Virtually Lossless Compression of Astrophysical Images

Cinzia Lastri

Institute of Applied Physics "Nello Carrara," Italian National Research Council (IFAC-CNR), 64 via Panciatichi, 50127 Florence, Italy
Email: c.lastri@ifac.cnr.it

Bruno Aiazzi

Institute of Applied Physics "Nello Carrara," Italian National Research Council (IFAC-CNR), 64 via Panciatichi, 50127 Florence, Italy
Email: b.aiazzi@ifac.cnr.it

Luciano Alparone

Department of Electronics and Telecommunications (DET), University of Florence, 3 via di Santa Marta, 50139 Florence, Italy
Email: alparone@lci.det.unifi.it

Stefano Baronti

Institute of Applied Physics "Nello Carrara," Italian National Research Council (IFAC-CNR), 64 via Panciatichi, 50127 Florence, Italy
Email: s.baronti@ifac.cnr.it

We describe an image compression strategy potentially capable of preserving the scientific quality of astrophysical data, simultaneously allowing a consistent bandwidth reduction to be achieved. Unlike strictly *lossless* techniques, by which moderate compression ratios are attainable, and conventional *lossy* techniques, in which the mean square error of the decoded data is *globally* controlled by users, *near-lossless* methods are capable of *locally* constraining the maximum absolute error, based on user's requirements. An advanced lossless/near-lossless differential pulse code modulation (DPCM) scheme, recently introduced by the authors and relying on a causal spatial *prediction*, is adjusted to the specific characteristics of astrophysical image data (high radiometric resolution, generally low noise, etc.). The background noise is preliminarily estimated to drive the quantization stage for high quality, which is the primary concern in most of astrophysical applications. Extensive experimental results of lossless, near-lossless, and lossy compression of astrophysical images acquired by the Hubble space telescope show the advantages of the proposed method compared to standard techniques like JPEG-LS and JPEG2000. Eventually, the rationale of *virtually lossless* compression, that is, a noise-adjusted lossles/near-lossless compression, is highlighted and found to be in accordance with concepts well established for the astronomers' community.

Keywords and phrases: astrophysical images, differential pulse code modulation, lossless compression, near-lossless compression, noise estimation, statistical context modeling.

1. INTRODUCTION

The volume of astrophysical data that is acquired and exchanged among users, either scientists or not, is rapidly increasing. This is partly owing to large digitized sky surveys in the visible and near-infrared spectral intervals. These surveys are made possible by the development of digital imaging arrays, such as charge-coupled devices (CCDs). The size of digital arrays is also increasing, pushed by astronomical research's demands for more data in less time. Compression of such images can reduce the volume of data that it is necessary to store (a concern for large-scale sky surveys) and can shorten the time required to transmit images. The latter issue is useful for remote observing of or remote access to data archives [1].

Astronomical images have some rather unusual characteristics that make many existing image compression techniques ineffective [2]. A typical image consists of a nearly flat

background sprinkled with point sources and occasional extended sources. Depending on acquisition bandwidths and exposure times, images may be more or less noisy; in the former case, lossless compression is ineffective for transmission bandwidth reduction because the coding bit rate is lower-bounded by the entropy of the noise [3]. Furthermore, the images are usually subjected to stringent quantitative analysis, so any lossy compression method must be proven not to discard useful information, but should in principle discard only the noise [4].

Data compression methods can be classified as either reversible, that is, *lossless*, or irreversible, *lossy*, depending on whether the original data may be exactly reconstructed from the compressed data, or the decompressed data is not exactly the same as the original, because some *distortion* has been introduced by compression. Astronomers often insist that they can accept only lossless compression, in part because of conservatism, and in part because the familiar lossy compression methods sacrifice some information that is needed for accurate analysis of image data. In fact, for an astronomer a scientific frame is not simply a *scene* to be reproduced with a more or less high *fidelity*, but a 2D *measure* of a scalar field representing fluxes. Then, as for any other measure, random and systematic errors must be carefully assessed, quantified, and kept under strict control [5]. A common practice to achieve this goal, given the root mean square (RMS) of the noise introduced by the analog instrument, is that the step size of the uniform threshold quantizer (UTQ) is chosen accordingly, based on application requirements; for example, target detection, and the outcome quantization levels are transmitted without further loss. However, since all astronomical images contain noise, which is inherently incompressible, lossy compression methods may produce much better compression results and are thus worth being investigated, provided that a deep quantitative analysis of the impact of information loss on the scientific products expected from the observation is preliminarily carried out.

