

QUALITY AND INFORMATION CONTENT OF CHRIS HYPER-SPECTRAL DATA

B. Aiazzi, S. Baronti, P. Marcoionni, I. Pippi, and M. Selva

Inst. of Applied Physics "Nello Carrara", IFAC-CNR, Via Panciatichi, 64, 50127 Florence, Italy

ABSTRACT

The work focuses on evaluating quality and estimating information of CHRIS hyper-spectral images. Quality is assessed through the characterisation of the noise while information is estimated by means of an operative definition according to which the information content of a data set is given by the amount of information that cannot be predicted from the data that have already been acquired and, thus, by the entropy of the prediction errors. The noise model is first verified and the parameters of the model are then estimated. Afterwards lossless data compression is exploited to measure the entropy of the prediction errors through their bit-rate. The information content of the data is estimated by taking into account that the bit-rate achieved by the reversible compression process is due to both the contribution of the noise, whose relevance is null to a user, and of the hypothetically noise-free data. Since our goal is to estimate the amount of information of the ideal noise-free data, an entropy-variance model is assumed for the ideal image. Once all the parameters of the model have been estimated, the entropy of the noise-free source is derived. Results are reported and discussed for hyper-spectral data sets acquired by CHRIS spectrometer. Information assessment is first assessed before and after the radiometric correction process in order to evaluate any effect introduced in the processed data. Then different areas of the same image are processed in order to assess the noise model. Eventually, the procedure is utilised to characterise data sets that have been acquired in different times in order to verify and assess any potential operational change occurred in the instrument set-up or in the processing chain.

Key words: CHRIS hyperspectral data; Correlation analysis; Generalised Gaussian PDF; Information theoretic assessment; Lossless compression; Noise estimation; Noise modelling; Parametric entropy modelling.

1. INTRODUCTION

Information-theoretic assessment is a branch of image analysis aimed at defining and measuring the quality of

digital images (Ref. 1). In the field of optical remote sensing, the fidelity of the data produced by the sensor to the underlying radiance field is another concern (Ref. 2), which can be related to the concept of quality as well. Although this term may be intended as the capability to fulfil the user's expectations and thus cannot be defined in a universal context, being application dependent, quality may be related to the information content of the data. Hence, it can be defined in an objective manner jointly from the signal to noise ratio (SNR) and from the entropy of the digitised images (Ref. 3). On the other hand, the availability of an objective quality measurement may be helpful also for application tasks, especially when hyperspectral images are concerned. Due to the huge amount of data, an extraction of features reflecting the main aspects of spectral behaviour may be preferable to a straightforward classification (Ref. 4). A preliminary information-theoretic assessment may suggest which bands are more significant, i.e., potentially capable of conveying an amount of information larger than that of other bands, so as to reduce the volume of the data to be processed without noticeable penalty.

This work is based on a model suitable for quantifying the information content of digitised multi-dimensional signals, more specifically hyperspectral images. Accurate estimates of the entropy of an image source can only be obtained provided that the data are uncorrelated. Hence, data decorrelation must be considered in order to suppress or largely reduce the correlation existing in natural images. Indeed, entropy is a measure of statistical information that is of uncertainty of symbols emitted by a source. Hence, any observation noise introduced by the imaging sensor will result in an increment in entropy, that is accompanied by a decrement of the information content useful in application contexts, according to Shannon's Information Theory (Ref. 5). An estimation of the noise must be preliminarily carried out in order to quantify its contribution to the overall source entropy rate. By assuming an additive noise independent of the signal and spatially stationary, i.e., statistically homogeneous, the noise parameters (variance and correlation coefficients across track, along track and along wavelength) can be estimated on the homogeneous areas of the signal. Once the standard deviation and the correlation coefficients of the noise (CCs) have been measured, the bit rate pro-

duced by the reversible encoder can be utilised to yield an estimate of the true information content of the multi-spectral source, i.e., of the entropy that the source would have if it were noise-free. To this purpose, a model is devised from the rate distortion theory describing how the relationships between entropy and variance of an uncorrelated non-Gaussian source changes when a stationary white Gaussian random process is superimposed. Such a model can be inverted to yield the entropy of the noise-free source from that of the observed source and from the estimated parameters of the noise.

The remainder of this paper is organised as follows. Sect. 2 presents the information-theoretic procedure step by step, starting from the assumed noise model, source decorrelation by DPCM, parametric entropy modelling of memory-less information sources via generalised Gaussian densities. Sect. 3 reports experimental results on several sets of CHRIS hyperspectral images. Concluding remarks are drawn in Sect. 4.

2. INFORMATION ASSESSMENT PROCEDURE

2.1. Noise modelling

This section focuses on modelling the noise affecting digitised observed signal samples. Unlike coherent or systematic disturbances, which may occur in some kind of data, the noise is assumed to be due to a fully stochastic process. Let us assume for the noise an additive signal-independent non-Gaussian model:

$$g(i) = f(i) + n(i) \quad (1)$$

in which $g(i)$ is the recorded noisy signal level at position (i) and $f(i)$ the noise-free signal. Both $g(i)$ and $f(i)$ are regarded as non-stationary non-Gaussian autocorrelated stochastic processes. The term $n(i)$ is a zero-mean process, independent of f , stationary and autocorrelated. Let its variance σ_n^2 and the correlation coefficient (CC) ρ be constant.

Let us assume for the stationary zero-mean noise a first-order Markov model, uniquely defined by the ρ and the σ_n^2

$$n(i) = \rho \cdot n(i-1) + \epsilon_n(i) \quad (2)$$

in which $\epsilon_n(i)$ is an uncorrelated random process having variance

$$\sigma_{\epsilon_n}^2 = \sigma_n^2 \cdot (1 - \rho^2). \quad (3)$$

The variance of (1) can be easily calculated as

$$\sigma_g^2(i) = \sigma_f^2(i) + \sigma_n^2 \quad (4)$$

thanks to the independence between signal and noise components and to the spatial stationarity of the latter. From (2) it stems that the autocorrelation of $n(i)$ is an exponentially decaying function of the correlation coefficient:

$$R_{nn}(m) \triangleq E[n(i)n(i+m)] = \rho^{|m|} \sigma_n^2. \quad (5)$$

The zero-mean additive signal-independent correlated noise model (2) is relatively simple and mathematically tractable. Its accuracy has been validated for 2D and 3D signals produced by incoherent systems (Ref. 6), by measuring the exponential decay of the correlation function (5).

