INFORMATION-THEORETIC ASSESSMENT OF OPTICAL REMOTE-SENSING IMAGERY

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

[(1)] *IFAC–CNR: Institute of Applied Physics "Nello Carrara", Italian National Research Council*
Via Panciatichi, 64, 50127 Firenze (Italy), {*B.Aiazzi S.Baronti C.Lastri L.Santurri M.Selva*}*@ifac.cnr.it*
[(2)] *DET-UniFI: Department of Electronics and Telecommunications, University of Firenze*
Via Santa Marta, 3, 50139 Firenze (Italy), Alparone@lci.det.unifi.it

ABSTRACT

This work focuses on estimating the information conveyed to a user by multi-band remotely sensed optical data, either *multi-spectral* or *hyper-spectral*. A trade-off exists between *spatial* and *spectral* resolution, due to physical constraints of sensors imaging with a prefixed SNR. Lossless data compression is exploited to measure the useful information content of the data. The bit-rate achieved by the reversible compression process takes into account both the contribution of the "observation" noise, i.e. information regarded as statistical uncertainty, whose relevance is null to a user, and the intrinsic information of hypothetically *noise-free* radiance data. An entropy model of the image source is defined and, once the standard deviation of the noise, assumed to be Gaussian, has been preliminary measured, such a model is inverted to yield an estimate of the information content of the noise-free source from the code rate. Results of mutual information assessment are reported and discussed on Landsat TM data and on AVIRIS data.

1 INTRODUCTION

Image quality definition and assessment is a recent branch of image analysis aiming at measuring the information conveyed by digital images. As a matter of fact, information-theoretic assessment of imaging systems is presently an open problem [1, 2]. Quality may be objectively defined jointly from the *signal-to-noise ratio* (SNR) and from the entropy content [3]. Whereas several methods exist for SNR assessments, the latter is still an open problem. Accurate estimates of the *entropy rate* of an image source can be obtained provided that the data is uncorrelated. As a consequence, data de-correlation should be considered in order to suppress or, at least, largely reduce the correlation existing in natural images.

When multi-spectral images are concerned, the de-correlation algorithms should take into account not only the *spatial* but also the *spectral* correlation [4, 5, 6, 7]. As a matter of fact the entropy rate is a measure of *statistical* information, that is of *uncertainty*. Therefore any observation noise introduced by the imaging sensor will result in an increase of the entropy rate, without a corresponding enhancement of the *useful* information content [8]. An estimation of the noise must be preliminarily carried out in order to quantify its contribution to the overall source entropy rate.

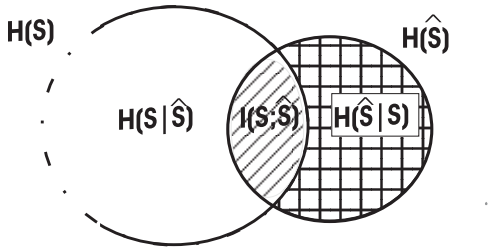


Figure 1: Relations between entropy and mutual information.

If we consider an image source S, its information content is given by the entropy rate $H(S)$. An acquisition (scanning) procedure originates an observed image $\hat{S}$, whose information is the entropy rate $H(\hat{S})$. $H(\hat{S})$ may not be adequate to measure acquired information, since the observed image will generally not coincide with the digitised source, mainly because of the observation noise. Furthermore, the source may be not exactly band-limited by half of the sampling frequency; hence, the non-ideal sampling is responsible for an additional amount of noise generated by aliasing. Therefore, only a fraction of the source information is conveyed by the digitised image.

The amount of source information that is not contained in the digitised image is measured by the conditional entropy $H(S|\hat{S})$, or equivocation, that is the residual uncertainty on the original source when the observed source is known. The contribution of the overall noise (i.e. *aliasing* and acquisition) to the entropy rate of the digitised source is measured by the conditional entropy $H(\hat{S}|S)$, which represents the uncertainty on the observed source $\hat{S}$ when the original source S is known. Therefore, the larger the noise, the larger $H(\hat{S})$, even if the information

on the original source is not increased. A suitable measure of the information content of an acquired source is represented by the mutual information [9]

$$I(S;\hat{S}) \;=\; H(S) - H(S|\hat{S}) \;=\; H(\hat{S}) - H(\hat{S}|S). \quad (1)$$

Fig. 1 describes the relationship existing between the entropy rate of the original source and of the observed source.

Once the standard deviation of the observation noise has been measured, the bit-rate produced by the reversible encoder will be utilised to yield an estimate of the true information content of the multi-spectral source, i.e. of the entropy 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 change when a stationary white/non-white Gaussian noise is superimposed. Such a model can be inverted to yield the entropy of the noise-free source from that of the noisy source and the variance of the noise.

Experimental results carried out both on Landsat Thematic Mapper (TM) data, and on data from the *Airborne Visible InfraRed Imaging Spectrometer* (AVIRIS) are reported and discussed to validate the proposed model, to corroborate its underlying assumptions, and to provide a deeper insight on the features of widespread imaging sensors.