The classical image compression scheme consists of a decorrelator, followed by a quantizer and an entropy coding stage [6]. The decorrelator has the purpose of removing spatial redundancy; hence it must be tailored to the specific characteristics of the data to be compressed. Examples are orthogonal transforms, for example, discrete cosine transform (DCT) [6] and Mallat's discrete wavelet transform (DWT) [7], and differential pulse code modulation (DPCM) [6]. The quantizer introduces a distortion to allow a decrement in entropy rate to be achieved. Once an image has been decorrelated and possibly quantized, it is necessary to find a compact representation of its coefficients, which may be sparse. Thus, an entropy coding algorithm maps such coefficients into codewords, aiming at minimizing the average code length.

Decorrelation is crucial for compression. DWT calculated on the whole image (full-frame DWT) allows long-range correlation to be effectively removed, unlike DCT, in which full-frame processing leads to a spread of energy in the transformed plane, because DCT is not suitable for the analysis of nonstationary signals. Also computational issues make DCT usually applied not to the whole frame,

but to small blocks only, in which the assumption of stationarity approximately holds [6]. Hence, it fails in exploiting long-range correlation and can effectively remove only short-range correlation. The *dc* component that is encoded stand-alone (e.g., by spatial DPCM) is a fundamental drawback of first-generation transform coders [6]. The new standard JPEG2000 proposed by the Joint Photographic Experts Group [8] was devised to overcome such limitations, thereby leading to substantial benefits. The use of critically decimated decompositions, like Mallat's octave wavelet pyramid [7], is motivated by a twofold requirement: lack of redundancy, the reason for which undecimated decompositions, like the "à trous" wavelet transform, widely used in astrophysical image processing [9], are little suitable for data compression; orthogonality, thanks to which the variance of quantization errors in the transformed domain is preserved when the data is transformed back to the spatial domain.[1] Thus, the *mean square error* (MSE) can be easily controlled through the step sizes of quantizers. However, quantization errors in the transformed domain, which are likely to be uniformly distributed, at least if the step size is not greater than the RMS of the data, and are upper bounded in modulus by half of the step size, are spread by the inverse transformation and may yield heavy-tailed distributions, whose maximum absolute amplitude cannot be generally known a priori. Therefore, lossy transform-based encoders are unable to control the distortion but in the MSE sense, which means that in the lossy case relevant image features may be locally distorted by an unquantifiable extent.

Let $\{g(i,j)\}$, with $0 \leq g(i,j) \leq g_{fs}$, g_{fs} being the full scale, denote an integer-valued N-pixel digital image and $\{\tilde{g}(i,j)\}$ its distorted version, integer valued as well, achieved by compressing $\{g(i,j)\}$ and decoding the outcome bit stream. All values are intended to be expressed either in an unspecified unit, or simply in digital counts. Widely used distortion measurements are MSE, or squared L_2 distance between original and distorted image (L_2^2),

$$\text{MSE} = \frac{1}{N} \sum_i \sum_j \left[g(i,j) - \tilde{g}(i,j) \right]^2; \qquad (1)$$

maximum absolute distortion (MAD), or *peak error*, or L_∞-distance between original and distorted image,

$$\text{MAD} = \max_{i,j} \left\{ \left| g(i,j) - \tilde{g}(i,j) \right| \right\}; \qquad (2)$$

peak signal-to-noise ratio (PSNR),

$$\text{PSNR}_{(\text{dB})} = 10 \log_{10} \frac{g_{fs}^2}{\text{MSE} + 1/12} \qquad (3)$$

---

[1]If the transformation is not orthogonal, like the biorthogonal wavelet transform used by JPEG2000, MSE distortions coming from quantized subbands must be multiplied by constant coefficients depending on filters synthesizing each subband, before being summed together to yield the total distortion.

in which the MSE at denominator is incremented by the variance of the integer roundoff error, to handle the limit lossless case, when MSE = 0. Thus, PSNR will be upper bounded by $10 \log_{10}(12 \cdot g_{fs}^2)$, in the lossless case, to indicate that the signal detected by the sensor has been quantized before being reversibly compressed.