The noise samples $n(i)$ may be estimated on homogeneous signal segments, in which $f(i)$ is constant, by taking the difference between $g(i)$ and its average $\bar{g}(i)$ on a sliding window of length $2m+1$.

Once the CC of the noise, ρ , and the most homogeneous image pixels have been found by means of robust bivariate regression procedures (Ref. 6), the noise samples may be estimated in the following way. If (2) and (5) are utilised to calculate the correlation of the noise affecting g and \bar{g} on a homogeneous window, the estimated noise sample at the i -th position can be written as

$$\hat{n}(i) = \sqrt{\frac{(2m+1)}{(2m+1) - \left(1 + 2\rho \frac{1-\rho^m}{1-\rho}\right)}} \cdot [g(i) - \bar{g}(i)]. \quad (6)$$

Eq. 6 is a generalisation of the corresponding expression reported in (Ref. 7). In fact, when $\rho = 0$, both equations give the same result, if the equation in (Ref. 7) is written to represent 1-D signals.

The resulting set $\{\hat{n}(i)\}$ is made available to find the noise PDF, either empirical (histogram) or parametric, via proper modelling.

2.2. Source de-correlation via DPCM

Differential Pulse Code Modulation (DPCM) is usually employed for *reversible* data compression. DPCM basically consists of a prediction followed by entropy coding of the resulting prediction errors. For sake of clarity, we will develop the analysis for a 1D *fixed* DPCM and will extend its results to the case of 2D and 3D *adaptive* prediction (Ref. 8, 9, 10, 11).

Let $\hat{g}(i)$ denote the prediction at pixel i obtained as a linear regression of the values of P previous pixels:

$$\hat{g}(i) = \sum_{j=1}^P \phi(j) \cdot g(i-j) \quad (7)$$

in which $\{\phi(j), j = 1, \dots, P\}$ are the coefficients of the linear predictor and are constant throughout the image.

By replacing the additive noise model one obtains:

$$\hat{g}(i) = \hat{f}(i) + \sum_{j=1}^P \phi(j) \cdot n(i-j) \quad (8)$$

in which

$$\hat{f}(i) = \sum_{j=1}^P \phi(j) \cdot f(i-j) \quad (9)$$

represents the prediction for the noise-free signal as formulated from its previous samples. Prediction errors of g are

$$e_g(i) \triangleq g(i) - \hat{g}(i) = e_f(i) + n(i) - \sum_{j=1}^P \phi(j) \cdot n(i-j) \quad (10)$$

in which $e_f(i) \triangleq f(i) - \hat{f}(i)$ is the error the predictor would produce starting from noise-free data. Both $e_g(i)$ and $e_f(i)$ are zero-mean processes, uncorrelated, and non-stationary. The zero-mean property stems from an assumption of local first-order stationarity, within the $(P+1)$ -pixel window comprising the current pixel and its prediction support.

Eq. (10) may be written as

$$e_g(i) = e_f(i) + e_n(i) \quad (11)$$

in which

$$e_n(i) \triangleq n(i) - \hat{n}(i) = n(i) - \sum_{j=1}^P \phi(j) \cdot n(i-j) \quad (12)$$

is the error produced when the correlated noise is being predicted. The term $e_n(i)$ is assumed to be zero-mean, stationary and independent of $e_f(i)$, since f and n are assumed to be independent of each other. Thus, the relationship among the variances of the three types of prediction errors becomes

$$\sigma_{e_g}^2(i) = \sigma_{e_f}^2(i) + \sigma_{e_n}^2. \quad (13)$$

From the noise model (2) it is easily noticed that the term $\sigma_{e_n}^2$ is lower bounded by $\sigma_{\epsilon_n}^2$, which means that $\sigma_{e_n}^2 \geq \sigma_n^2 \cdot (1 - \rho^2)$. The optimum MMSE predictor for a first-order Markov model like (2) is $\phi(1) = \rho$ and $\phi(j) = 0$, $j = 2, \dots, P$; it yields $\sigma_{e_n}^2 = \sigma_n^2 \cdot (1 - \rho^2) = \sigma_{\epsilon_n}^2$, as it can be easily verified. Thus, the residual variance of the noise after de-correlation may be approximated from the estimated variance of the correlated noise, i.e. $\hat{\sigma}_n^2$, and from its estimated CC, $\hat{\rho}$, as

$$\sigma_{e_n}^2 \cong \hat{\sigma}_n^2 \cdot (1 - \hat{\rho}^2) \quad (14)$$

the approximation being as more accurate as the predictor attains the optimal MMSE performance.

2.3. Entropy modelling

Given a stationary memory-less source S uniquely defined by its PDF, $p(x)$, having zero-mean and variance σ^2 , linearly quantised with a step size Δ , the minimum bit-rate needed to encode one of its symbols is (Ref. 12):

$$R \cong h(S) - \log_2 \Delta \quad (15)$$

in which $h(S)$ is the differential entropy of S defined as

$$h(S) = - \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx = \frac{1}{2} \log_2(c \cdot \sigma^2) \quad (16)$$

with $0 < c \leq 2\pi e$ a positive constant accounting for the shape of the PDF and attaining its maximum for a Gaussian function. Such a constant will be referred in the following as *entropy factor*. The approximation in (15) holds for $\sigma \gg \Delta$, but is still acceptable for $\sigma > \Delta$ (Ref. 13).

Now, the minimum average bit-rate R_g necessary to reversibly encode an integer-valued sample of g , may be approximated as in Eq. (15) in which prediction errors are regarded as an uncorrelated source $G \equiv \{e_g(i)\}$ and are linearly quantised with a step size $\Delta = 1$:

$$R_g \cong h(G) = \frac{1}{2} \log_2(c_g \cdot \bar{\sigma}_{e_g}^2) \quad (17)$$

in which $\bar{\sigma}_{e_g}^2$ is the *average* variance of $e_g(i)$. By averaging (13) and replacing it into (17), R_g may be written as

$$R_g \cong \frac{1}{2} \log_2[c_g \cdot (\bar{\sigma}_{e_f}^2 + \sigma_{e_n}^2)] \quad (18)$$

where $\bar{\sigma}_{e_f}^2$ is the *average* variance of $\sigma_{e_f}^2(i)$. If $\sigma_{e_f}^2 = 0$, then (18) reduces to

$$R_g \equiv R_n = \frac{1}{2} \log_2(c_n \cdot \sigma_{e_n}^2) \quad (19)$$

in which $c_n = 2\pi e$ is the entropy factor of the PDF of e_n , if n is Gaussian. Analogously, if $\sigma_{e_n}^2 = 0$, then (18) becomes:

$$R_g \equiv R_f = \frac{1}{2} \log_2(c_f \cdot \bar{\sigma}_{e_f}^2) \quad (20)$$

in which $c_f \leq 2\pi e$ is the entropy factor of prediction errors of the noise-free image, that are generally non-Gaussian.

