Evaluation of Near Lossless Feature of Two On-board Data Compression Algorithms: SAMVQ and HSOCVQ

Shen-En Qian, Martin Bergeron and Allan Hollinger
Canadian Space Agency
6767 Route de l'Aeroport, St-Hubert, Quebec, J3Y 8Y9
CANADA

Abstract: - To deal with the large data volume of hyperspectral sensors the Canadian Space Agency (CSA) developed two near lossless compression algorithms. This paper analyzes compression errors introduced by the two compression algorithms and assesses their near lossless feature. Experimental results show that errors introduced by the two compression algorithms are smaller than the intrinsic noise of the original data caused by the instrument noise and other noise sources such as calibration and atmospheric correction errors. This level of compression errors is expected to have small to negligible impact on remote sensing applications compared to the intrinsic noise of the original data.

Key-Words: - Hyperspectral imagery, data compression, near lossless, vector quantization.

1 Introduction

In the development of a spaceborne hyperspectral imager, one of the challenges is the large volume of data generated by the imager on-board the satellite, which exceeds the downlink capacity, and may quickly exhaust the on-board storage capacity. To deal with this problem, the Canadian Space Agency has been developing data compression technology for hyperspectral data for a number of years. Compression techniques for operational use of multi-dimensional data have been developed [1-9]. Recently, two data compression techniques for multi-dimensional sensor data on-board satellites have been developed and patented. They are the Successive Approximation Multi-stage Vector Quantization (SAMVQ) [10, 11] and the Hierarchical Self-Organizing Cluster Vector Quantization (HSOCVQ) [12, 13]. These techniques were designed specially for the near lossless compression of hyperspectral imagery. They are simple, fast and near lossless even at high compression ratio. Both of them can be used for on-board and on-ground data compression of multi-dimensional data and are optimized for hyperspectral data. The CSA is planning to place a near lossless data compressor on-board a proposed Canadian hyperspectral satellite using these techniques to reduce the requirement for on-board storage and to better match the available downlinking capacity.

One of the unique features of SAMVQ and HSOCVQ is that they allow control of the errors (noise) introduced during the compression process. For remote sensing applications, compression is considered near lossless provided the errors introduced by the compression are no larger than the intrinsic noise in the original data caused by the instrument noise and other noise sources from pre-processing such as calibration and atmospheric correction. This is critical and highly valuable for data compression of satellite hyperspectral data, where compression error is expected to have a small or negligible impact on remote sensing applications. Since an original datacube is not exempt from instrument noise and other noise sources, these errors propagate into the remote sensing products derived from the original data.

In order to evaluate the impact of the compression techniques on remote sensing applications a multi-disciplinary user acceptability study has been carried out [14]. Eleven hyperspectral data users, covering a wide range of application areas and a variety of hyperspectral sensors, assessed the acceptability of the compressed data qualitatively and quantitatively using their well-understood hyperspectral datacubes and predefined evaluation criteria. For their applications, the users accepted most of the compressed datacubes as they provided the same amount of information as the original datacube.

In this paper, we analyze the error introduced by the two compression algorithms and evaluate their near lossless feature. The errors introduced by the compression algorithms will be evaluated by comparing them with the intrinsic noise of the original data that is caused by the instrument noise and other noise resources.

2 Test Hyperspectral Data Set

A low altitude data set acquired using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in the Greater Victoria Watershed District, Canada on August 12, 2002 (the information on the data set is available at http://aviris.jpl.nasa.gov/ql/list02.html) was used. The spatial resolution of the data set is 4m x 4m with a peak SNR of 1000:1 in the visible and near infrared (VNIR) region. A spectral subset was selected to remove redundant and bad bands. This reduced the data from 224 bands to 204 bands, including the original bands 6-31 (423.04 nm - 664.79 nm, VNIR), bands 35-96 (673.64 nm - 1258.39 nm) and bands 98-213 (1263.72 nm - 2399.48 nm) for short wavelength infrared (SWIR). A 28m x 28m spatial resolution data set was derived by spatially averaging the 4m x 4m data set. The signal-to-noise ratio (SNR) of the derived data set is $1000 ? \sqrt{49}$ =7000:1. Figure 1 (a) shows the derived datacube. The spatial size of the datacube is 292 lines with 121 pixels per line. The datacube is stored in 16-bit digital numbers (DN). This datacube is viewed as a noise free datacube in this paper, as the noise is too small to have a significant impact on the evaluation.