Noteworthy are those lossy methods that allow to settle "a priori" the maximum reconstruction error, not only on the whole, that is, *globally*, but also *locally*, that is, at each pixel location. Control of the maximum value of the absolute error, that is, of MAD, is capable to ensure constant quality throughout the reconstructed image. If the L_∞-error is constrained to be not greater than a user-defined value, the current definition of *near-lossless* compression, established for the medical community [10], applies.

The evaluation of the maximum allowable distortion is an open problem. In astrophysical applications, the data acquired from the instrument, after being preliminarily processed (preprocessed), for example, reduced and corrected for acquisition distortions, and calibrated, is usually postprocessed to extract information that may not be immediately available by visual inspection. Under this perspective, an attractive facility of near-lossless compression methods is that, if the user-defined L_∞-error is properly related to the RMS value of the background noise (assumed to be additive and signal-independent) the decompressed image, even though not identical to the original, may be *virtually lossless* [11]. Originally introduced for remote-sensing data compression, this term indicates not only that the decoded image is visually indistinguishable from the original, but also that possible outcomes of postprocessing (e.g., features extraction, target detection, data modeling, classification, etc.) are substantially unchanged from those calculated from the original data. Thus, the drawback of compression will be a small and predictable increment in the equivalent sensor's noisiness.

To conclude this section, we wish to recall that the introduction of data compression can alleviate bandwidth requirements at the price of a computational effort for encoding (images can be extremely large in size and processing power is generally limited on spaceborne platforms) and decoding, as well as of a possible loss of quality. The goal of this paper is investigating state-of-the-art and advanced compression algorithms from the viewpoint of their potential suitability to preserve the scientific quality of astrophysical imagery. To this purpose, a statistical analysis of the compression-induced distortion, when compression is lossy, will be carried out.

The remainder of this paper is organized as follows. Section 2 briefly reviews the theoretic fundamentals of differential pulse code modulation (DPCM) and state-of-the-art data compression methods. Section 3 describes an advanced DPCM encoder, recently introduced by the authors, whose characteristics of adaptivity make it suitable for astrophysical image compression. Section 4 reports extensive coding results on a large set of astrophysical images in a comparison with such compression standards as JPEG-LS and JPEG2000. Concluding remarks are drawn in Section 5.

2. LOSSLESS/NEAR-LOSSLESS IMAGE COMPRESSION

2.1. *Adaptive prediction*

Differential pulse code modulation (DPCM) schemes are indeed the sole algorithms suitable for lossless/near-lossless image compression, or more exactly for L_∞-constrained compression. DPCM basically consists of a decorrelation followed by entropy coding of the outcome residues, given as differences between true and *estimated* pixel values. If estimation of the current sample is carried out from past samples, according to the image scan fashion, DPCM is said to be spatially *causal* and the estimation is a *prediction*, that is, an extrapolation, driven by the previous samples. Conversely, estimation may be carried out *hierarchically*, that is, by increasing resolution: a low-resolution coarse image version is interpolated to a finer scale and differences between true and interpolated samples are progressively encoded. In this way DPCM is said to be spatially *noncausal*, or interpolation-based, and the outcome decoded bit stream resembles a *pyramid* [12], whose basis is the decompressed image. Both the causal and noncausal DPCM schemes may be L_∞-constrained. However, the former is not redundant, the number of residues being identical to that of image pixels, whereas the latter is redundant. Therefore, causal DPCM performs better than noncausal DPCM for medium/high bit rates, that is, close-to-lossless compression. Noncausal DPCM, which has the attractive characteristic of progressive decoding, is preferable for low bit rates, where its performance plots lie in the middle between those of JPEG and of JPEG2000.

Figure 1 outlines the flowcharts of causal DPCM *encoder* and *decoder*, featuring context modeling for entropy coding, which will be described in Section 2.2 . For the sake of clarity, notation is one-dimensional. The difference $e(n)$ between the current sample $g(n)$ and its estimation $\hat{g}(n)$ is quantized by the block labeled with Q to yield the quantized prediction error $e_\Delta(n)$, which is sent to the entropy coder (featuring context modeling in the example), which outputs the encoded prediction error $\varepsilon(n)$ and the array of data-dependent context thresholds Θ, as side information. At the same time, $e_\Delta(n)$ is inversely quantized (Q^{-1}) to the reconstructed prediction error $\tilde{e}(n)$, which is added to the output of predictor to yield the reconstructed sample $\tilde{g}(n)$. The latter is delayed by one sample, for the causality constraint, which states that the predicted value $\hat{g}(n)$ may not depend on $\tilde{g}(n)$, but only on $\tilde{g}(n-1)$, $\tilde{g}(n-2)$, and so on. The quantization noise feedback loop at the encoder allows the L_∞-error to be constrained, by forcing prediction at the encoder to be carried out from the same distorted samples that will be available at the decoder, where an identical predictor is placed. In a lossless implementation, $g(n)$ is integer valued, the output of predictor is rounded to integer as well, the quantizer block is missing together with the feedback loop, and the predictor is straightforwardly fed by the delayed sequence of $g(n)$.

