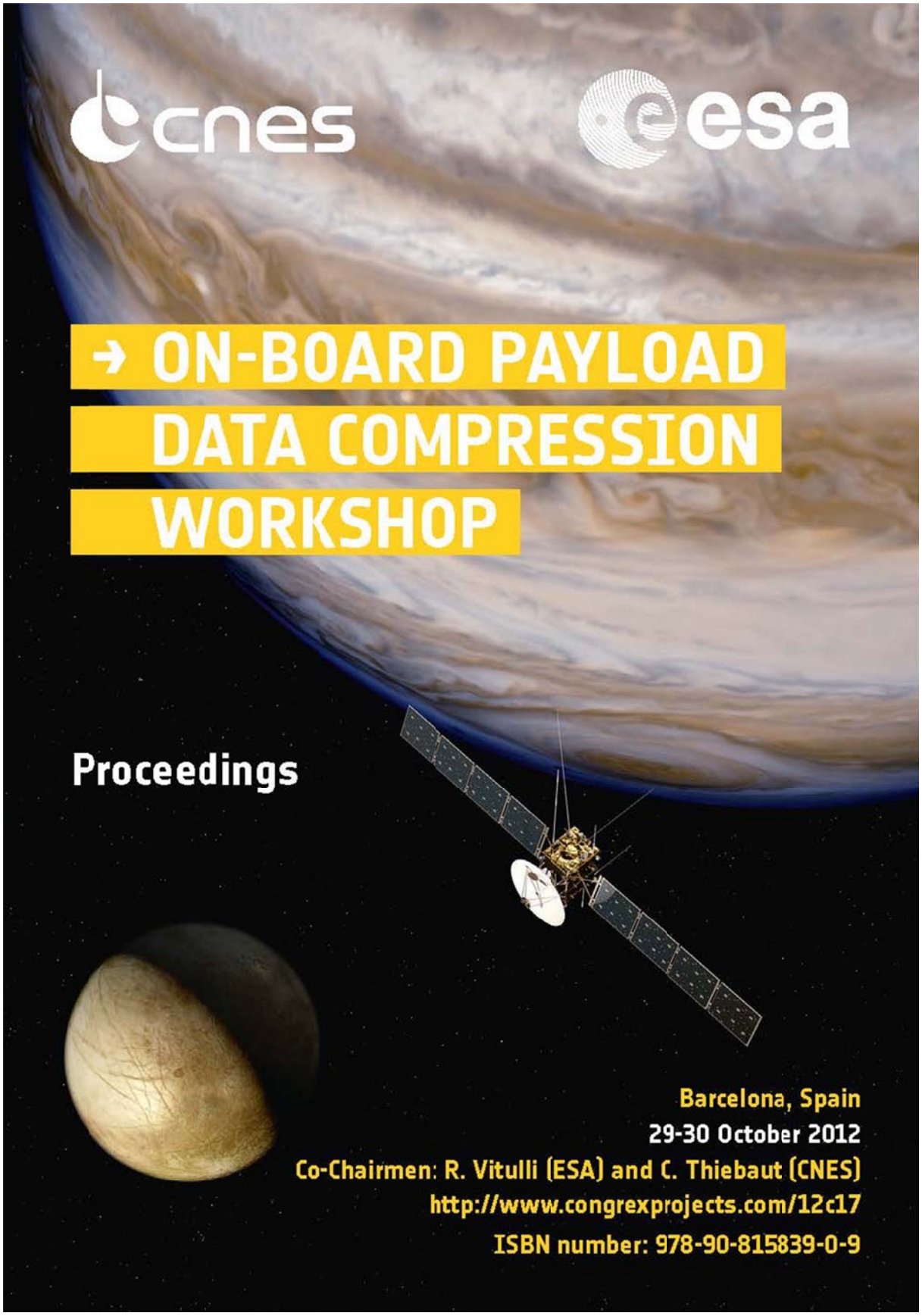
cnes
esa
→ ON-BOARD PAYLOAD
DATA COMPRESSION
WORKSHOP
Proceedings
Barcelona, Spain
29-30 October 2012
Co-Chairmen: R. Vitulli (ESA) and C. Thiebaut (CNES)
http://www.congrexprojects.com/12c17
ISBN number: 978-90-815839-0-9

INFORMATION-THEORETIC ASSESSMENT OF LOSSY AND NEAR-LOSSLESS ON-BOARD HYPERSPECTRAL DATA COMPRESSION

Bruno Aiazzi[(1)], **Luciano Alparone**[(1,2)], **Stefano Baronti**[(1)], **Leonardo Santurri**[(1)]