Figure 1 (b) shows a datacube the same as (a) except that its SNR is 600:1 generated by adding simulated instrument noise and other possible noise sources to (a). An additive noise model was used. This noise-added datacube was used as an original datacube for compression and evaluation of the near lossless feature of the compression algorithms in this paper. It is considered to be representative of a real satellite hyperspectral data set, since SNR for such an instrument is likely to be around that level. The noise of the original datacube caused by the instrument noise and other noise sources is referred to as intrinsic noise in this paper. Figure 1 (c) shows an intrinsic noise image at band 38

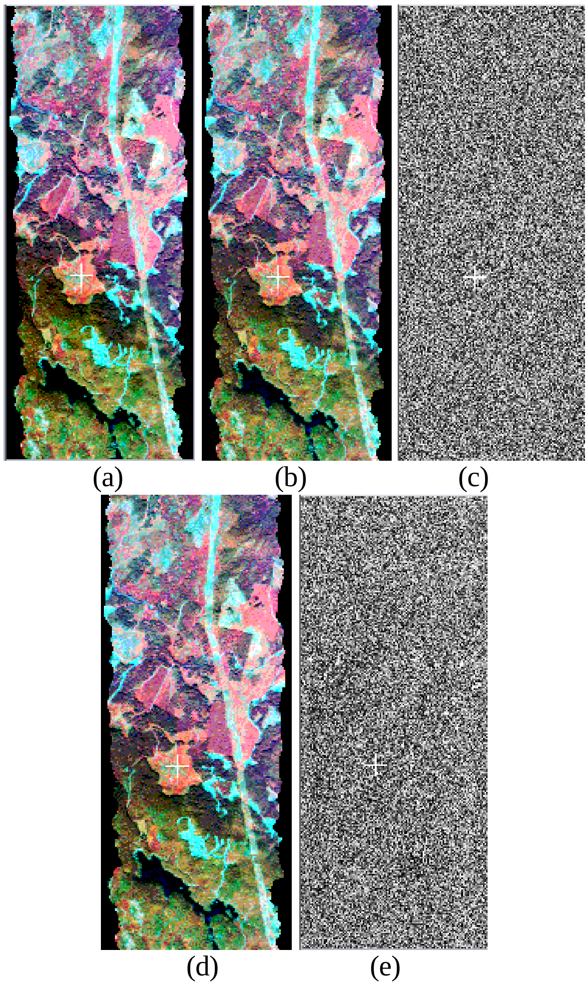


Figure 1 AVIRIS Greater Victoria Watershed District datacube, intrinsic noise and compression error images

(a) AVIRIS Greater Victoria Watershed District datacube with SNR of 7000:1 after aggregation, which is viewed as noise free (signal only), displayed band 38 (702.2 nm) as Red, band 20 (557.9 nm) as Green and band 2 (432.6 nm) as Blue;
(b) Noise added datacube from (a) with SNR of 600:1, which is used as an original datacube for compression;
(c) Intrinsic noise image of the original datacube (with SNR 600:1) displayed at band 38;
(d) Reconstructed datacube compressed using SAMVQ at compression ratio of 20:1;
(e) Difference image (compression error) between the original datacube and the reconstructed datacube compressed using SAMVQ at compression ratio of 20:1 displayed at band 38.

(702.2 nm) of the noise datacube, which was obtained by subtracting the datacube of SNR 600:1 from the noise free datacube.