The simplest way to design a predictor is to take a linear or nonlinear combination of the values of pixels lying within a causal neighborhood of the current pixel, that is,
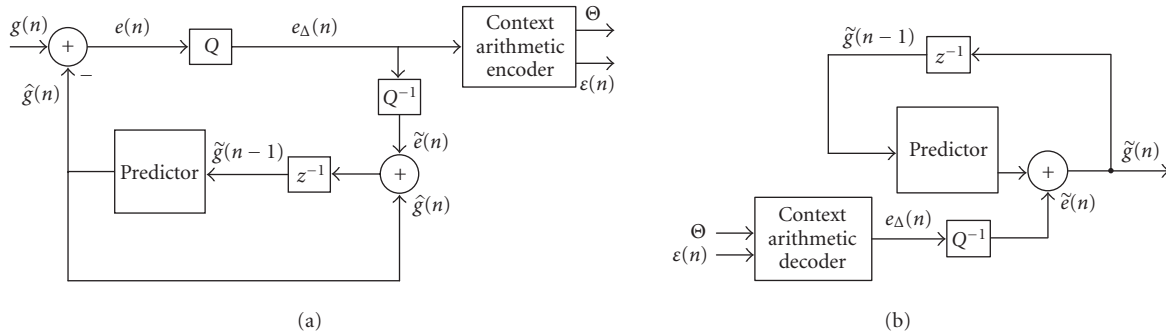
FIGURE 1: Flowchart of DPCM with quantization noise feedback loop at the encoder, suitable for error-bounded near-lossless compression: (a) encoder; (b) decoder.

surrounding the current pixel and such that they have been previously encountered along the image scan path, thereby representing *past* pixels. A linear combination, or regression, with fixed coefficients usually provides limited decorrelation. For better performance, the coefficients, whose number represents the *order* of prediction, may be calculated so as to yield *minimum* MSE (MMSE) over the whole image. Such coefficients are constant throughout an image, but change from an image to another image. The globally MMSE prediction, however, is optimal only for stationary signals. To overcome this drawback, two variations have been proposed: *adaptive* DPCM (ADPCM) [6], in which the coefficients of the MMSE predictor are continuously recalculated from the incoming new data at each pixel location on a subset of past pixels (the procedure is symmetrical at the decoder, so that the coefficients need not to be transmitted); *classified* DPCM [13], in which a preliminary training phase is aimed at recognizing some statistical classes of pixels and at calculating an MMSE predictor optimized for each class. Once such predictors are available, the most performing (in the MMSE sense) on a block of pixels may be *selected* to encode the current block [14]. Alternatively, predictors may be adaptively combined [15], also based on fuzzy-logic concepts [16], to attain an MMSE space-varying prediction. The two strategies of classified prediction will be referred to as *adaptive selection/combination of adaptive predictors* (ASAP/ACAP).

Eventually, we wish to remind the reader that a forerunner of the ACAP paradigm is the fuzzy 3D DPCM developed by some of the authors for lossless compression of multispectral and hyperspectral remotely sensed images [17]. In this case, the prototype MMSE spatial/spectral linear predictors constituting the dictionary were calculated on *clustered* data, an idea successfully developed in later works [18].

2.2. Context modeling

A notable feature of all advanced image compression methods [19] is statistical context modeling for entropy coding. The underlying rationale is that prediction errors should be similar to stationary white noise as much as possible. As a matter of fact, they are still spatially correlated to a certain extent and especially are non-stationary, which means that they exhibit space-varying variance. The better the prediction, however, the more noise-like prediction errors will be.