[(1)] *IFAC–CNR: Institute of Applied Physics "Nello Carrara", Area della Ricerca di Firenze*
10 Via Madonna Del Piano, 50019 Sesto F.no (FI) (Italy), {B.Aiazzi S.Baronti }@ifac.cnr.it
[(2)] *DET-UniFI: Department of Electronics and Telecommunications, University of Florence*
3 Via Santa Marta, 50139 Firenze (Italy), Alparone@lci.det.unifi.it

ABSTRACT

This paper proposes a Rate-Distortion model to measure the impact of lossy and near-lossless compression of raw data on the information conveyed once the lossy/near-lossless decompressed raw data have been converted to radiance units. Input variables of the model are the original uncompressed raw data and their measured noise variances, according to a mixed photon + electronic noise model. Band-scaling gains and destriping coefficients for calibration, or equivalently radiance data obtained from the raw data, are also assumed to be available. The model makes use of advanced lossless/near lossless methods achieving the ultimate compression, regardless of their computational complexity. They are not to be implemented on board, but are used to measure the entropy of the data. Preliminary experiments on AVIRIS 2006 Yellowstone sequences show the trend of spectral information, i.e. information pertaining the ideal noise-free radiance source, without acquisition noise, versus either bit-rate/compression ratio or MAD/MSE distortion.

INTRODUCTION

Technological advances in imaging spectrometry have lead to acquisition of data that exhibit extremely high spatial, spectral, and radiometric resolution. In particular, the increment in spectral resolution has motivated the extension of vector signal/image processing techniques to hyperspectral data, for both data analysis and compression [1]. As a matter of fact, a challenge of satellite hyperspectral imaging is data compression for dissemination to users and especially for transmission to ground station from the orbiting platform. Data compression generally performs a decorrelation of the correlated information source, before entropy coding is carried out. To meet the quality issues of hyperspectral image analysis, differential pulse code modulation (DPCM) is usually employed for lossless/near-lossless compression, i.e., the decompressed data have a user-defined maximum absolute error, being zero in the lossless case, nonzero otherwise. DPCM basically consists of a prediction followed by entropy coding of quantized differences between original and predicted values. A unity quantization step size allows reversible compression as a limit case. Several variants exist in DPCM prediction schemes, the most sophisticated being adaptive [2, 3, 4, 5, 6].

When the hyperspectral imaging instrument is placed on a satellite, data compression is crucial. To meet the quality issues of hyperspectral imaging, differential pulse code modulation (DPCM) is usually employed for either lossless or near lossless compression. The latter indicates that the decompressed data have a user-defined maximum absolute error, being zero in the lossless case. Several variants exist in prediction schemes, the most performing being adaptive [7, 8]. Lossless compression thoroughly preserves the information of the data but allows a moderate decrement in transmission rate to be achieved [9]. The bottleneck of downlink to ground stations may hamper the coverage capabilities of modern satellite instruments. If strictly lossless techniques are not used, a certain amount of information of the data will be lost. However, such a statistical information may be mostly due to random fluctuations of the instrumental noise. The rationale that compression-induced distortion is less harmful, in those bands, in which the noise is higher constitutes the *virtually lossless* paradigm [10].

This paper faces the problem of quantifying the trade-off between compression and decrement in the spectral information content. A rate distortion model is used to quantify the compression of the noisy version of an ideal information source, analogously to audio/video transcoding from different formats. Experiments carried out on AVIRIS 2006 raw data show that the proposed approach is practically feasible and yields reasonably agreeable results.

SATELLITE HYPERSPECTRAL REMOTE SENSING

Since 2001, the pioneering mission Hyperion has opened new possibilities of global Earth coverage, hyperspectral imaging from satellite has progressively grown in interest up to motivate the upcoming missions EnMAP and PRISMA.

The hyperspectral processing chain consists of three segments: satellite segment, ground segment and user segment. The on-board instruments produces data in raw format. Raw data are digital count from the analog-to-digital converter (ADC) diminished by the dark signal that has been averaged in time immediately before the acquisition. Raw data are compressed, with or without loss, and downloaded to the ground station(s), where the data are decompressed, converted to radiance values and corrected for instrumental effects (e.g., striping of push-broom sensors). The calibrated data are geometrically corrected for orbital effects, georeferenced and possibly orthorectified. All geometric operations subsequent to calibration have little impact on the quality of data products. Eventually, data products are stored in archives, generally with highly redundant formats, e.g., double precision floating point per pixel radiance value, with spectral radiance measured in $W \cdot sr^{-1} \cdot m^{-2} \cdot nm^{-1}$.