The average entropy of the noise-free signal f in the case of correlated noise will be given by replacing (14) in (20) to yield

$$R_f = \frac{1}{2} \log_2\{c_f \cdot [\bar{\sigma}_{e_g}^2 - (1 - \rho^2) \cdot \sigma_n^2]\}. \quad (21)$$

Since $\bar{\sigma}_{e_g}^2$ can be measured during compression by averaging $\sigma_{e_g}^2$, c_f is the only unknown parameter and its determination is crucial for the estimation accuracy of R_f . c_f is obtained by modelling e_f as a Generalised Gaussian density (GGD) function whose parameters are to be estimated. After reporting the definition of GGD, next two sections show how model parameters of GGDs can be estimated and how correlated noise can be modelled by GGDs.

2.4. Generalised Gaussian PDF

A model suitable for describing unimodal non-Gaussian amplitude distributions may be achieved by varying the parameters ν (shape factor) and σ (standard deviation) of

the *Generalised Gaussian* density (GGD) (Ref. 14, 15, 16), which is defined as

$$p_{GG}(x) = \left[\frac{\nu \cdot \eta(\nu, \sigma)}{2 \cdot \Gamma(1/\nu)} \right] \exp\{-[\eta(\nu, \sigma) \cdot |x|]^\nu\} \quad (22)$$

in which

$$\eta(\nu, \sigma) = \frac{1}{\sigma} \left[\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)} \right]^{1/2} \quad (23)$$

and $\Gamma(\cdot)$ is the Gamma function, i.e., $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$, $z > 0$. Since $\Gamma(n) = (n-1)!$, when $\nu = 1$ a Laplacian law is obtained; $\nu = 2$ yields a Gaussian distribution. As limit cases, for $\nu \rightarrow 0$, $p_{GG}(x)$ becomes an impulse function, yet having extremely heavy tails and thus nonzero σ^2 variance, whereas for $\nu \rightarrow \infty$, $p_{GG}(x)$ approaches a uniform distribution having variance σ^2 as well. The shape parameter ν rules the exponential rate of decay: the larger the ν the flatter the PDF; the smaller the ν , the more peaked the PDF. Fig. 1(a) shows the trend of the GG function for different values of ν .

The matching between a GGD and the empirical data distribution can be obtained following a maximum likelihood (ML) approach (Ref. 17), having the disadvantage of a cumbersome numerical solution. In this work, three simple and effective methods, suitable for real-time applications and based on fitting a parametric function of the modelled source to statistics calculated from the observed data, will be reviewed hereafter.

2.4.1. Higher-order moment method

This method falls into the category of moment methods, being based on matching the moments of the data set with those of the assumed distribution, in order to estimate variance and shape factor. For a GGD, the ratio of second-order moment to square root of fourth-order moment is a steadily increasing function of the shape factor ν (Ref. 18):

$$\mathcal{F}_{HM}(\nu) = \frac{\Gamma(3/\nu)}{\sqrt{\Gamma(1/\nu) \cdot \Gamma(5/\nu)}}. \quad (24)$$

Given N *i.i.d.* zero-mean random variables, $\{x_1, x_2, \dots, x_N\}$, following a GGD $p(x)$, let $\hat{\mu}_2 = \frac{1}{N} \sum_{i=1}^N x_i^2$ be the estimated second-order moment and $\hat{\mu}_4 = \frac{1}{N} \sum_{i=1}^N x_i^4$ the estimated fourth-order moment. The parameter ν is estimated by inverting (24), that is by solving

$$\hat{\nu} = \mathcal{F}_{HM}^{-1} \left(\frac{\hat{\mu}_2}{\sqrt{\hat{\mu}_4}} \right). \quad (25)$$

In practical implementations, the values of (24) are calculated at uniform steps sizes and stored in a look-up table whose entries are the corresponding values of $\hat{\mu}_2/\sqrt{\hat{\mu}_4}$.

2.4.2. Mallat's method

Another moment-based method was briefly introduced by Mallat (Ref. 19) and then developed in greater detail by Sharifi and Leon-Garcia (Ref. 20). Again, the ratio of mean absolute value to standard deviation of a GGD is a steadily increasing function of the shape factor ν :

$$\mathcal{F}_M(\nu) = \frac{\Gamma(2/\nu)}{\sqrt{\Gamma(1/\nu) \cdot \Gamma(3/\nu)}}. \quad (26)$$

Given N *i.i.d.* zero-mean random variables, obeying to a GGD $p(x)$, let $m_1 = \frac{1}{N} \sum_{i=1}^N |x_i|$ be the estimate of the mean absolute value and $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N x_i^2$ the sample variance. The parameter ν is estimated by inverting (26), that is by solving

$$\hat{\nu} = \mathcal{F}_M^{-1} \left(\frac{m_1}{\hat{\sigma}} \right). \quad (27)$$

Again, the values of (26) are pre-calculated and stored in a look-up table indexed by the values of $m_1/\hat{\sigma}$.