Following a trend established in the literature, first in the medical field [20], then for lossless coding in general [21, 22], and recently for *near-lossless* coding [23, 24], prediction errors are entropy coded by means of a *classified* implementation of an entropy coder, generally arithmetic [25] or Golomb-Rice [26]. For this purpose, they are arranged into a user-defined number of statistical classes based on the spatial *context* that can be a measure of *magnitude* or *activity* of past surrounding pixel values and/or prediction errors. If such classes are statistically discriminated, then the entropy of a *context-conditioned* model of prediction errors will be lower than that of a stationary memoryless model of the (decorrelated) source [27].

2.3. Review of standards and state-of-the-art methods

Considerable efforts have recently been spent on the development of lossless and near-lossless image compression techniques. The first specific standard has been the lossless version of JPEG [6], which relies on a set of linear predictors with fixed coefficients. A new standard, which provides also near-lossless compression, has been released under the name JPEG-LS [22]. It is based on an adaptive nonlinear prediction, potentially capable to fit contours, and exploits statistical context modeling of prediction errors followed by Golomb-Rice entropy coding. A similar context-based algorithm named CALIC has also been recently proposed [21]. The simple adaptive predictors used by JPEG-LS and CALIC, however, the *median adaptive predictor* (MAP) and the *gradient adjusted predictor* (GAP), have been empirically tailored to the average characteristics of gray-scale images. Thorough comparisons with methods following the ASAP and ACAP paradigms [14, 16] have revealed that their performance is limited and still far from the entropy bounds. In fact, the original 2D encoder following the ACAP paradigm [16] achieves lossless compression ratios 5% better than CALIC and 10% than JPEG-LS, on average. Although the 2D ASAP encoder [14] is slightly less performing than the former, its feature of real-time decoding is highly valuable in application contexts, since an image is usually encoded only once, but decoded many times. Furthermore, the crisp algorithm takes more advantage than the fuzzy one from a low noisiness of the data to compress.
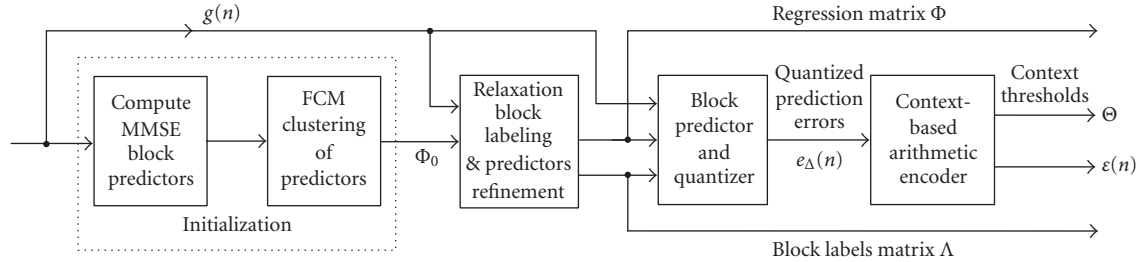
FIGURE 2: Flowchart of the *relaxation-labeled prediction encoder* (RLPE). The box marked as "block predictor and quantizer" includes a quantization noise feedback loop.

Eventually, the JPEG2000 image coding standard [8] incorporates a lossless mode, based on reversible integer wavelets, and is capable of providing a scalable bit stream that can be decoded from the lossy up to the lossless level. The possibility of defining *regions of interest* (ROI) is another facility of JPEG2000: for example, compression can be lossless inside ROIs and lossy elsewhere. However, despite its advanced and powerful facilities, JPEG2000 is an L_2-constrained encoder and thus not capable of providing near-lossless compression, except for the limit lossless case.

3. RELAXATION-LABELED PREDICTION ENCODER

The DPCM encoder proposed for astrophysical image compression [14] follows the ASAP paradigm, being based on a classified linear-regression prediction, with context-based arithmetic coding of prediction errors. The image is partitioned into blocks, typically 8×8, and an MMSE linear predictor is calculated for each block. Given a prefixed number of classes, a clustering algorithm produces an initial guess of as many classified predictors that are delivered to an iterative labeling procedure, which classifies pixel blocks simultaneously refining the associated predictors. In order to achieve reduction in bit rate within the constraint of a near-lossless compression, prediction errors are quantized with odd-valued step sizes, $\Delta = 2\delta + 1$, with δ denoting the induced L_∞-error. Quantized prediction errors are then arranged into activity classes based on the spatial *context*, that are entropy coded by means of arithmetic coding. Figure 2 shows the flowchart of the encoder. Besides encoded prediction errors $\varepsilon(n)$, the refined predictors are transmitted along with the label of each block and the set of thresholds defining the context classes for entropy coding.