2 NOISE MODELLING

This section focuses on noise modelling from image data. 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. Solutions will be devised for the noise typically introduced by multi-spectral scanners and imaging spectrometers.

Let us assume for the noise an additive signal-independent model:

$$g(i,j,k) \;=\; f(i,j,k) + n(i,j,k) \quad (2)$$

in which $g(i,j,k)$ is the recorded intensity at pixel position (i,j) of the kth spectral band and $f(i,j,k)$ the kth component of the spectral reflectance at (i,j). Both $g(i,j,k)$ and $f(i,j,k)$ are regarded as non-stationary non-Gaussian (auto)correlated stochastic processes. The term $n(i,j,k)$ is a zero-mean Gaussian process independent of f, stationary along (i,j) but not along k and correlated, both spatially and spectrally. The variance is $\sigma_n^2(k)$, and the CC's are ρ_x and ρ_y, along and across the scan line, respectively, and ρ_λ in the spectral direction. The CC's are assumed to be constant throughout.

The variance of (2) can be easily calculated as

$$\sigma_g^2(i,j,k) \;=\; \sigma_f^2(i,j,k) + \sigma_n^2(k) \quad (3)$$

thanks to the independence between radiance and noise components and to the spatial stationarity of the latter. From (3) it appears that $\sigma_n^2(k)$ can be estimated by averaging $\sigma_g^2(i,j,k)$ in homogeneous areas of the kth band, where $\sigma_f^2(i,j,k) \equiv 0$, by definition. The homogeneity requirement is crucial because the radiance components are generally cross-correlated with one another, besides being *spatially* auto-correlated. Since an area may be homogeneous in one band, but not in another, in the following the analysis will be carried out on separate bands, notwithstanding they may exhibit noise components that are dependent on one another. For sake of clarity the index k will be omitted to indicate a generic band.

By assuming a uniform signal distribution between the levels 0 and $2^L - 1$, the SNR can be estimated as a *peak* SNR:

$$PSNR_{(\mathrm{dB})} \;\triangleq\; 10 \cdot \log_{10} \frac{(2^L-1)^2}{12 \cdot \sigma_n^2} \quad (4)$$

The assumption of uniformly distributed signal, however, is unlikely for multi-spectral scanners. The average variance of the signal f, $\bar{\sigma}_f^2$, which is given by $\bar{\sigma}_g^2 - \sigma_n^2$, is a more likely measurement of signal power

$$SNR_{(\mathrm{dB})} \;\triangleq\; 10 \cdot \log_{10} \frac{\bar{\sigma}_g^2 - \hat{\sigma}_n^2}{\sigma_n^2} \quad (5)$$

The variance and CC's of the noise have been estimated by means of the scatter-plot based methods described in [3, 10].

3 MUTUAL INFORMATION ASSESSMENT

3.1 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 outcome 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 a 3D *adaptive* prediction [6].

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

$$\hat{g}(i) \;=\; \sum_{j=1}^{P} \phi(j) \cdot g(i-j) \quad (6)$$

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

By replacing the additive noise model one obtains:

$$\hat{g}(i) \;=\; \hat{f}(i) + \sum_{j=1}^{P} \phi(j) \cdot n(i-j) \tag{7}$$

in which

$$\hat{f}(i) \;=\; \sum_{j=1}^{P} \phi(j) \cdot f(i-j) \tag{8}$$

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

$$e_g(i) \;\triangleq\; g(i) - \hat{g}(i) \;=\; e_f(i) + n(i) - \sum_{j=1}^{P} \phi(j) \cdot n(i-j) \tag{9}$$

in which $e_f(i) \triangleq f(i) - \hat{f}(i)$ is the error the predictor would produce if it were feed by 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.

3.1.1 White noise

Let us assume that the noise $n(i)$ is white and Gaussian, i.e. a stationary and uncorrelated random process with zero mean and variance σ_n^2. Under this assumption the variance of $e_g(i)$ (9) becomes

$$\sigma_{e_g}^2(i) \;=\; \sigma_{e_f}^2(i) + \left[1 + \sum_{j=1}^{P} \phi^2(j)\right] \cdot \sigma_n^2 \tag{10}$$

in which the summation term at right-hand stems from the statistical independence and stationarity of noise samples. The term $\sum_{j=1}^{P} \phi^2(j)$ represents the power gain of the linear filter responsible for the prediction and will be referred in the following as G_P. Eq. (10) states that the variance of prediction errors calculated on the *noisy* image equals that of prediction errors calculated on the *noise-free* image incremented by the noise variance multiplied by $1 + G_P$.