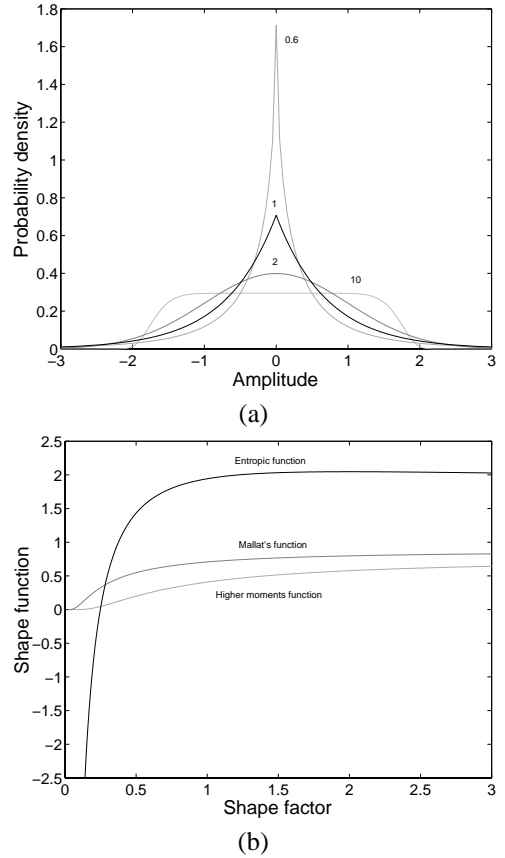


Figure 1. (a) unity-variance GG density plotted for several ν 's; (b) shape functions \mathcal{F}_{HM} (24), \mathcal{F}_M (26) and \mathcal{F}_H (29) of a GG PDF as a function of the shape factor ν .

2.4.3. Entropy matching method

The method developed by the authors (Ref. 21) relies on fitting the *entropy* of the modelled source to that of the empirical data. Since entropy is not a moment, whose accuracy of estimation requires larger and larger sample sizes, as the order increases, this method is particularly suitable for small sample sizes.

Given a GGD (22), its differential entropy (16), h_{GG} , is a function of both σ and ν

$$h_{GG}(\sigma, \nu) = -\log_2 \left[\frac{\nu \cdot \eta(\nu, \sigma)}{2 \cdot \Gamma(1/\nu)} \right] + \frac{1}{\nu \ln 2}. \quad (28)$$

Let $H = -\sum_l p_l \log_2 p_l$ be the entropy of the memory-less source, in which the p_l 's are the probabilities of the integer-valued data obtained after quantising with $\Delta = 1$. Replacing (23) in (28) and equating (28) to H yield

$$H - \log_2 \sigma = -\log_2 \left[\frac{\nu \cdot \Gamma(3/\nu)^{1/2}}{\nu \cdot \Gamma(1/\nu)^{3/2}} \right] + \frac{1}{\nu \ln 2} \triangleq \mathcal{F}_H(\nu). \quad (29)$$

Since (29) is a steadily increasing function, for $0 < \nu < 2$, once H and $\hat{\sigma}$ have been calculated from the sample data, the estimated shape factor $\hat{\nu}$ is found by inverting (29):

$$\hat{\nu} = \mathcal{F}_H^{-1}(H - \log_2 \hat{\sigma}). \quad (30)$$

The values of (29) are calculated at non-uniform step sizes (increasing with ν) and stored in a look-up table whose entries are the values of $H - \log_2 \hat{\sigma}$.

Fig. 1(b) shows the trends of higher-moment, Mallat's and entropy-matching functions against the GGD shape factor ν . In particular, the entropy-matching function measures the entropy of a stationary memory-less source that emits unity-variance GG-distributed symbols quantised with a unity step size (Ref. 21). It attains the maximum for $\nu = 2$, $\mathcal{F}_H(2) = \log_2 \sqrt{2\pi e} \approx 2.04$, and yields $\log_2(\sqrt{2e}) \approx 1.94$ when $\nu = 1$. Eventually, $\lim_{\nu \rightarrow \infty} \mathcal{F}_H(\nu) = \log_2(2\sqrt{3}) \approx 1.79$.

The entropy-matching method cannot be used for ν close to 2, because the function (29) cannot be univocally inverted around $\nu = 2$. Since the slope of the function to be inverted is related to the accuracy of inversion, the moments method is preferable to Mallat's method around the Gaussian case. The latter, however, is more accurate for small ν . The entropy method is more accurate than the other two methods for $\nu < 1$, especially when $\nu < 0.5$. The moments method fails for $\nu < 0.15$ because its function turns out to be practically flat in that case. Conversely, the entropy method yields the most accurate results for very small ν , provided that the function is non-uniformly sampled.

2.5. GG modelling of correlated noise

Once the estimated samples of correlated noise $\hat{n}(i)$ have been found by means of (6), their amplitude can be easily

matched by a GGD. However, $n(i)$ are not realizations of a memory-less source, if the noise is correlated. In that case, the decorrelated noise residues, $e_n(i)$, are better suitable for describing the entropy of the noise through a parametric model.

Eq. (2) shows that if the noise residue $\epsilon_n(i)$ is Gaussian, the noise $n(i)$ will be Gaussian as well, regardless of correlation. Hence, if $n(i)$ is estimated and found to be Gaussian, $\epsilon_n(i)$ will be white and Gaussian, thereby having parametric entropy exactly known (19). Conversely, if $n(i)$ is not Gaussian, but its amplitude is described by a GGD with standard deviation σ_n and shape factor $\nu_n \neq 2$, then also $\epsilon_n(i)$ is GG-distributed with variance $\sigma_n^2(1 - \rho^2)$ and shape factor $\nu_{\epsilon_n} \neq \nu_n$, that depends on both ν_n and ρ (Ref. 18). A relationship found between the fourth-order moments of $n(i)$ and $\epsilon_n(i)$ yields:

$$\nu_{\epsilon_n} = \mathcal{F}_{\mu_4}^{-1} \left[\left(\mathcal{F}_{\mu_4}(\nu_n) - \frac{6\rho^2}{1 + \rho^2} \right) \cdot \frac{1 + \rho^2}{1 - \rho^2} \right] \quad (31)$$

where

$$\mathcal{F}_{\mu_4}(\nu) \triangleq \frac{\Gamma(5/\nu)\Gamma(1/\nu)}{\Gamma^2(3/\nu)}. \quad (32)$$

Therefore, once ν_n and ρ are estimated, the shape factor of the correlated noise residue, ν_{ϵ_n} is found by calculating $\mathcal{F}_{\mu_4}(\nu_n)$ (32), replacing its value in (31) and inverting \mathcal{F}_{μ_4} by means of a look-up table.