3.1. Initialization

Patterns of pixel values occurring within the causal *neighborhood* of each pixel, also known as *prediction support*, reflect local image features, for example, edges, textures, and shadings. An efficient prediction should be capable of embodying and reflecting such features as much as possible. After preliminarily partitioning the input image into square blocks, for example, 8×8, a prediction support of size S (i.e., containing S samples) is set, and the S coefficients of an MMSE linear predictor are calculated for each block by means of a

least squares (LS) algorithm. Thus, a large number of predictors, each optimized for a single block, is produced.

The S coefficients of each predictor are arranged into an S-dimensional space. Since the coefficients of any predictor sum to one, all predictors lie on the hyperplane passing through the unit vectors of the coordinate axes. It can be noticed that statistically similar blocks exhibit similar predictors. Thus, the MMSE predictors calculated for each block are clustered on the hyperplane, instead of being spread.

A user-provided number M of representative predictors are identified by the *fuzzy-C-means* (FCM) clustering algorithm [28]. Such *dominant* predictors are calculated as *centroids* of as many clusters in the predictors space, according to a Euclidean metrics. Thus, an $S \times M$ matrix $\mathbf{\Phi}^{(0)} = \{\vec{\phi}_m^{(0)}, m = 1, \ldots, M\}$ containing the coefficients of the M predictors is produced. The superscript (0) highlights that such predictors are start-up values of an iterative refinement procedure.

3.2. Relaxation labeling and predictors refinement

Once M predictors have been found out through fuzzy clustering, they are used to initialize an iterative procedure in which image blocks are assigned to M classes and an optimized predictor is obtained for each class.

Step 0. Classify blocks based on their *mean square prediction error* (MSPE). The label of the predictor minimizing MSPE for a block is assigned to the block itself. This operation has the effect of partitioning the set of blocks into M classes that are best matched by the currently available predictors.

Step 1. Recalculate each of the M predictors from the data belonging to the blocks of each class. The new set of predictors is thus designed so as to minimize MSPE for the current block partition into M classes.

Step 2. Reclassify blocks: the label of the new predictor minimizing MSPE for a block is assigned to the block itself. This operation has the effect of moving some blocks from one class to another, thus repartitioning the set of blocks into M new classes that are best matched by the current predictors.

Step 3. Check convergence; if realized, stop; otherwise, go to Step 1.

3.3. Blockwise prediction and quantization

Once all blocks have been classified and labeled, together with the optimized predictors, the image is raster scanned and the M refined predictors are *activated* based on the classes of crossed blocks. Thus, each pixel value $g(n)$ belonging to one block of the partition that has been labeled to the mth class is predicted by using the mth predictor. Since $g(n)$ is integer, $\hat{g}(n)$ is rounded to integer as well and the outcome integer-valued prediction error, $e(n) = g(n) - \hat{g}(n)$, is uniformly quantized with a step size Δ as $e_\Delta(n)$ = round $[e(n)/\Delta]$ and delivered to the *context-coding* section.

The operation of inverse quantization $\tilde{e}(n) = e_\Delta(n) \cdot \Delta$ introduces an error, whose variance is approximately $(\Delta^2 - 1)/12$ (provided that Δ is lower than the RMS value of $e(n)$) and whose maximum absolute value is $L_\infty = \lfloor \Delta/2 \rfloor$. Since MSE is a quadratic function of Δ, odd-valued step sizes yield lower L_∞ than even sizes do. The step size Δ is set identical for all blocks, both to minimize L_∞ and to avoid blocking artifacts in reconstructed images.

3.4. Context-based arithmetic coding

Prediction errors are classified into a predefined number of statistically homogeneous classes based on the spatial *context*. A context function is defined and measured on prediction errors lying within a circular neighborhood of the current pixel, possibly larger than the prediction support, as the RMS value of prediction errors (RMSPE). Again, causality of neighborhood is necessary in order to make the same information available both at the encoder and at the decoder. At the former, the probability density function (PDF) of RMSPE is calculated and partitioned into a number of intervals chosen as equally populated to yield equiprobable contexts. This choice is motivated by the subsequent use for *residues* belonging to each class of *adaptive* arithmetic coding, which benefits from a number of data in each class large enough for training, which happens simultaneously with coding. The residue in each class is split into sign bit and magnitude. The former is strictly random and is coded as it stands, the latter exhibits a reduced variance in each class; thus, it may be coded with fewer bits than the original residue. From the PDF of context, $L - 1$ thresholds $\boldsymbol{\Theta} = \{\theta_l \in \mathbb{R}, l = 1, \ldots, L - 1\}$, that define the decision intervals of each class, are calculated. $\boldsymbol{\Theta}$, as well as $\boldsymbol{\Phi}$, is stored in the file header as *overhead*.