3.1.2 Correlated noise

When the noise is correlated (10) no longer holds, because of the statistical dependence of $n(i)$ from the previous noise samples. Let us assume for the stationary zero-mean Gaussian noise a first-order Markov model, uniquely defined by the ρ and the σ_n^2

$$n(i) \;=\; \rho \cdot n(i-1) + \epsilon_n(i) \tag{11}$$

in which $\epsilon_n(i)$ is a white Gaussian random process having variance $\sigma_{\epsilon_n}^2 \;=\; \sigma_n^2 \cdot (1-\rho^2)$. From (11) it stems that

$$E[n(i) \cdot n(i+m)] = \rho^{|m|} \cdot \sigma_n^2. \tag{12}$$

In the case of correlated noise (9) may be written as

$$e_g(i) \;=\; e_f(i) + e_n(i) \tag{13}$$

in which

$$e_n(i) \;\triangleq\; n(i) - \hat{n}(i) \;=\; n(i) - \sum_{j=1}^{P} \phi(j) \cdot n(i-j) \tag{14}$$

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

$$\sigma_{e_g}^2(i) \;=\; \sigma_{e_f}^2(i) + \sigma_{e_n}^2. \tag{15}$$

By replacing the noise model (11) into (14) and making use of (12) the variance of $e_n(i)$ may be calculated as

$$\sigma_{e_n}^2 = E\{[n(i) - \hat{n}(i)]^2\}$$
$$= \sigma_n^2 \left(1 - 2\sum_{j=1}^{P} \phi(j)\rho^j + \sum_{j=1}^{P}\sum_{l=1}^{P} \phi(j)\phi(l)\rho^{|j-l|}\right) \tag{16}$$

If the double summation term in (16) is split into the sum of two contributions, one for $j = l$, which is equal to G_P, and another for $j \neq l$, Eq. (16) may be rewritten as

$$\sigma_{e_n}^2 = \sigma_n^2 \times$$
$$\left(1 + G_P - 2\sum_{j=1}^{P} \phi(j)\rho^j + \sum_{\substack{j=1 \\ j\neq l}}^{P}\sum_{l=1}^{P} \phi(j)\phi(l)\rho^{|j-l|}\right) \tag{17}$$

It can be easily noticed that (17) reduces to (10) when the noise is white, i.e. $\rho \equiv 0$. On the other hand, the $\sigma_{e_n}^2$ 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 (11) is $\phi(1) = \rho$ and $\phi(j) = 0,\ j = 2, \cdots, P$; 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

$$\sigma_{e_n}^2 \;\cong\; \hat{\sigma}_n^2 \cdot (1 - \hat{\rho}^2) \tag{18}$$

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

3.2 Entropy Modelling

Given a stationary memoryless source S uniquely defined by its PDF $p(x)$, e.g. having zero-mean and variance σ^2, and linearly quantised with a step size Δ, the minimum bit-rate needed to encode one of its samples is [11]:

$$R \;\cong\; h(S) - log_2\Delta \tag{19}$$

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

$$h(S) \;=\; -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx \;=\; \frac{1}{2}\cdot\log_2(c\cdot\sigma^2) \tag{20}$$

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 (19) holds for $\sigma \gg \Delta$, but is still acceptable for $\sigma > \Delta$.

Now, the minimum average bit-rate R_g necessary to reversibly encode an integer-valued pixel of g, may be approximated as in Eq. (19) in which prediction errors are regarded as an uncorrelated source $G \equiv \{e_g(i)\}$ and are linearly quantised with a step size $\Delta = 1$:

$$R_g \;\cong\; h(G) \;=\; \frac{1}{2}\cdot\log_2(c_g\cdot\bar{\sigma}^2_{e_g}) \tag{21}$$

in which $\bar{\sigma}^2_{e_g}$ is the *average* variance of $e_g(i)$. By averaging (15) and replacing it into (21), R_g may be written as

$$R_g \;=\; \frac{1}{2}\log_2[c_g\cdot(\bar{\sigma}^2_{e_f}+\sigma^2_{e_n})] \tag{22}$$

where $\bar{\sigma}^2_{e_f}$ is the *average* variance of $\sigma^2_{e_f}(i)$. If $\sigma^2_{e_f}=0$, then (22) reduces to

$$R_g \;\equiv\; R_n \;=\; \frac{1}{2}\log_2(c_n\cdot\sigma^2_{e_n}) \tag{23}$$

in which $c_n = 2\pi e$ is the entropy factor of the Gaussian PDF of e_n. Analogously, if $\sigma^2_{e_n}=0$, then (22) becomes:

$$R_g \;\equiv\; R_f \;=\; \frac{1}{2}\log_2(c_f\cdot\bar{\sigma}^2_{e_f}) \tag{24}$$

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.

Thus, the average entropy rate of the noise-free signal f in the case of white noise will be given from (10) and (24) by

$$R_f \;=\; \frac{1}{2}\cdot\log_2\{c_f\cdot[\bar{\sigma}^2_{e_g}-(1+G_P)\cdot\sigma^2_n]\} \tag{25}$$

while in the case of correlated noise, (18) and (24) yield

$$R_f \;=\; \frac{1}{2}\cdot\log_2\{c_f\cdot[\bar{\sigma}^2_{e_g}-(1-\rho^2)\cdot\sigma^2_n]\}. \tag{26}$$

Since $\bar{\sigma}^2_{e_g}$ can be measured during the compression procedure by averaging $\sigma^2_{e_g}$, c_f is the only unknown parameter. Its determination is crucial for the accuracy of the estimation of R_f. A model suitable for prediction errors from DPCM may be achieved by varying the parameters ν (shape factor) and σ (standard deviation) of the *Generalised Gaussian* density (GGD). The matching between a GGD and the actual data has been obtained through an original method, recently developed by the authors [12], based on fitting the *entropy* of the modelled source to that of the empirical data.