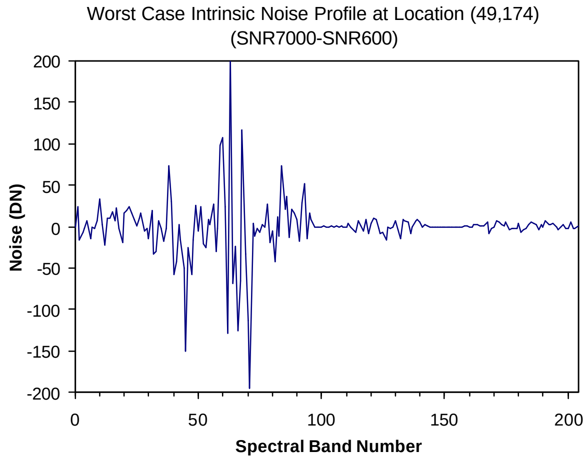
3 Near Lossless Feature of SAMVQ

The original datacube was compressed using the SAMVQ algorithm at compression ratio 20:1. Then the compressed data was decompressed to obtain the reconstructed version for evaluation. Figure 1 (d) shows the reconstructed datacube. It is difficult to visually distinguish the difference between the original and the reconstructed datacubes. Figure 1 (e) shows the difference (compression error) between the original datacube and the reconstructed datacube at band 38. The pattern of the compression error image looks similar to that of the intrinsic noise image. There are no apparent structures to the error image.

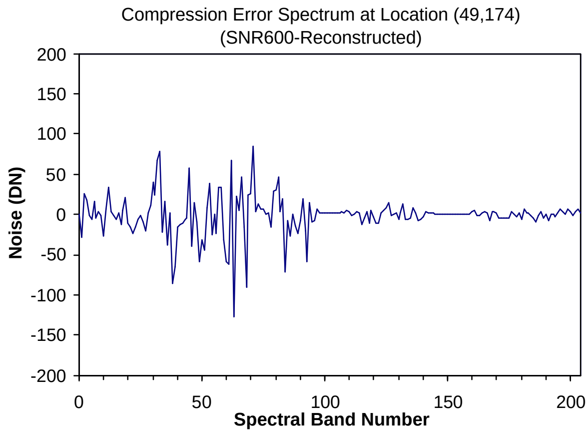
The intrinsic noise of the original datacube was analyzed band-by-band and pixel-by-pixel. Figure 2 (a) shows the worst-case noise profile at location (49, 174) of the intrinsic noise datacube as a function of spectral band number. The noise magnitudes for the VNIR bands and the beginning of SWIR bands are large. The maximum value of the noise is 204 DN at band 66, and minimum value of the noise is -196 DN at band 71. The noise values are between -15 DN and 15 DN for the bands between 100 and 204.

Figure 2 (b) shows the compression error (or noise) profile of the reconstructed datacube with a compression ratio of 20:1 at the same location as Figure 2 (a). It can be seen that the error introduced due to the compression is smaller than the intrinsic noise of the original datacube across the spectral bands. The maximum value of the compression error is 85 DN at band 71, and minimum value of the compression error is -127 DN at band 63. The compression error values for bands between 100 and 204 are in the same range as the intrinsic noise of the original datacube. They are between -15 DN and 15 DN.

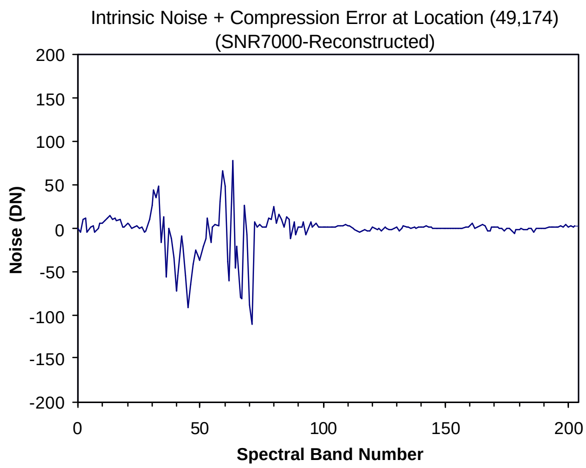
After compression/decompression, the reconstructed data contains both the intrinsic noise of the original datacube and the compression error (or compression noise). In this paper the combination of the intrinsic noise and compression error is referred to as overall noise. This is the final noise budget of the datacube, if the reconstructed data is



(a)



(b)



(c)

Figure 2 Profiles of (a) intrinsic noise of the original data, (b) compression error introduced by SAMVQ at compression ratio 20:1, and (c) overall noise (i.e. intrinsic noise + compression error) at location (49,174).

sent to a hyperspectral data user for use. Figure 2 (c) shows the overall noise profile at the same location (49, 174). It was obtained by subtracting the spectrum of the reconstructed datacube from the spectrum of the noise free datacube at the same location. Interestedly, the overall noise profile shows that the maximum value of the noise is reduced to 77 DN at band 63, and the minimum value of the noise is reduced -111 DN at band 71. The overall noise value for bands between 100 and 204 is reduced to between –5 DN and 5 DN rather than between –15 DN and 15 DN as in the original datacube. The range of the overall noise in VNIR bands is also smaller than in the intrinsic noise of original datacube.

