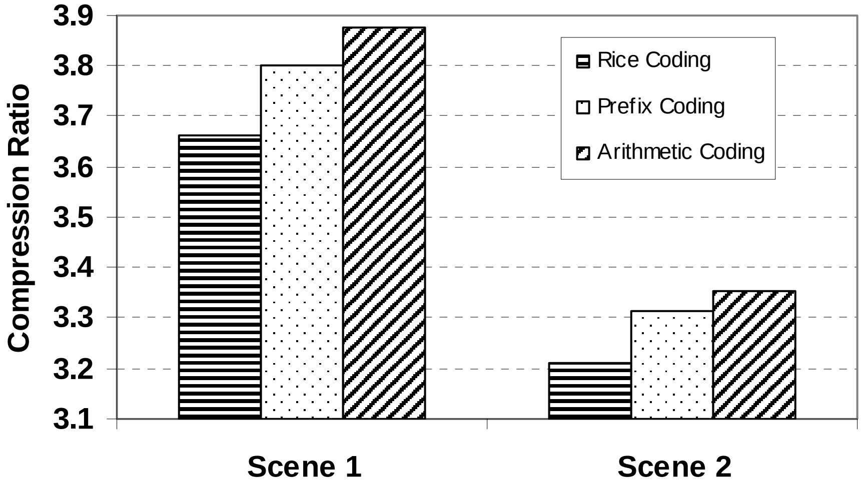


Figure 4. Average compression ratios for 17 bands of two MODIS Level 1A granules at 1km spatial resolution.

	Rice Coding	**Arithmetic Coding**
Scene 1	**0.9794**	**1.1646**
Scene 2	**1.0063**	**1.1899**

Table 2. Compression time (seconds) for 17 bands of MODIS Level 1A granules.