3.3 Entropy Estimation Procedure

First, let us assume that the real-valued $e_g(i)$ may be modelled as a GGD. From (21) and [12] the entropy function is

$$\frac{1}{2}\cdot\log_2(c_g) \;=\; R_g-\log_2(\bar{\sigma}_{e_g}) \;=\; \mathcal{F}_H(\nu_g) \tag{27}$$

in which ν_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_g is found by inverting the entropy function $\mathcal{F}_H(\nu_g)$, as shown in [12]. Thus, the (average) PDF of $e_g(i)$ can be analytically formulated.

The term $e_g(i)$ is obtained by adding to $e_f(i)$ a sample of white Gaussian noise of variance $\tilde{\sigma}^2_n$ equal to either $(1+G_P)\cdot\sigma^2_n$, or to $(1-\rho^2)\cdot\sigma^2_n$, depending on whether the data noise is white or not. Furthermore, $e_f(i)$ and $n(i)$ are independent of each other. Thus, the GG PDF previously found out for p_{e_g} will be given by the linear convolution of the unknown p_{e_f} with a zero-mean Gaussian PDF having variance $\tilde{\sigma}^2_n$:

$$p_{GG}[\bar{\sigma}_{e_g},\nu_g](x) \;=\; p_{e_f}(x)\otimes N[0,\tilde{\sigma}^2_n](x) \tag{28}$$

The PDF of the noise-free residuals, $p_{e_f}(x)$, can be obtained by numerically de-convolving the analytical PDF of the noise from $p_{e_g}(x)$, which is analytical as well, being GG-modelled.

Eventually, a numerical function $\{p_l\}$ is found out for $p_{e_f}(x)$, the entropy function of which, namely

$$\mathcal{F}_H(\nu_f) \;\equiv\; \frac{1}{2}\log_2(c_f) \tag{29}$$

is estimated by inverting the entropy model and replaced in either (25) or (26) to yield the rate of the noise-free data R_f.

3.4 Extension of the Model to the 3D Case

The analysis reported above holds for a 1D predictor having fixed coefficients. Now it will be extended to the case of a 3D prediction, i.e. jointly *spatial* and *spectral* operated by an adaptive predictor, the coefficients of which are continuously recalculated from the incoming data at each pixel position in each spectral band [6].

3.4.1 White noise

Let $\sigma^2_n(k)$ denote the variance of the random observation noise superimposed to the kth band of a multi-spectral image. Let also $R(k) \equiv R_g(k)$ denote the average bit-rate needed to reversibly encode one pixel of the kth band.

Prediction errors produced by the de-correlation algorithm may still be modelled as a random process given by the sum of two independent contributions: the random prediction errors calculated to predict a hypothetically

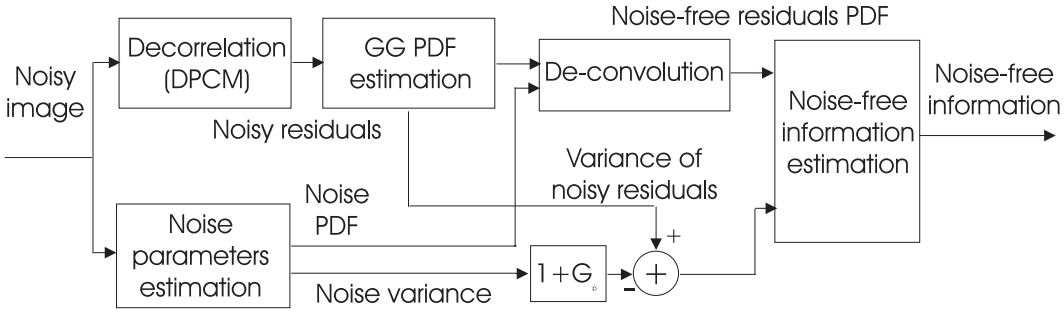


Figure 2: Flowchart of the proposed information-assessment procedure from a digital image.

noise-free image from the actual, i.e. noisy, image, plus the observation noise propagated by the predictor. The difference is that the predictor now spans more than one band, say, Q bands, each of which may exhibit a different noise variance. Furthermore, the predictor may be *adaptive* or *classified*, in both cases *space-varying* (denoted with a superscript *. Therefore, let $\bar{G}_P^*(m)$ denote the average sum of the squares of the space-varying linear predictor's coefficients lying on the band at the mth relative position with respect to the current band, i.e. the kth. Thus, $m = 0$ indicates the current band, $m = -1$ the previous one, $m = -2$ the next to previous one, and so on. Trivially, $\sum_{m=-Q+1}^{0} \bar{G}_P^*(m) = \bar{G}_P^*$.