In order to assess the compression error of the entire reconstructed datacube, the standard deviation of the compression error datacube at each band was calculated and plotted as a function of spectral band number. These standard deviations were used to estimate the noise level of the compressed data. The standard deviations of the intrinsic noise and of the overall noise at each band were also calculated for the purpose of comparison. They are shown in Figure 3. It is observed that the standard deviations of compression error images (solid line) are much smaller than those of intrinsic noise images (doted line) for the bands with high magnitude noise (bands between 35 and 75). For the rest of the bands they are very close. It is also observed that the standard deviations of the overall noise images that include both the intrinsic noise and the compression error (thick broken line) are smaller than those of intrinsic noise images (doted line) for all the bands. This observation indicates that the overall noise level of the compressed datacube is even lower than the noise level of the original datacube. This is probably due to the random error (or noise) introduced by the compression algorithm which cancelled the intrinsic noise in the original data. This result shows that the vector quantization based compression algorithm evaluated here can act as a low-pass filter, suppressing the high frequency noise during the compression [3].

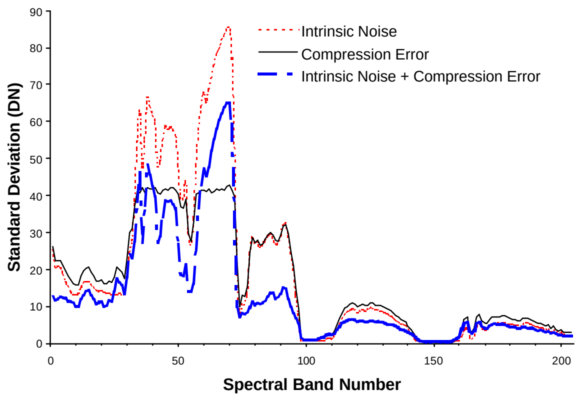


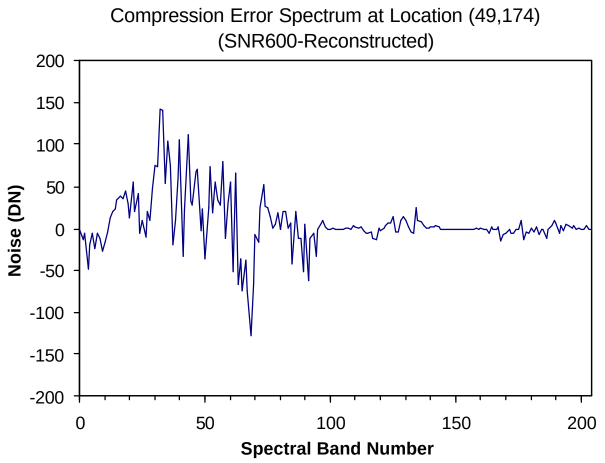
Figure 3 Standard deviations of single band images for intrinsic noise, compression error (using SAMVQ at compression ratio 20:1) and overall noise (intrinsic noise + compression error).