2.6. Information-theoretic assessment

Let us assume that the real-valued $e_g(i)$ may be modelled as a GGD. From (17) the entropy function is

$$\frac{1}{2} \log_2(c_g) = R_g - \log_2(\bar{\sigma}_{e_g}) = \mathcal{F}_H(\nu_{e_g}) \quad (33)$$

in which ν_{e_g} is the shape factor of $e_g(i)$, the average rate of which, R_g , has been set equal to the entropy H of the discrete source. The ν_{e_g} is found by inverting either the entropy function $\mathcal{F}_H(\nu_{e_g})$, or any other shape function; in that case the $e_g(i)$ produced by DPCM, instead of its variance, is directly used. Eventually, the parametric PDF of the uncorrelated observed source $e_g(i)$ is available.

The term $e_g(i)$ is obtained by adding a sample of white non-Gaussian noise of variance $\sigma_{e_n}^2$, approximately equal to $(1 - \rho^2) \cdot \sigma_n^2$, to a sample of the noise-free uncorrelated non-Gaussian signal $e_f(i)$. Furthermore, $e_f(i)$ and $e_n(i)$ are independent of each other.

Therefore, the GG PDF of e_g previously found will be given by the linear convolution of the unknown $p_{e_f}(x)$ with a GG PDF having variance $\sigma_{e_n}^2$ and shape factor ν_{e_n} . By assuming that the PDF of the noise-free residue, $p_{e_f}(x)$, is GG as well, its shape factor ν_{e_f} can be obtained starting from the forward relationship

$$p_{GG}[\bar{\sigma}_{e_g}, \nu_{e_g}](x) = p_{GG}[\sqrt{\bar{\sigma}_{e_g}^2 - \sigma_{e_n}^2}, \nu_{e_f}](x) \otimes p_{GG}[\sigma_{e_n}, \nu_{e_n}](x) \quad (34)$$

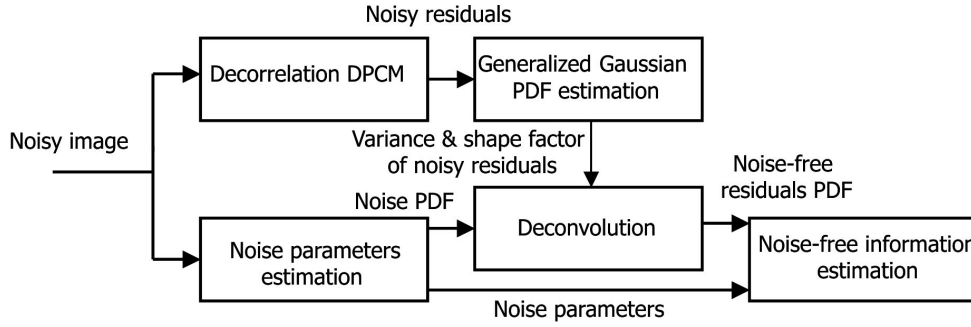


Figure 2. Flowchart of the information-theoretic assessment procedure for a digital signal.

by de-convolving the PDF of noise residue from that of noisy-signal residue.

In a practical implementation, the estimated value of ν_{e_f} is found such that the direct convolution at right side of (34) yields a GGD, whose shape factor matches ν_{e_g} as much as possible.

Eventually, the estimated shape factor $\hat{\nu}_{e_f}$ is used to determine the entropy function

$$\frac{1}{2} \log_2(c_f) = \mathcal{F}_H(\hat{\nu}_{e_f}) \quad (35)$$

which is replaced in (21) to yield the entropy of the noise-free signal R_f .

Fig. 2 summarises the overall procedure.

Extension of the procedure to two-dimensional (2D) and three-dimensional (3D) signals, i.e. to digital images and sequences of digital images is straightforward. In the former case, 2D prediction is used to find e_g , two correlation coefficients, ρ_x and ρ_y , are estimated for the noise, whose variance after decorrelation is approximated as $\sigma_n^2(1 - \rho_x^2)(1 - \rho_y^2)$, by assuming a separable 2D Markov model. Analogously, Eq. (6), defining the estimated value of a sample of correlated noise, is extended as

$$\hat{n}(i, j) = \sqrt{\frac{N^2}{(N^2 - \left(1 + 2\rho_x \frac{1 - \rho_x^m}{1 - \rho_x}\right) \left(1 + 2\rho_y \frac{1 - \rho_y^m}{1 - \rho_y}\right))}} \cdot [g(i, j) - \bar{g}(i, j)]$$

and generalises the model reported in (Ref. 7), as previously discussed in sect. 2.1. $N = 2m + 1$ is the length of the side of the square window on which the average $\bar{g}(i, j)$ is calculated.

The 3D extension is more critical because a sequence of images may have noise variances and spatial CCs different for each image. Moreover, it is often desirable to estimate entropy of the individual images of the sequence. Therefore, each image is de-correlated both spatially and along the third dimension by using a 3D prediction.

3. EXPERIMENTAL RESULTS

3.1. CHRIS hyperspectral data



Figure 3. Colour composite of CHRIS data acquired on 18 September 2003: red 661 nm, green 502 nm, blue 442 nm.

The proposed information-theoretic procedure was run on several hyperspectral sequences collected by CHRIS over the San Rossore test site, in Central Italy on different dates. The sequences are constituted by 18 bands with a mean width of 14.7 nm, in the range 442–1015 nm and a ground resolution of about 17 m. The size of each image is 744 pixels across track and 748 along track. The data have a dynamic range of 19 bit. The RCI (Restored Corrected Images) data have been provided by Sira Electro-Optics Ltd. Company, whereas the LIB data have been radiometrically corrected at IFAC-CNR; an efficient de-striping algorithm has also been applied to mitigate the striping effect due to the diversity in gain and offset of the imaging sensor elements. Three main objectives were considered. The first was aimed at evaluating the quality of the radiometric corrections performed at IFAC-CNR.

The second was devoted at assessing the noise model on different classes of data while the third consisted in verifying the variations of the noise parameters for different observations.