The *average* variance of the multi-spectral prediction errors on the kth band may be written as:

$$\bar{\sigma}_{e_g}^2(k) \;=\; \bar{\sigma}_{e_f}^2(k) + \sum_{m=-Q+1}^{0} [\delta_m + G_P^*(m)] \cdot \sigma_n^2(k+m) \tag{30}$$

in which $\bar{\sigma}_f^2(k)$ is the *average* variance of prediction errors calculated on the kth band of the noise-free signal and δ_m is the Kronecker operator.

By deriving $\bar{\sigma}_f^2(k)$ from (30) and replacing it into (25), the average entropy of the noise-free kth band, $R_f(k)$, may be estimated as

$$R_f(k) \;=\; \frac{1}{2} \cdot \log_2[c_f(k) \cdot \bar{\sigma}_f^2(k)] \tag{31}$$

in which $\frac{1}{2} \cdot \log_2[c_f(k)]$ is the entropy function of the kth band and is found by de-convolving the GG-modelled PDF of the prediction residuals of the kth band, namely $p_{GG}[\bar{\sigma}_{e_g}(k), \nu_g(k)](x)$, by a zero-mean Gaussian density with variance equal to $\sum_{m=-Q+1}^{0} [\delta_m + G_P^*(m)] \cdot \sigma_n^2(k+m)$. Fig. 2 shows the flowchart of the overall procedure for information-theoretic assessment of a digitised image.

3.4.2 Correlated noise

Let also ρ_x, ρ_y, and ρ_λ denote the noise CC's which are assumed to be space-invariant as well as constant throughout the sequence of spectral bands. According to (13), prediction errors produced by the de-correlation algorithm may still be modelled as a random process given by the sum of two independent contributions: the random prediction errors calculated to predict a hypothetically *noise-free* image, plus the prediction errors of the observation noise. By assuming a 3D separable Markov model for the noise, which may be both spatially and spectrally correlated, the latter term is easily lower-bounded as

$$\sigma_{e_n}^2(k) \;=\; \hat{\sigma}_n^2(k) \cdot (1 - \hat{\rho}_x^2)(1 - \hat{\rho}_y^2)(1 - \hat{\rho}_\lambda^2) \tag{32}$$

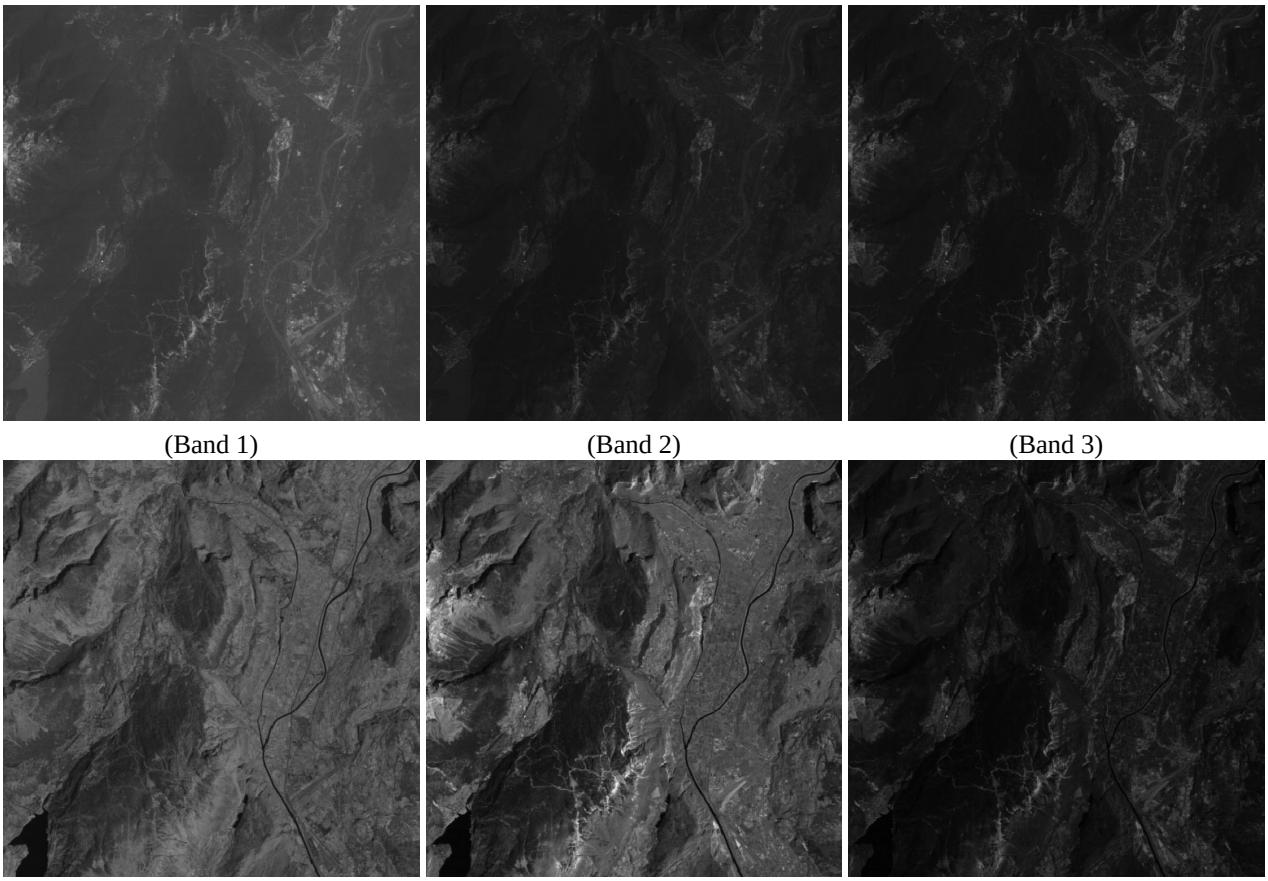
which is calculated from the estimated values of the noise parameters. Since $\sigma_n^2(k)$ varies with k, according to (32), $\sigma_{e_n}^2(k)$ varies with k as well, although the CC's are assumed to be constant. It is noteworthy that the adaptive predictor utilised in this work and thoroughly described in [6] attains a prediction very close to the optimum MMSE one. Furthermore, on homogeneous areas the predictor modifies itself in order to predict the correlated noise as much as possible. Thus, the approximation (32) is likely, at least when the CC's are not too small. In that case the noise variance is slightly propagated by the 3-D predictor, same as for white noise, because the adaptive coefficients may not be all identically null.