When the data are distributed to users, they are usually converted to fixed-point formats (e.g., 16-bit per component, including a sign bit). This conversion may lead to a loss of information, especially because the radiance unit in the fixed point format is $\mu W \cdot sr^{-1} \cdot m^{-2} \cdot nm^{-1}$. A finer radiance step would be 10 times smaller and would require 20 bits instead of 16. Fixed point radiance data are compressed, possibly with loss, and delivered to users. After decompression of radiance data, in a typical user application, solar irradiance and atmospheric transmittance are corrected by user to produce reflectance spectra that may be matched to library spectra in order to recognize and classify materials. At this step, it may be important to investigate the effects of a lossy compression of radiance data in terms of changes in spectral angle with respect to reflectance spectra obtained from uncompressed radiance data [11]. Another approach is to investigate the loss in spectral information due to an irreversible compression. Such a study is complicated by the fact that the available data are a noisy realization of the spectral information source, which is assumed to be noise-free.

SIGNAL-DEPENDENT NOISE MODELING AND ESTIMATION

A generalized signal-dependent noise model has been proposed to deal with several different acquisition systems. Many types of noise can be described by using the following parametric model [12]

$$\begin{aligned} g(m,n) &= f(m,n) + f(m,n)^{\gamma} \cdot u(m,n) + w(m,n) \\ &= f(m,n) + v(m,n) + w(m,n) \end{aligned} \tag{1}$$

where (m,n) is the pixel location, $g(m,n)$ the observed noisy image, $f(m,n)$ the noise-free image, modeled as a non-stationary correlated random process, $u(m,n)$ a stationary, zero-mean uncorrelated random process independent of $f(m,n)$ with variance σ_u^2, and $w(m,n)$ is electronics noise (zero-mean white and Gaussian, with variance σ_w^2). For a great variety of images, this model has been proven to hold for values of the parameter γ such that $|\gamma| \leq 1$. The additive term $v = f^{\gamma} \cdot u$ is the *generalized signal-dependent* (GSD) noise. Since f is generally non-stationary, the noise v will be non-stationary as well. The term w is the signal-independent noise component and is generally assumed to be Gaussian distributed.

The model (1) applies also to images produced by optoelectronic devices, such as CCD cameras, multispectral scanners and imaging spectrometers. In that case the exponent γ is equal to 0.5. The term $\sqrt{f}u$ stems from the Poisson-distributed number of photons captured by each pixel and is therefore denoted as *photon* noise [13].

Let us rewrite the model (1) with $\gamma = 0.5$:

$$g(m,n) = f(m,n) + \sqrt{f(m,n)} \cdot u(m,n) + w(m,n). \tag{2}$$

Eq. (2) represents the electrical signal resulting from the photon conversion and from the dark current. The mean dark current has been preliminarily subtracted to yield $g(m,n)$. However, its statistical fluctuations around the mean constitute most of the zero-mean electronic noise $w(m,n)$. The term $\sqrt{f(m,n)} \cdot u(m,n)$ is the photon noise, whose mean is zero and whose variance is proportional to $E[f(m,n)]$. It represents a statistical fluctuation of the photon signal around its noise-free, $f(m,n)$, due to the granularity of photons originating electric charge.

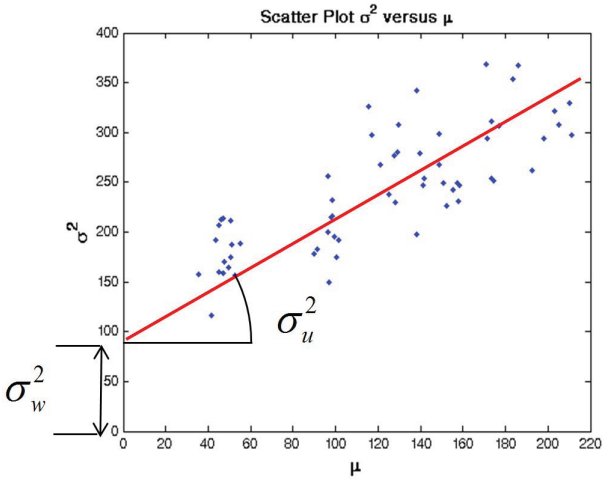

Figure 1: Calculation of slope and intercept of mixed photon/electronic noise model, with regression line superimposed.

If the variance of (2) is calculated on homogeneous pixels, in which $\sigma_f^2(m,n) = 0$, by definition, thanks to the independence of f, u and w and the fact that both u and w have null mean and are stationary, we can write:

$$\sigma_g^2(m,n) = \sigma_u^2 \cdot \mu_f(m,n) + \sigma_w^2 \tag{3}$$

in which $\mu_f(m,n) \triangleq E[f(m,n)]$ is the nonstationary mean of f. The term $\mu_f(m,n)$ equals $\mu_g(m,n)$, from (2).