4 Near Lossless Feature of HSOCVQ

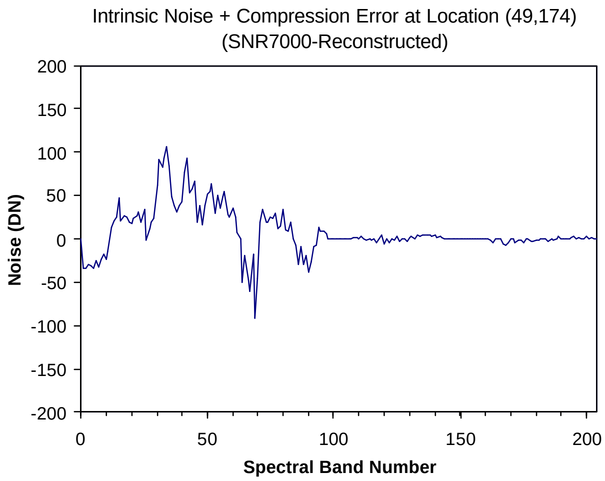
The original datacube was compressed using HSOCVQ algorithm at compression ratio 10:1. The compression error and the overall noise profiles of a single spatial sample are shown in Figure 4. It can be seen that the compression error introduced by the HSOCVQ is smaller than the intrinsic noise of the original datacube across the spectral bands. The maximum value of the compression error is 143 DN at band 32, and the minimum value of the compression error is -128 DN at band 68. The error value for bands between 100 and 204 is between -15 DN and 15 DN, the same level as for the intrinsic noise.

The overall noise profile of the reconstructed datacube at location (49, 174), which takes into account both the intrinsic noise and the compression error, shows that the maximum value of the noise is reduced to 107 DN at band 34, and the minimum value of the noise is reduced -92 DN at band 69. The noise range for bands between 100 and 204 is reduced to between –5 DN and 5 DN rather than between –15 DN and 15 DN as in the original datacube.

Figure 5 shows standard deviations of band images for compression error datacube (solid line) and for the overall noise datacube that includes both the intrinsic noise and the compression error (thick broken line) as a function of spectral band number. The standard deviations of band images for compression error (solid line) are about 5 to 10 DN larger than those for intrinsic noise (doted line) for most bands, but they are close to or smaller for the bands with high magnitude noise (bands between 35 and 75). The standard deviations of band images for the overall noise (thick broken line) are smaller than those for intrinsic noise (doted line) for the bands between 35 and 105.



(a)



(b)

Figure 4 Profiles of (a) compression error introduced by HSOCVQ at ratio 10:1, and (b) overall noise (i.e. intrinsic noise + compression error) at location (49,174).

5 Conclusion

In summary, the experimental results show that the compression errors introduced by SAMVQ and HSOCVQ are similar to the level of intrinsic noise caused by the instrument noise and other noise sources contained in the original datacube. The compression errors are smaller than the intrinsic noise in bands with high magnitude noise.

The reconstructed data after compression contains both the intrinsic noise of the original datacube and the compression error. They are the final noise budget of the datacube, if the reconstructed data is sent to a hyperspectral data user for derivation of his applications. The noise contained in the reconstructed data is the overall noise. The overall noise of the compressed datacube is even smaller than the intrinsic noise for all the bands when the data is compressed using SAMVQ and for most of the bands when the data is compressed using HSOCVQ. This is because the compression algorithms act like a low-pass filter, suppressing the high frequency noise during the compression.

The experimental results justify the claim that SAMVQ and HSOCVQ algorithms are near lossless for remote sensing applications, compared to the intrinsic noise of the original datacube caused by the instrument noise and other noise sources. Statistically, the SAMVQ algorithm shows better near lossless performance than the HSOCVQ algorithm.

An additive noise model was used to simulate the intrinsic noise in this paper. Other noise modes such as multiplicative or shot noise will be investigated in a future evaluation study.

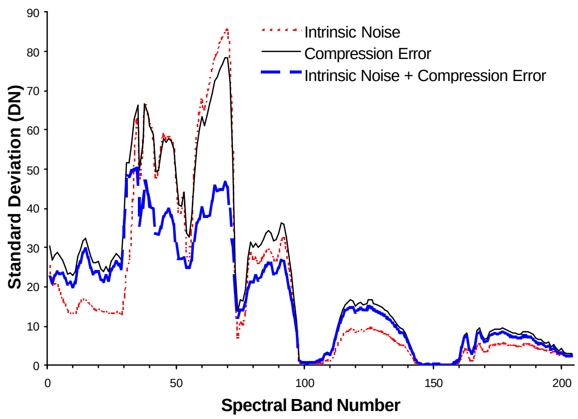


Figure 5 Standard deviations of single band images for intrinsic noise, compression error (using HSOCVQ at compression ratio 10:1) and overall noise (i.e. intrinsic noise + compression error).