The *average* variance of the multi-spectral prediction errors on the kth band may be measured as:

$$\bar{\sigma}_{e_g}^2(k) \;=\; \bar{\sigma}_{e_f}^2(k) + \hat{\sigma}_n^2(k) \cdot (1-\hat{\rho}_x^2)(1-\hat{\rho}_y^2)(1-\hat{\rho}_\lambda^2) \tag{33}$$

in which $\bar{\sigma}_{e_f}^2(k)$ is the *average* variance of prediction errors on the kth band of the noise-free image.

By deriving $\bar{\sigma}_{e_f}^2(k)$ from (33) and replacing it into (21), the average entropy of the noise-free kth band, $R_f(k)$, is still given by (31), in which the entropy function of the kth band $\frac{1}{2}\log_2[c_f(k)]$ and is found by de-convolving the GG-modelled PDF of the prediction residuals of the kth band by a zero-mean Gaussian density with variance equal to $\hat{\sigma}_n^2(k) \cdot (1 - \hat{\rho}_x^2)(1 - \hat{\rho}_y^2)(1 - \hat{\rho}_\lambda^2)$.



(Band 1) (Band 2) (Band 3)

(Band 4) (Band 5) (Band 7)

Figure 3: 512×512 details from Landsat TM image of *Trento*.

4 EXPERIMENTAL RESULTS & DISCUSSION

4.1 Landsat TM Data

The test set comprises a Landsat TM image, with $8\ bit/pel$ and 6 bands out of the 7 available. In fact, the Band 6 (thermal infrared) was omitted mainly because of its poor resolution ($120m \times 120m$ vs. $30m \times 30m$) and scarce spectral correlation with the other bands. The test site, whose bands are portrayed in Fig. 3 as examples of visible and infrared observations, is part of the valley of the Adige river, near Trento, in Northern Italy. To achieve an optimal multi-spectral de-correlation, the different bands available should be arranged to form a sequence that maximizes the average cross-correlation between any couple of consecutive bands [4]. The optimal causal sequence was found to be $1 \rightarrow 2 \rightarrow 3 \rightarrow 7 \rightarrow 5 \rightarrow 4$. Band 6, when utilised, is always encoded stand-alone. Actually, a bidirectional, i.e. non-causal, prediction yields bit-rates that are slightly lower, on an average [10]. Non-causal prediction means that some of the middle bands of the causal sequence are first skipped and then bidirectionally predicted. The optimum bidirectional sequence, was found to be: $1 \rightarrow 3$, $1 \rightarrow 2 \leftarrow 3 \rightarrow 7 \rightarrow 4 \rightarrow 5 \leftarrow 7$. The difference in rate between causal and non-causal prediction, however, is moderate: the latter provides an improvement of four hundredths of bit for the optical bands and nearly eight hundredths for the infrared channels.

Table 1 reports the estimated parameters for the six bands, the first of which is encoded in *intra* mode, i.e. without reference to any other band. The PSNR is larger in the visible than in the infrared wave-lengths; instead the SNR values are all comparable. Bands 2 and 5 are the least informative, in the sense that they may be bidirectionally predicted to a larger extent than the others.