Eq. (3) represents a straight line in the plane $(x,y) = (\mu_f, \sigma_g^2) = (\mu_g, \sigma_g^2)$, whose slope and intercept are equal to σ_u^2 and σ_w^2, respectively. The interpretation of (3) is that on statistically homogeneous pixels the theoretical nonstationary ensemble statistics (mean and variance) of the observed noisy image $g(m,n)$ lie upon a straight line. In practice, homogeneous pixels with $\sigma_f^2(m,n) \equiv 0$ may be extremely rare and theoretical expectation are approximated with local averages. Hence, the most homogeneous pixels in the scene appear in the mean-variance plane to be clustered along the straight line $y = mx + y_0$, in which $m = \sigma_u^2$ and $y_0 = \sigma_w^2$.

INFORMATION-THEORETIC PROBLEM STATEMENT

Let us denote with A the (unknown) source ideally obtained by means of an acquisition with a noiseless device. Quantization of the data produced by the on-board instrument is set on the basis of instrument noise and downlink constraints. For satellite imaging spectrometers it is usually 12 bits per pixel per band (bpppb). Let B denote the noisy acquired source, e.g. quantized with 12 bpppb. The unavaiable source A is assumed to be quantized with 12 bpppb, but its acquisition is noise-free. If the source B is lossy compressed, a new source C is obtained by decompressing the compressed bit stream at the ground station. If C is converted to radiance units, band scaling gains and destriping coefficients are applied for each wavelength and the outcome is quantized to radiance units (typically W/(m2 sr nm)), so that any physically attainable value can be mapped with a 16-bit wordlength (15 bits + one sign bit, because radiances may take small negative values after on-board removal of time averaged dark current from raw data). The radiance source, D, is the result of reversible deterministic operations (calibration and destriping), which produce real valued data, followed by quantization to an integer number of radiance units. The transformation to pass from C to D is assumed to be the same for both reversible and irreversible compression of the raw data. A simplified model of the rate distortion chain that does not include losses introduced by conversion to radiance of decompressed raw data is displayed in Fig. 2.

The problem can be stated in terms of Rate-Distortion Theory. The entropy of A, $H(A)$, is unknown, but the entropy of B, $H(B)$ may be estimated by compressing the raw data without any loss by means of an advanced DPCM compressor. Once the noise of B has been measured, the irrelevance $H(B|A)$ can be estimated from the noise model and measurements. Hence, the mutual information between A and B, $I(A;B)$, which is the amount of the entropy of A that is contained in B, is calculated as $I(A;B) = I(B;A) = H(B) - H(B|A)$. $I(A;B)$ is the mean amount of useful "spectral" information coming from A that will be contained in B and, if compression is strictly lossless, also in C, because in that case $C = B$. The lossless case is useful to calibrate the model. The ratio between the cubes of lossless radiance (D) and raw data (B) constitutes a kind of transfer function, useful to pass from raw to radiance data.

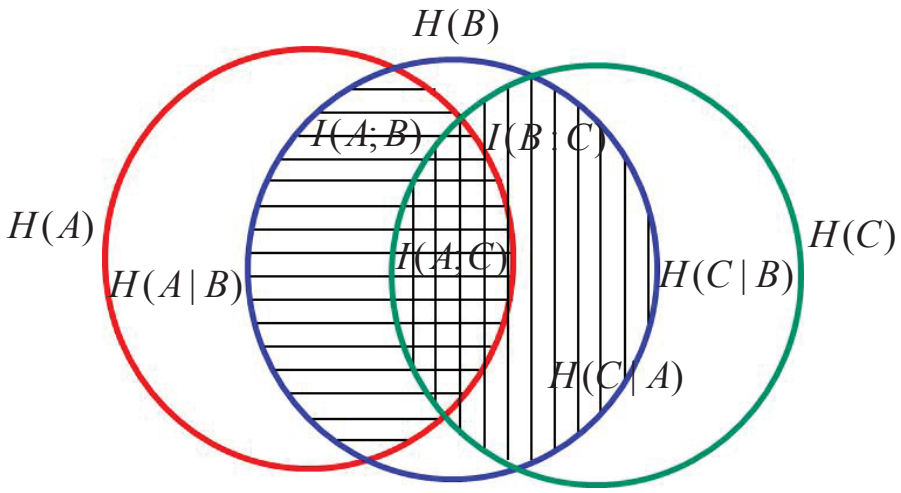

Figure 2: Rate-distortion chain for the information-theoretic assessment of the lossy compression of the noisy version of an information source.