3.1.1. Radiometric correction assessment

The first step of the procedure concerns the estimation of the noise parameters for each band. Fig. 4 reports the plots of noise standard deviation σ_n before and after radiometric corrections. The amount of noise is larger for

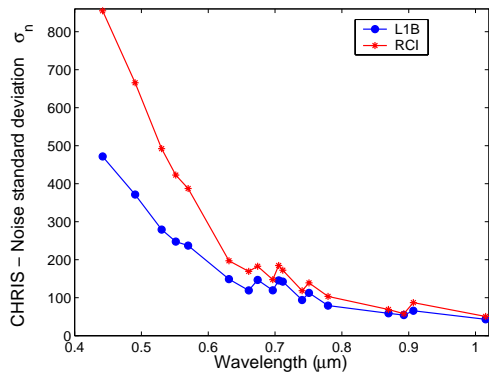


Figure 4. Noise standard deviation of CHRIS test image plotted vs. wavelength.

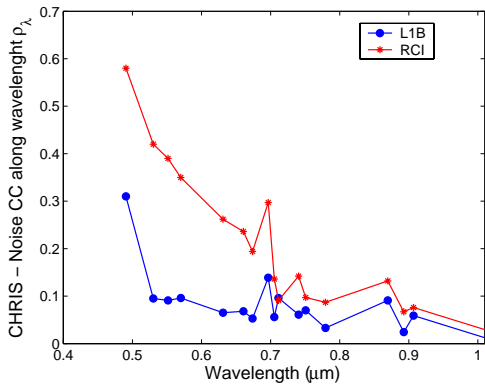


Figure 5. Spectral CC ρ_λ plotted vs. wavelength

shorter wavelengths. The de-striping process has the effect of reducing the contribution of noise. That is apparent for shorter wavelength. As a consequence, the curve tends to go down and becomes more flat in wavelength.

A similar trend in wavelength appears in Fig. 5, where the spectral CC, ρ_λ , of the noise is plotted. Analogously to the standard deviation, ρ_λ is larger in the VIS than in the other parts of the spectrum. As expected, the de-striping process tends to reduce the spectral correlation, because it processes each band independently of each other.

In a similar way, we can justify the trend of the plots of the CCs of the noise across and along track, reported in

Fig. 6 and 7, respectively. In fact, filtering across track has the effect of reducing the correlation along track and, conversely, to increase the correlation across track.

Eventually, the shape factor of the GG-modelled noise is reported in Fig. 8 for L1B data. Fig. 8 shows that the noise of CHRIS images, after radiometric correction and de-striping, is Gaussian with a very good approximation.

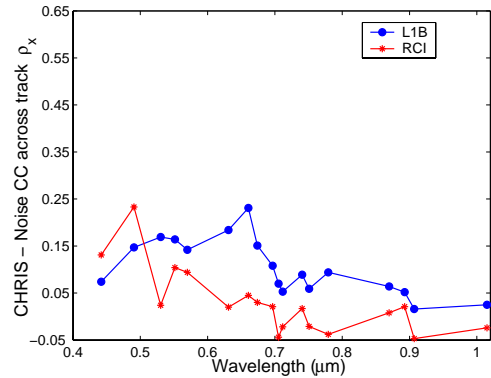


Figure 6. Across-track CC ρ_x plotted vs. wavelength.

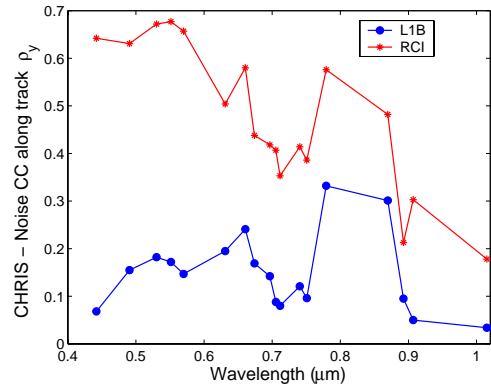


Figure 7. Along-track CC ρ_y plotted vs. wavelength.

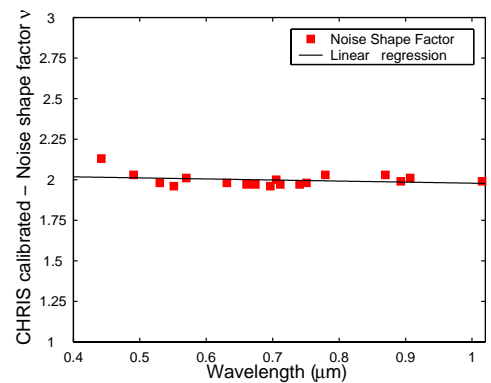


Figure 8. Noise shape factor ν vs. wavelength.

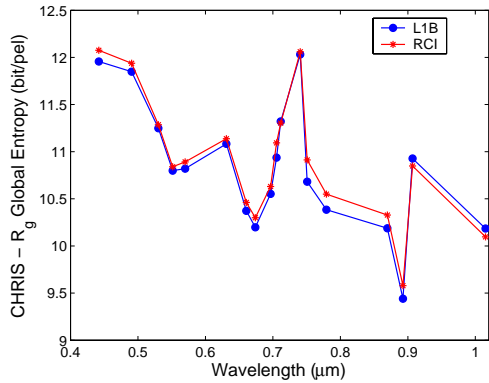


Figure 9. Estimated information R_g of observed radiance data plotted vs. wavelength.

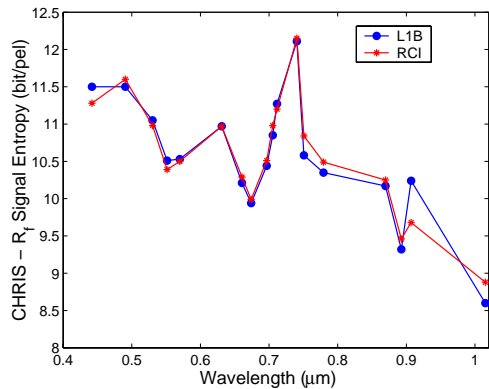


Figure 10. Estimated information R_f of ideal noise-free radiance data plotted vs. wavelength.

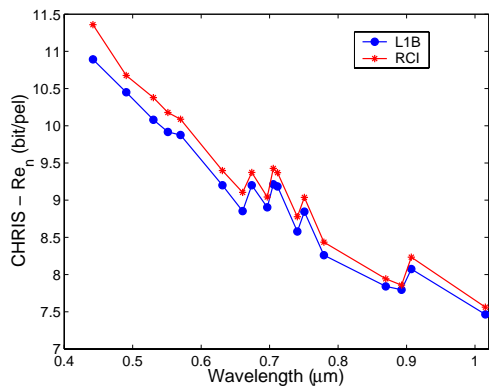


Figure 11. Estimated information R_n due to noise for the observed radiance data plotted vs. wavelength.