It is noteworthy that the context-coding procedure introduced by the authors [23] is independent of the particular method used to decorrelate the data. Unlike most of the schemes, for example, CALIC [21], in which context coding is embedded in the decorrelation, it can be applied to any DPCM scheme, either lossless or near-lossless, without adjustments for the near-lossless case [24], as a patch between decorrelation and entropy coding stages.

4. EXPERIMENTAL RESULTS

4.1. Dataset

All the images used in the following experiments have been acquired by the Wide Field and Planetary Camera 2 (WFPC2) and are available at http://archive.eso.org, courtesy of the European Southern Observatory (ESO).

The WFPC2 is a two-dimensional imaging photometer, whose field of view (FOV) is located at the center of focal plane of the Hubble Space Telescope (HST) and covers the spectral range between approximately 1150 Å and 10500 Å. The central portion of the $f/24$ beam coming from the Optical Telescope Assembly (OTA) is intercepted by a steerable pick-off mirror attached to the WFPC2 and is diverted through an open port entry into the instrument. The beam then passes through a shutter and interposable filters. A total of 48 spectral elements and polarizers are contained in an assembly of 12 filter wheels. The light then falls onto a shallow-angle, four-faceted pyramid, located at the aberrated OTA focus. Each face of the pyramid is a concave spherical surface, dividing the OTA image of the sky into four parts. After leaving the pyramid, each quarter of the full field of view is relayed by an optically flat mirror to a Cassegrain relay that forms a second field image on a charge-coupled device (CCD) of 800×800 pixels. Each of these four detectors is housed in a cell sealed by an MgF2 window, which is figured to serve as a field flattener. The optics of three of the four cameras —the Wide Field Cameras (WF2, WF3, WF4)— are essentially identical and produce a final focal ratio of $f/12.9$. The fourth camera, known as the Planetary Camera (PC or PC1), has a focal ratio of $f/28.3$. The WFPC2 simultaneously images a $150'' \times 150''$ L-shaped region with a spatial sampling of $0.1''$ per pixel, and a smaller $34'' \times 34''$ square field with $0.046''$ per pixel. Figure 3 shows the field of view of WFPC2 projected onto the sky. The four operational configurations of WFPC2 are described in Table 1.

The total system quantum efficiency (WFPC2+HST) ranges from 5% to 13% at visual wavelengths, and drops to $\approx 0.5\%$ in the far UV. Detection of faint targets will be limited by either the sky background (for broad filters) or by noise in the read-out electronics (for narrow and UV filters) with an RMS equivalent to 5 detected photons. Bright targets can cause saturation (more than 53 000 detected photons per pixel), but there are no related safety issues.

A large test set of images acquired by WFPC2 was used for lossless compression experiments. The subjects are Globular Cluster M30 (NGC7099), Irregular Galaxy Small Magellanic Cloud (SMC), and Ring Nebula (NGC6720). For each scene, several observations, differing by spectral filter and exposure time, were considered. All the images downloaded from the archive are *raw* data that have been neither *reduced* nor calibrated, have 12 bit dynamic range, and are packed in 16 bit words. Units are digital counts, which are converted into physical measure units once the calibration process is accomplished. A subset of images —one for each scene— on which lossy compression experiments have been carried out, is shown in Figure 4. Acquisition parameters are summarized in Table 2 and statistics, including the measured noise RMS value, in Table 3.