All the coefficients in the transfer data-cube are expected to be greater than one, in order to preserve the information of the integer valued raw data in the integer valued radiance data obtained after multiplication by the transfer function and roundoff to integer.

Now, let us consider the lossy case, i.e. $C \neq B$. The lossy compression bit rate achieved by an optimized coder will approximate the mutual information between B and C, $I(B;C)$, It is note worth that, regardless of the entropy of the radiance data (D) obtained from lossy compressed raw data, (C), the mutual information $I(B;D)$ will be upper bounded by $I(B;C)$. In summary. $I(A;B) \geq I(A;C) \geq I(A;D)$. The term $I(A;D)$ may be estimated from the noise model and measurements of B and from the model of compression noise, assumed to be Gaussian for lossy compression or uniformly distributed for near-lossless compression. Eventually $H(A)$, whenever of interest, can be estimated by following the procedure described in [3, 14].

EXPERIMENTAL SETUP

The operational steps implementing the proposed rate-distortion model for quality assessment of on-board near-lossless compression are the following:

1. take a raw (uncalibrated) hyperspectral sequence;
2. measure its noise variance (both signal independent and signal dependent terms) for each spectral band;
3. extract the spatial noise pattern of each band, e.g. by taking the difference between original and denoised band;
4. compress the sequence with desired loss, e.g. MAD, by means of an optimized coder. Incidentally the bit-rate produced by such a coder is an estimate of $I(B;C)$;
5. decompress the compressed sequence and recompress the outcome without loss, to obtain $H(C)$;
6. take the difference between original and decompressed band, i.e. the pattern of compression-induced pixel error values;
7. add the two patterns of instrumental noise and of compression errors to yield the overall error map;
8. calculate the entropy of the overall error map as an estimate of $H(C|A)$;
9. calculate $I(A;C)$ as $I(A;C) = H(C)H(C|A)$.

COMPUTER SIMULATIONS

A set of calibrated and raw images acquired in 2006 by the *Airborne Visible InfraRed Imaging Spectrometer* (AVIRIS) has been provided by NASA/JPL to Consultative Committee for Space Data Systems (CCSDS) and is available for compression experiments. This data set consists of five 16-bit calibrated images and the corresponding 16-bit raw images acquired over Yellowstone, WY. Each image is composed by 224 bands and each scene (the scene numbers are 0, 3, 10, 11, 18) has 512 lines [8]. All data have been clipped to 14 bits (a negligible percentage of outliers has been affected by clipping) and remapped to 12 bits to simulate a space-borne instrument, e.g. Hyperion.

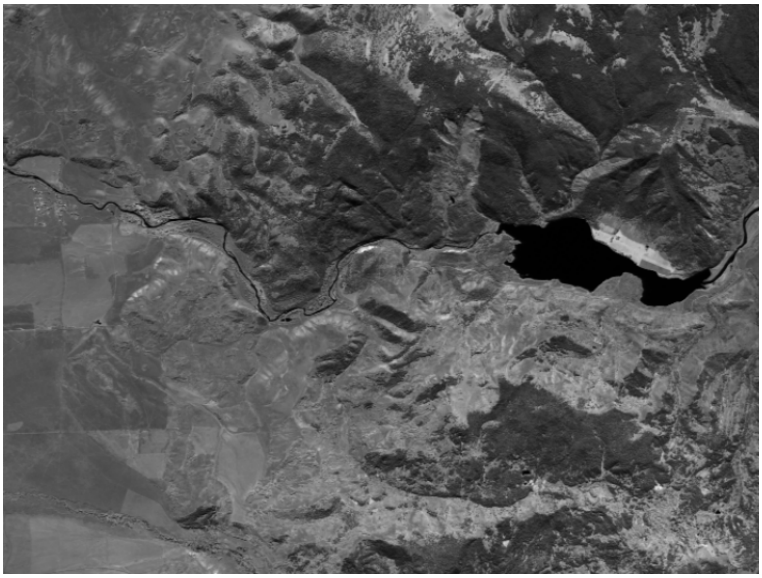

Figure 3: Sample NIR band of AVIRIS 2006 Yellowstone scene 0.