4.2 AVIRIS Data

The data set comprises two sequences of hyper-spectral images collected by the *Airborne Visible InfraRed Imaging Spectrometer* (AVIRIS) operated by NASA/JPL, in 1989 and 1997, respectively, on the *Cuprite Mine* test-site, in Nevada. Each sequence is constituted by 224 bands recorded at different wave-lengths in the range $380 \div 2500$ nm, with a spectral separation between two bands nominally of 10 nm. The size of each image is 614×512 pixels. The earlier sequence was acquired by the sensor with the 10-bit analogue-to-digital converter (ADC) originally introduced. In 1995 the sensor was

Table 1: Signal variance ($\bar{\sigma}_f^2$), noise parameters ($\hat{\sigma}_n$, $\hat{\sigma}_n^2$, PSNR, and SNR) and information parameters ($R(k)$, $R_n(k)$ and $R_f(k)$, in bit/pel) measured on the six $30\ m$ bands of the test TM image *Trento*.

Band	$\bar{\sigma}_f^2$	$\hat{\sigma}_n$	$\hat{\sigma}_n^2$	PSNR (dB)	SNR (dB)	$R(k)$	$R_n(k)$	$R_f(k)$
TM-1	130.49	1.33	1.77	34.86	18.67	3.63	2.72	2.95
TM-2	52.82	0.61	0.37	41.66	21.55	1.94	1.64	0.78
TM-3	106.72	0.70	0.49	40.44	23.38	2.76	2.22	1.77
TM-4	448.09	2.28	5.20	30.18	19.35	4.60	3.68	4.26
TM-5	769.79	2.36	5.57	29.88	21.41	3.79	3.77	0.72
TM-7	178.44	1.41	1.99	34.35	19.53	3.43	2.83	2.57
Avg.	281.06	1.60	2.56	33.26	20.41	3.36	2.81	2.17

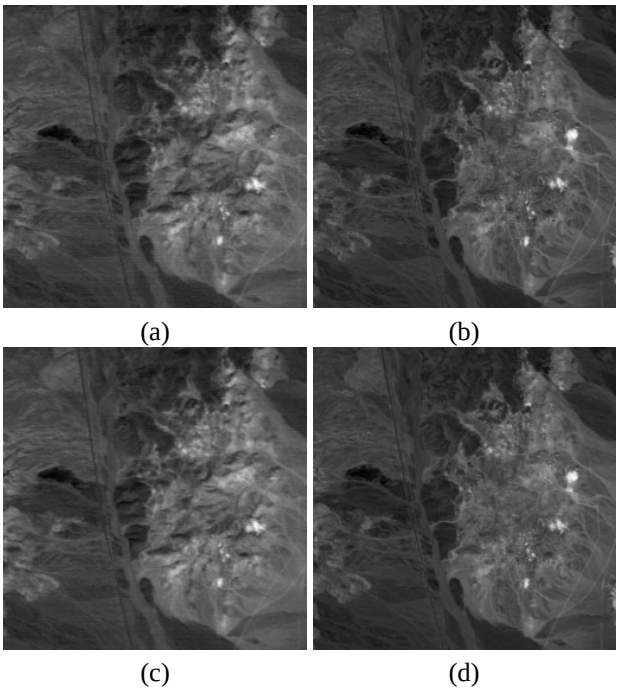


Figure 4: Two 256×256 details taken from two bands of the two AVIRIS test sequences portraying the *Cuprite Mine*: (a) band 48 (808 nm) from Cuprite'89; (b) band 48 (808 nm) from Cuprite'97; (c) band 60 (923 nm) from Cuprite'89; (d) band 60 (923 nm) from Cuprite'97.

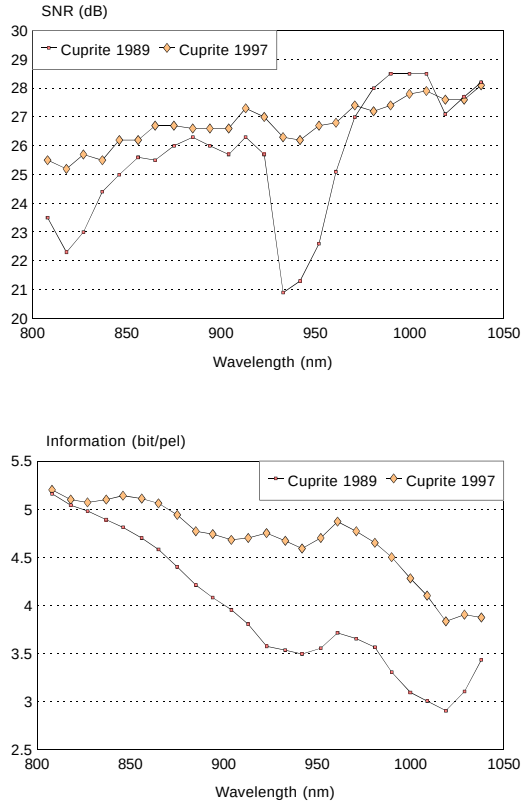


Figure 5: SNR and mutual information of 25 bands from the two AVIRIS sequences, varying with the wave-length.

Table 2: Measured CC's in horizontal, vertical, and spectral directions, for the two test AVIRIS sequences.

	$\hat{\rho}_x$	$\hat{\rho}_y$	$\hat{\rho}_\lambda$
Cuprite '89	0.65	0.25	0.35
Cuprite '97	0.75	0.75	0.80

equipped with a 12-bit ADC; so, the later sequence was digitised with a word-length larger by two bits. All the data have been radiometrically calibrated and expressed as radiance values.