After the noise parameters have been determined, the information-theoretic assessment procedure has been run. Fig. 9, 10, and 11, report the scores relative to information content varying with the wavelength. R_g and R_f are perfectly in trend, whereas R_{e_n} has been slightly reduced by the filtering step of the de-stripping process.

Notwithstanding the useful information content is un-

changed after the calibration procedure, the user can extract information more easily from the calibrated images because the de-stripping process has diminished the noise level that has the effect to make the information content less intelligible.

It is noticeable that the results obtained both for the RCI and calibrated data are very similar, confirming, as expected, that the calibration process does not change the information content of the acquired data.

3.1.2. Variability with landscape

Notwithstanding the information assessment procedure is able to find homogeneous areas occurring at any place of the image, it resulted that all the regions selected by the program were located mostly on the sea. That happened in particular when a restricted number of regions was selected in order to improve the reliability of the estimation.

To investigate on this aspect two subimages were initially considered. The first, namely *Sea*, was constituted by sea pixel only, while the latter, namely *Land*, by land pixel only. On both subimages the information assessment procedure was run. As expected no difference was found between the whole image and the *Sea* subimage. Concerning the *Land* subimage, we found that, depending on the band, the standard deviation of noise, σ_n , substantially increased with respect to the *Sea* image. Initially we attributed this behaviour to a possible *bias* in σ_n caused by the presence of texture on areas that were believed homogeneous but that probably were not. In fact this conjecture was also supported by a general increase of the values of the correlation coefficients as it appears in Fig. 12 where ρ_x and ρ_y for the *Land* subimage are reported.

In order to verify the presence of such a bias we decided to estimate σ_n also with the *bit plane* algorithm reported in (Ref. 7). This approach is less sharp than the procedure that adopts homogeneous regions but has the advantage to be independent of region homogeneity.

Surprisingly, the results obtained by the *bit plane* procedure confirmed that σ_n is somewhat greater on the *Land* than on the *Sea* image, thus justifying the behaviour of σ_n reported in Fig. 13 where σ_n is reported for the *Land*, *Sea*, *River1*, and *River2* subimages, the last two being narrow rectangular areas around the *Arno* river that appears in the bottom part of Fig. 3.

River1, and *River2* subimages have different geometric size. The difference in size has been deliberately fixed in order to force the estimation algorithm to work on homogeneous areas clustered in different conditions. In particular, the estimation that is more reliable (on *River2*) is obtained on the smallest subimage. In fact, on *River2* the algorithm is able to clusterise the homogeneous areas of the river in all bands and thus derive an estimation that is extremely similar to that of the *Sea* image.

Concerning the explanation of the bias, it might be due to

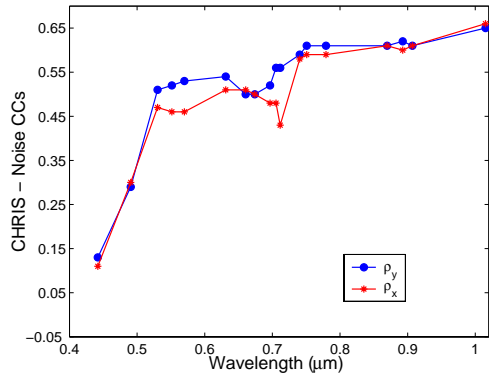


Figure 12. CC ρ_x and ρ_y estimated on Land subimage and plotted vs. wavelength.

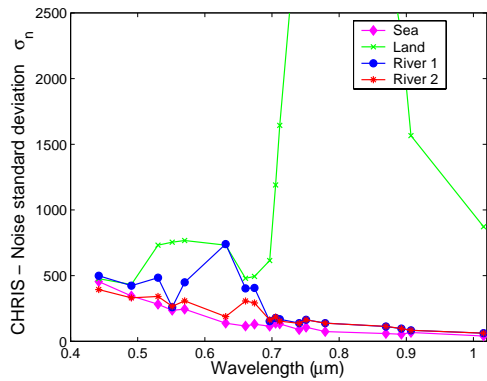


Figure 13. Noise standard deviation plotted vs. wavelength for different landscapes; some values of σ_n for Land subimage are out of scale and have been clipped

a *non-linear* processing of the image occurred when passing from the raw *LO* data to the instrument corrected *RCI* data. A not perfectly additive noise component, which could be modelled with the introduction of a multiplicative signal-dependent noise component, might also contribute to this effect. In fact, due to the relatively large dynamic range, also a slight multiplicative noise component could easily explain the increase of σ_n on land areas.

In order to clarify this point, further analysis should be developed on the raw data since *RCI* images have been corrected by a *sensor-dependent* gain whose effect is to modify the statistics of the acquired data.

The presence of a slight multiplicative noise component could be easily taken into account by the noise model but would also cause implications in the estimation of the entropy of the noise. Actually, since σ_n is mainly estimated by means of the contribution of sea areas, and thus probably under estimated, R_f might consequently be over estimated and the reported scores might represent an upper bound.

3.1.3. Temporal variation

A preliminary analysis was also carried out to investigate on temporal variation of the data. The analysis was performed on the images acquired on July 2003, September 2003, and January 2004. The different meteorological conditions that occurred during the observations obviously impose some caution in the interpretation of the results. In particular, as it appears in Fig. 14, the behaviour of the estimated σ_n in the image acquired in January is influenced by the conditions of the sea that is particularly rough. Notwithstanding this premise some not negligible differences appear in the plots of σ_n at all wavelengths. Further investigations should be considered with further information to be provided on the acquisition process.

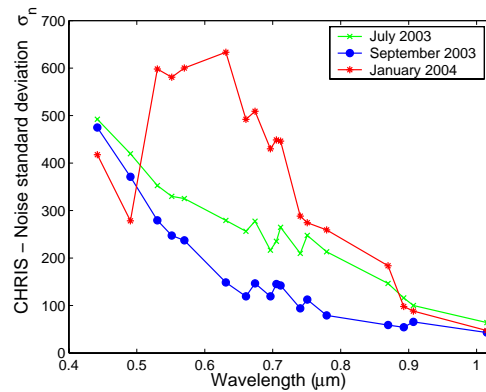


Figure 14. σ_n vs. wavelength for three different dates.