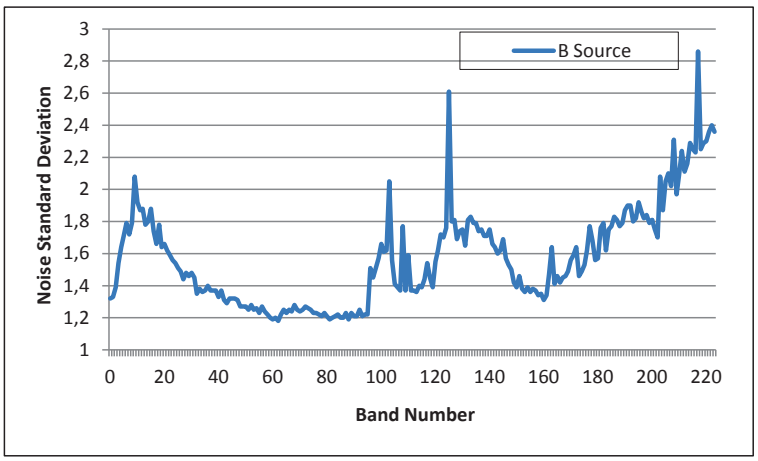

Figure 4: Noise standard deviation (thermal/electronic component only) measured on AVIRIS 2006 Yellowstone scene 10.

The noise standard deviation has been measured for Yellowstone 10 by using the algorithm described in [15] and is reported in Fig. 4. Apart from marked increments at the edges of the spectral interval due to loss of sensibility of instruments, the noise standard deviation is approximately constant and always greater than 1.2. The average is approximately 1.7. However, only the electronic, or dark, component of the noise is reliably estimated. The photon component is undetermined because bright areas do not produce clusters in the variance-to-mean scatterplot. The photon noise variance is expected to be lower than the electronic one for whisk-broom instruments, but not for push-broom ones, at least in V-NIR wavelengths.

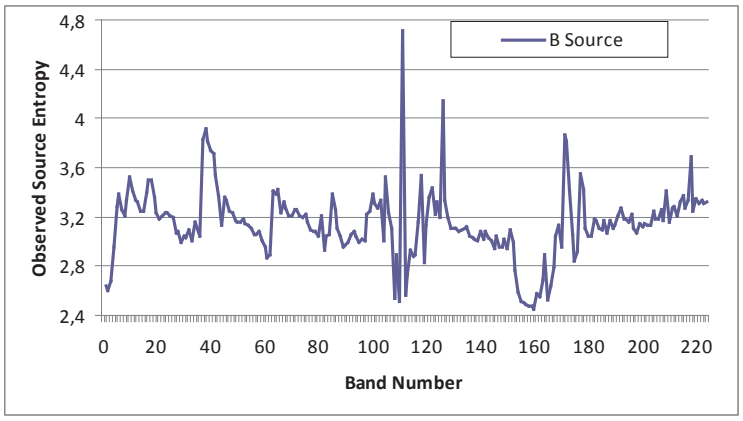

Figure 5: Bit-rate in bits per pixel per band (bpppb) of AVIRIS 2006 Yellowstone scene 10 produced by the S-RLP codec.

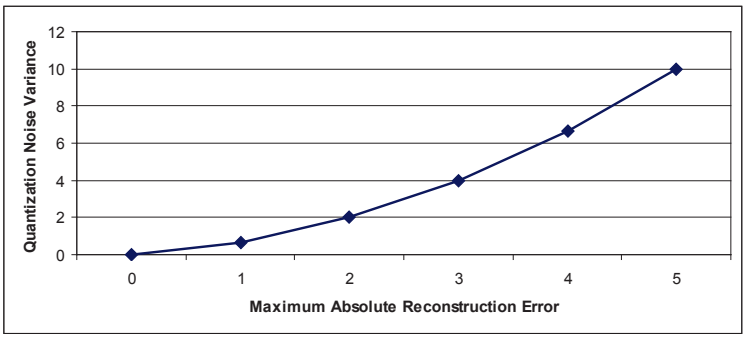

Figure 6: Average quadratic distortion due to quantization of prediction errors as a function of the integer valued maximum allowed absolute error for a DPCM coder.

Fig. 5 reports lossless compression bit rates of AVIRIS 2006 Yellowstone raw scene 10 (clipped and remapped to 12 bpppb). The compression algorithm is S-RLP [7], which is a MAD-bounded, i.e., near lossless algorithm, providing the ultimate compression attainable for hyperspectral data. In a possible on-board implementation, MA-DPCM [16], a simplified version of S-RLP, is preferable. However, S-RLP is used only to measure the entropy of the various sources and is not required for a practical implementation.

Let ϵ denote the Maximum Absolute Reconstruction Error, a.k.a. Maximum Absolute Deviation (MAD), a.k.a. peak error (PE). The maximum quantization step size Δ yielding MAD equal to ϵ is $\Delta = 2\epsilon + 1$ and a distortion $D = (\Delta^2 - 1)/12 = ((2\epsilon + 1)^2 - 1)/12$. The term $-1/12$ occurs when integers are requantized. The trend is visible in Fig. 6.