Out of the four spectrometers covering the visible and infrared spectrum, the second one, covering the near-infrared wave-lengths, was analyzed. The goal was twofold: on one side to highlight how the SNR and the information vary with the wave-length; on the other side to emphasize different features between the two sequences, due to progresses in sensor setup and possible spatial/spectral post-processing of the raw data.

Two bands of each sequence portraying the detail of the mine are shown in Fig. 4. The more recent sequence appears significantly different from the earlier one. The CC's of the noise have been measured for the two sequences and are shown in Table 2. Values of CC relatively higher for the '97 sequence than for the '89 one

suggest possible post-processing undergone by the raw data both spatial and especially in the spectral direction.

Once the noise variances of each band have been estimated, the SNR is calculated. The trend in SNR to wave-length is shown in Fig. 5(top) for the AVIRIS data. The later sequence exhibits a trend somewhat regular and steady, slightly increasing with the wave-length. The negative peaks encountered in the earlier sequence and due to absorption phenomena (e.g. water vapour around 940 nm) appear mitigated in the later observation, possibly because of different atmospheric conditions (a small cloud was visible in the '89 image), and also to post-processing in the spectral direction which has the effect of "smearing" the noise, and thus the SNR, along the λ, thereby increasing its spectral correlation. An average gain around 1.5 dB is achieved.

Fig. 5(bottom) displays the spectral trend of mutual information for the two flight dates. Such an information was assessed from the bit-rate of Fig. 5(b), from the estimated noise variances used to calculate the SNR and from the CC's reported in Table 2, by means of the information-theoretic procedure. Although the trends of information are substantially similar to those of bit-rate, the values are different and all lower, since the bit-rate is diminished of an amount which is a highly non-linear function of the SNR and of the CC's. The two plots are closer for lower SNR (left side of the plots) and are far apart for higher SNR, as one can see on the right side. Furthermore, the plots of information are more *regular* than those of bit-rate, because the effects of the spectral-varying noise have been removed to yield a quantity which can be related to a true spectral information more easily than the bit-rate.

5 CONCLUDING REMARKS

A procedure for information-theoretic assessment of digitised images has been described. It relies on robust noise estimation and on parametric entropy modelling to calculate the mutual information between the noise-free band-limited analogue image and the digitised image. 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 of spectral radiance. The conclusions are that the six 30 m bands of Landsat TM scanning a vegetated area convey about 13 bits of useful information on the whole, thanks to a moderate spectral correlation and a relatively high SNR ($\approx$ 20 dB). The information of each individual band depends on the order in which the bands are encoded: the bidirectional prediction of the middle visible and infrared bands, i.e. B2 and B5, yields the minimum attainable code rate and, thus, the best entropy approximation. Results on AVIRIS images demonstrate progresses in the setup of the sensor during almost a decade (1989–1997), in that about 0.6 bit of usable information have been gained, thanks to an enhanced SNR ($\approx$ 1.5 dB), and two extra bits of the ADC. The proposed method yields an estimation of the information content of each spectral band; thus, it may be useful to assess the significance of different wave-lengths for a given test area and application.

ACKNOWLEDGEMENTS

This research was carried out under the financial support of the Italian Space Agency (ASI).

References

[1] F. O. Huck, C. L. Fales, R. Alter-Ganterberg, S. K. Park, and Z. Rahman, "Information-theoretic assessment of sampled imaging systems," *J. Optical Engin.*, vol. 38, no. 5, pp. 742–762, May 1999.

[2] S. K. Park and Z. Rahman, "Fidelity analysis of sampled imaging systems," *J. Optical Engin.*, vol. 38, no. 5, pp. 786–800, May 1999.

[3] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi, "Information-theoretic assessment of sampled hyperspectral imagers," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 7, pp. 1447–1458, July 2001.

[4] J. Wang, K. Zhang, and S. Tang, "Spectral and spatial decorrelation of Landsat-TM data for lossless compression," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, no. 5, pp. 1277–1285, Sep. 1995.

[5] R. E. Roger and M. C. Cavenor, "Lossless compression of AVIRIS images," *IEEE Trans. Image Processing*, vol. 5, no. 5, pp. 713–719, May 1996.

[6] 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.

[7] X. Wu and N. Memon, "Context-based lossless interband compression–Extending CALIC," *IEEE Trans. Image Processing*, vol. 9, no. 6, pp. 994–1001, Jun. 2000.

[8] R. E. Roger and J. F. Arnold, "Reversible image compression bounded by noise," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 1, pp. 19–24, Jan. 1994.

[9] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA, 1987.

[10] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi, "Estimating noise and information of multispectral imagery," *J. Optical Engin.*, vol. 41, no. 3, pp. 656–668, Mar. 2002.

[11] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[12] B. Aiazzi, L. Alparone, and S. Baronti, "Estimation based on entropy matching for generalized Gaussian PDF modeling," *IEEE Signal Processing Lett.*, vol. 6, no. 6, pp. 138–140, June 1999.