Acknowledgement

The authors would like to thank Dr. David Goodenough at Pacific Forestry Centre, Natural Resources Canada for providing the AVIRIS Greater Victoria Watershed District data set, and Dr. David Mates at MacDonald, Dettwiler and Associates, Canada for generating the noise-added datacube.

Reference:

[1] S.-E. Qian, A.B. Hollinger, D. Williams and D. Manak, Fast 3D data compression of hyperspectral imagery using vector quantization with spectral-feature-based binary coding, *Journal of Optical Engineering*, Vol. **35**, 1996, pp.3242-3249.

[2] S.-E. Qian, A.B. Hollinger, D. Williams and D. Manak, A near lossless 3-dimensional data compression system for hyperspectral imagery using correlation vector quantization, *Proceedings of 47th International Astronautical Congress*, Beijing, China, 1996.

[3] S.-E. Qian, A.B. Hollinger, D. Williams and D. Manak, 3D data compression system based on vector quantization for reducing the data rate of hyperspectral imagery, In *Applications of Photonic Technology 2*, (New York: Plenum Press), 1997, pp. 641-654.

[4] S.-E. Qian, A.B. Hollinger, D. Williams and D. Manak, 3D data compression of hyperspectral imagery using vector quantization with NDVI-based multiple codebooks, *Proceedings IGARSS'98, IEEE International Geoscience and Remote Sensing Symposium*, Seattle, Washington, USA, Vol. **3**, 1998, pp.2680-2684.

[5] D. Manak, S.-E. Qian, A.B. Hollinger and D. Williams, Efficient Hyperspectral Data Compression using vector Quantization and Scene Segmentation, *Canadian Journal of Remote Sensing*, Vol. **24**, 1998, pp.133-143.

[6] S.-E. Qian, A.B. Hollinger and Y. Hamiaux, Study of real-time lossless data compression for hyperspectral imagery, *Proceedings of IGARSS'99, IEEE Int. Geoscience and Remote Sensing Symposium*, Hamburg, Germany, 1999.

[7] S.-E. Qian, A.B. Hollinger, D. Williams and D. Manak, Vector quantization using spectral index based multiple sub-codebooks for hyperspectral data compression, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. **38, No. 3**, 2000, pp.1183-1190.

[8] S.-E. Qian, A. B. Hollinger, M. Dutkiewicz, H. Tsang and H. Zwick, Fast Spectral Processing of Hyperspectral Image Cubes, *14th International Conference and Workshop on Applied Geologic Remote Sensing*, Las Vegas, USA, 2000.

[9] S.-E. Qian and A.B. Hollinger, Applications of wavelet data compression using modified zerotrees in remotely sensed data, *Proc. of IGARSS'2000, IEEE Int. Geoscience and Remote Sensing Symp.*, Honolulu, Hawaii, USA, 2000.

[10] S.-E. Qian and A.B. Hollinger, System and method for encoding/decoding multi-dimensional data using Successive Approximation Multi-stage Vector Quantization (SAMVQ), U. S. Patent No. 6,701,021 B1, awarded on March 2, 2004.

[11] S.-E. Qian and A.B. Hollinger, Method and System for Compressing a Continuous Data Flow in Real-Time Using Cluster Successive Approximation Multi-stage Vector Quantization (SAMVQ)", U.S. Patent Application No. 10/611,897 filed on July 3, 2003.

[12] S.-E. Qian and A.B. Hollinger, System and method for encoding multi-dimensional data using Hierarchical Self-Organizing Cluster Vector Quantization (HSOCVQ)", U. S. Patent No. 6,724,940 B1, awarded on April 20, 2004.

[13] S.-E. Qian and A.B. Hollinger, Method and System for Compressing a Continuous Data Flow in Real-Time Using Recursive Hierarchical Self-Organizing Cluster Vector Quantization (HSOCVQ)", U.S. Patent Application No. 10/606,761 filed on June 27, 2003.

[14] S.-E. Qian, A.B. Hollinger, M. Bergeron, I.Cunningham, C. Nadeau, G. Jolly and H. Zwick, A Multi-disciplinary User Acceptability Study of Hyperspectral Data Compressed Using Onboard Near Lossless Vector Quantization Algorithm, Submitted to *International Journal of Remote Sensing*, in Feb. 2004.