Fig. 7 reports information parameters for band n. 20 and the whole sequence of Yellowstone 10. Band 20 has been chosen as representative in terms of compression. Its bit-rate in inter-band mode is almost identical to the bit-rate of the whole sequence. Compression ratios (CR) are relative to the uncompressed word-length of 12 bits.

In the plot of information parameters relative to band 20, when the error is equal to zero, $C = B$, hence $I(B;C) = I(B;B) = H(B)$. $I(B;C)$ is monotonically decreasing with the mean distortion D, or better with the quantization step size $\Delta = (2\epsilon * 1)$, approximately as $I(B;C) = H(B)\log_2(\Delta)$. $H(C)$ depends on how much C has been *smoothed* by the lossy compression of B, in other words on how the compression algorithm works. The term $H(C|A)$ represents the entropy of the sum of the two events, both independent of A, that are acquisition noise and compression errors, which are not necessarily independent of one another. The trend of $H(C|A)$ vs. peak error reveals that compression errors, at least if they are not too small, tend to cancel acquisition noise. The plots of $I(A;C) = H(C)H(C|A)$ exhibit an apparent incongruence for errors equal to 3, 4 and 5, since the amount of information from A is higher than the coding bit rate. Such incongruence is mostly due to inaccurate noise measurements (the photon component can not be reliably measured).

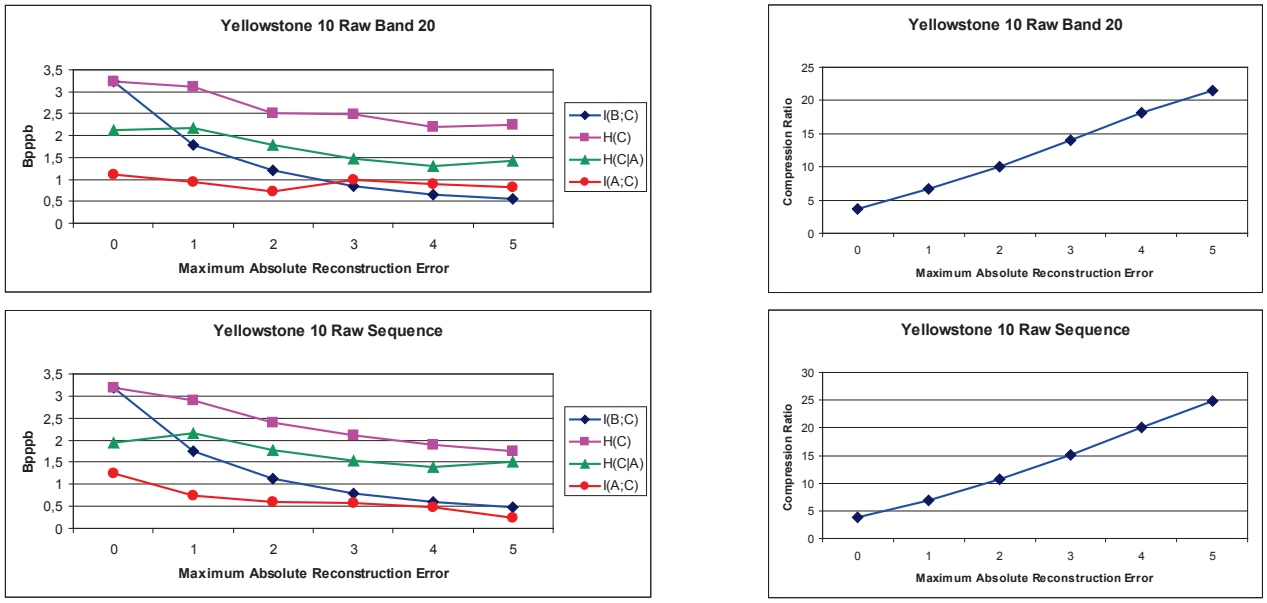

Figure 7: Estimated entropy and mutual information parameters for AVIRIS 2006 Yellowstone 10, band 20 and whole sequence.

Compression ratios are 10% higher for the whole sequence than for band 20. The CR vs MAD plots refer to S-RLP, but can be replaced by any other (DPCM) coder, possibly suitable for space environment, which will be likely produce lower CRs. On the whole sequence, the average $I(A;C)$ has been corrected by discarding all absorption bands, whose $I(A;C)$ turns out to be negative. The trend of $I(C|A)$ vs. MAD is now monotonic and reveals that compression errors, at least if they are not too small, tend to cancel acquisition noise, thereby resulting in an almost flat trend in [2,4]. When the $MAD = 1$, CR almost doubles with respect to lossless compression, but the *useful* information is approximately halved. Therefore, lossy compression might not be recommended.