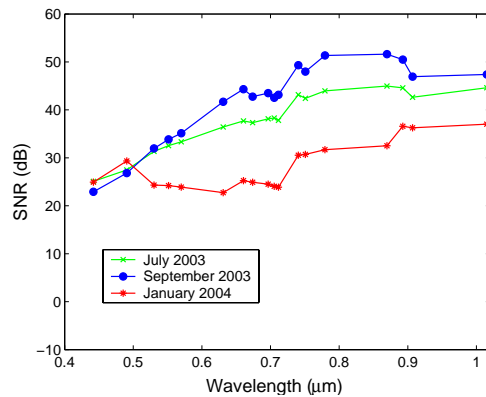


Figure 15. SNR vs. wavelength for the three acquisitions.

Eventually, the Signal to Noise Ratio (SNR),

$$SNR_{(dB)} \triangleq 10 \cdot \log_{10} \frac{\bar{\sigma}_g^2 - \hat{\sigma}_n^2}{\hat{\sigma}_n^2}. \quad (36)$$

is reported in Fig. 15 for the three different observations. Apart from the shorter wavelength in January's image, the three plots present a similar trend and confirm the reliability of the procedure. It should be noticed that the SNR is signal dependent and a direct comparison of its value for the three plots is not significant.

4. CONCLUDING REMARKS

A procedure for information-theoretic assessment of digitised multi-dimensional signal has been described. It relies on robust noise estimation and on parametric entropy modelling to calculate the information of the ideal noise-free signal starting from the digitised observed signal. From the code rate and the estimated noise variance and CC's a model was suggested to upper bound the amount of information generated by an ideally noise-free process of sampling and digitisation. The procedure has been utilised on CHRIS hyperspectral image data to estimate noise parameters and information content of each spectral band. Results between RCI and L1B data have been found consistent. Some discrepancy was found in the estimation of noise parameters on areas of different texture and mean grey value. Possible explanations have been preliminarily discussed. Concerning variations found in noise parameters of multitemporal observations, further and deeper investigations are needed to understand the possible causes.

ACKNOWLEDGEMENTS

The authors wish to warmly thank their former co-authors L. Alparone, of the University of Florence, and A. Barducci, of IFAC-CNR, for the valuable discussions on hyperspectral data analysis.

REFERENCES

1. Huck, F. O., Fales, C. L., Alter-Ganterberg, R., Park, S. K., and Rahman, Z. Information-theoretic assessment of sampled imaging systems. *J. Optical Engin.*, 38(5):742–762, May 1999.
2. Park, S. K. and Rahman, Z. Fidelity analysis of sampled imaging systems. *J. Optical Engin.*, 38(5):786–800, May 1999.
3. Aiazzi, B., Alparone, L., Barducci, A., Baronti, S., and Pippi, I. Assessment of noise variance and information content of multi-/hyper-spectral imagery. *ISPRS Internat. Archives Photogramm. Remote Sensing*, 32(7-4-3W6):167–174, July 1999.
4. Benediktsson, J. A., Sveinsson, J. R., and Arnason, K. Classification and feature extraction of AVIRIS data. *IEEE Trans. Geosci. Remote Sensing*, 33(5):1194–1205, Sep. 1995.
5. Shannon, C. E. and Weaver, W. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, 1949.
6. Aiazzi, B., Alparone, L., Barducci, A., Baronti, S., and Pippi, I. Information-theoretic assessment of sampled hyperspectral imagers. *IEEE Trans. Geosci. Remote Sensing*, 39(7):1447–1458, July 2001.
7. Aiazzi, B., Alparone, L., Barducci, A., Baronti, S., and Pippi, I. Estimating noise and information of multispectral imagery. *J. Optical Engin.*, 41(3):656–668, Mar. 2002.
8. Aiazzi, B., Alparone, L., and Baronti, S. Fuzzy logic-based matching pursuits for lossless predictive coding of still images. *IEEE Trans. Fuzzy Systems*, 10(4):473–483, Aug. 2002.
9. Aiazzi, B., Alparone, L., and Baronti, S. Near-lossless image compression by relaxation-labelled prediction. *Signal Processing*, 82(11):1619–1631, Nov. 2002.
10. Aiazzi, B., Alba, P., Alparone, L., and Baronti, S. Lossless compression of multi/hyper-spectral imagery based on a 3-D fuzzy prediction. *IEEE Trans. Geosci. Remote Sensing*, 37(5):2287–2294, Sep. 1999.
11. Aiazzi, B., Alparone, L., and Baronti, S. Near-lossless compression of 3-D optical data. *IEEE Trans. Geosci. Remote Sensing*, 39(11):2547–2557, Nov. 2001.
12. Jayant, N. S. and Noll, P. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall, Englewood Cliffs, NJ, 1984.
13. Roger, R. E. and Arnold, J. F. Reversible image compression bounded by noise. *IEEE Trans. Geosci. Remote Sensing*, 32(1):19–24, Jan. 1994.
14. Farvardin, N. and Modestino, J. W. Optimum quantizer performance for a class of non-Gaussian memoryless sources. *IEEE Trans. Inform. Theory*, IT-30(5):485–497, Sep. 1984.
15. Varanasi, M. K. and Aazhang, B. Parametric generalized Gaussian density estimation. *J. Acoust. Soc. Amer.*, 86(4):1404–1415, 1989.
16. Birney, K. A. and Fischer, T. R. On the modeling of DCT and subband image data for compression. *IEEE Trans. Image Processing*, 4(2):186–193, Feb. 1995.
17. Müller, F. Distribution shape of two-dimensional DCT coefficients of natural images. *Electronics Lett.*, 29(22):1935–1936, 1993.
18. Niehsen, W. Generalized Gaussian modeling of correlated signal sources. *IEEE Trans. Signal Processing*, 47(1):217–219, Jan. 1999.
19. Mallat, S. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-11(7):674–693, July 1989.
20. Sharifi, K. and Leon-Garcia, A. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Trans. Circuits Syst. Video Technol.*, 5(1):52–56, 1995.
21. Aiazzi, B., Alparone, L., and Baronti, S. Estimation based on entropy matching for generalized Gaussian PDF modeling. *IEEE Signal Processing Lett.*, 6(6):138–140, June 1999.