DISCUSSION

The analysis reported suffers from several shortcomings presented in the following sorted by decreasing relevance:

1. photon noise is disregarded because hard to estimate;
2. the cross-covariance between acquisition noise and compression errors is calculated only in the subset of pixels in which the noise was estimated (lake);
3. the entropy of the noise components has been replaced with differential entropy, which allows a correction for (auto)correlated noise to be applied (the noise σ^2 has been multiplied by $(1 - \rho^2)$ and hence decremented because the correlation coefficient ρ ranges in $[-1, 1]$. CC is measured via compression.

Nevertheless it seems that the above analysis is capable of capturing the essence of the problem and, once the above shortcomings have been fixed, the proposed model is expected to be suitable for investigating the trade-off between attainable compression and quality of decompressed data.

CONCLUDING REMARKS

This study has proposed a rate-distortion model to quantify the loss of useful spectral information that gets lost after near-lossless compression, and its experimental setup, such that the information-theoretical model may become a computational model. Key ingredient of this recipe is an advanced hyperspectral image coder providing the ultimate compression regardless of time and computation constraints (such a coder is not necessary to perform on-board compression, but only for simulations). Noise estimation is crucial because of the extremely low noisiness of hyperspectral data, that are also

richly textured. Noise filtering is also required to extract noise patterns. The drawback is that denoising filters generally work properly if they know reasonably exact values of the parameters of the noise model they have been designed for. Preliminary results on AVIRIS 2006 Yellowstone 10 sequence are encouraging, although absorption/anomalous bands are likely to contradict the main assumptions underlying the proposed model and should not be considered.

ACKNOWLEDGEMENT

The authors are grateful to CCDS for providing the AVIRIS raw data (http://compression.jpl.nasa.gov/hyperspectral).

References

[1] D. A. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] B. Aiazzi, P. Alba, L. Alparone, and S. Baronti, "Lossless compression of multi/hyper-spectral imagery based on a 3-D fuzzy prediction," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 5, pp. 2287–2294, Sep. 1999.

[3] B. Aiazzi, L. Alparone, and S. Baronti, "Near-lossless compression of 3-D optical data," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 11, pp. 2547–2557, Nov. 2001.

[4] J. Mielikainen and P. Toivanen, "Clustered DPCM for the lossless compression of hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 12, pp. 2943–2946, Dec. 2003.

[5] E. Magli, G. Olmo, and E. Quacchio, "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Geosci. Remote Sensing Lett.*, vol. 1, no. 1, pp. 21–25, Jan. 2004.

[6] F. Rizzo, B. Carpentieri, G. Motta, and J. A. Storer, "Low-complexity lossless compression of hyperspectral imagery via linear prediction," *IEEE Signal Processing Lett.*, vol. 12, no. 2, pp. 138–141, Feb. 2005.

[7] B. Aiazzi, L. Alparone, S. Baronti, and C. Lastri, "Crisp and fuzzy adaptive spectral predictions for lossless and near-lossless compression of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 532–536, Oct. 2007.

[8] A. B. Kiely and M. A. Klimesh, "Exploiting calibration-induced artifacts in lossless compression of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 8, pp. 2672–2678, Aug. 2009.

[9] B. Aiazzi, L. Alparone, and S. Baronti, "Lossless compression of hyperspectral images using multiband lookup tables," *IEEE Signal Processing Lett.*, vol. 16, no. 6, pp. 481–484, June 2009.

[10] C. Lastri, B. Aiazzi, L. Alparone, and S. Baronti, "Virtually lossless compression of astrophysical images," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 15, pp. 2521–2535, 2005.

[11] B. Aiazzi, L. Alparone, S. Baronti, C. Lastri, and M. Selva, "Spectral distortion in lossy compression of hyperspectral data," *Journal of Electrical and Computer Engineering*, vol. 2012, no. Article ID 850637, pp. 8.

[12] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.

[13] J. L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, New York, 1998.

[14] B. Aiazzi, , S. Baronti, L. Santurri, M. Selva, and L. Alparone, "Information-theoretic assessment of multidimensional signals," *Signal Processing*, vol. 85, no. 5, pp. 903–916, May 2005.

[15] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, P. Marcoionni, I. Pippi, and M. Selva, "Noise modelling and estimation of hyperspectral data from airborne imaging spectrometers," *Annals of Geophysics*, vol. 41, no. 1, pp. 1–9, Feb. 2006.

[16] B. Aiazzi, L. Alparone, and S. Baronti, "On-board DPCM compression of hyperspectral data," in *Proc. ESA OBPDC 2010, 2nd International Workshop on On-Board Payload Data Compression*, Vitulli R, Ed., 2010.