4.2. Lossless compression performance comparisons

The methods compared are RLPE with context modeling (CTX) and arithmetic coding (AC), RLPE without CTX and
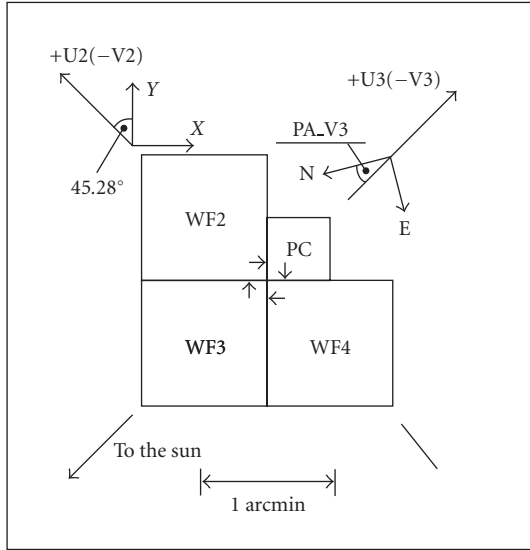
FIGURE 3: Field of view (FOV) of the Wide Field and Planetary Camera 2 (WFPC2) projected onto the sky. U2 and U3 axes are defined by the "nominal" Optical Telescope Assembly (OTA) axis, which is near the center of the FOV of WFPC2.

TABLE 1: Operational configurations of the Wide Field and Planetary Camera 2.

Camera	Pixels	Field of view	Scale	f/ratio
PC	800×800	$36'' \times 36''$	$0.0455''$/pixel	28.3
WF2,3,4	800×800	$80'' \times 80''$	$0.0966''$/pixel	12.9

with AC, RLPE with CCSDS-Rice coding [29] (including its own context model), JPEG-LS, lossless JPEG2000, and plain zero-order prediction with Rice coding (ZOP-Rice), standardized as baseline space encoder [30], though a more sophisticated predictor is left as an open concern. RLPE uses 8×8 blocks, 5 predictors (each with 4 coefficients, with values summing to one) refined with one iteration of relaxation labeling. Context modeling uses nine context classes, that is, eight thresholds, and context calculated on a circular causal neighborhood of radius three. We wish to point out that goal of the experimental section is comparing the *lossless*, *near-lossless*, and unconstrained *lossy* compression modalities, rather than providing a comprehensive comparison among compression algorithms, as it can be found, for example, in [16]. Most of state-of-the-art algorithms, like CALIC [21], are not available for image data having more than 8 bits per pixel, being developed for *multimedia* images rather than for scientific data.

Bit rates on disk including overhead and entropy coding are reported in Tables 4, 5, and 6 for a wide variety of observation of the three test scenes. A trend steady intra-table and inter-table shows that RLPE yields the lowest bit rates. Benefits stem from arithmetic coding and especially from context modeling. The coupling of RLPE with the CCSDS-Rice context-based entropy coding is slightly penalized with

respect to JPEG-LS, which exploits Golomb coding [31], together with a context model optimized to its nonlinear prediction. The baseline CCSDS scheme [30] is somewhat poorer, notwithstanding all predictors are relatively short (one-to-four-pixel neighborhoods), mainly because a *one-dimensional* predictor cannot adequately remove an intrinsically *two-dimensional* redundancy. Also, Rice entropy coding appears to be far less powerful than arithmetic coding. The advantage of the former over the latter for space application is that, at the time of its standardization, space-qualified hardware was already available for Rice coding, but not for arithmetic coding. Eventually, lossless JPEG2000, which is not based on DPCM, yields results somewhat poor on *Globular Cluster*, being superior to ZOP-Rice only; on *SMC* the average performance is identical to that of RLPE+Rice. However, JPEG2000 outperforms RLPE+Rice on *Ring Nebula* and closely approaches the performances of JPEG-LS. By watching Figure 4 the explanation of these trends is easily found. JPEG2000 is penalized with respect to advanced DPCM schemes on a dark background sprinkled by stars because of its compact-support oscillating functions, the decrement in performances against DPCM algorithms being directly related to the density of bright spots.

Given the intrinsically multispectral nature of the astrophysical data under concern, joint spectral and spatial decorrelation was investigated, by using the 3D version of RLPE [11]. It was found that the same strategy of adaptive prediction carried out from spectrally adjacent bands, unlike what happens for conventional remote-sensing data, is not rewarding in terms of compression performances, the average bit rate saving being less than one hundredth of bpp. This is not surprising, since astronomical bands, even if adjacent, are defined in order to select different physical emission mechanisms, with the consequence that images may be somewhat different.

Computationally speaking, ZOP-Rice is obviously the fastest scheme, closely followed by JPEG-LS, RLPE-Rice, plain RLPE without context, and full RLPE (with context and arithmetic coding). Table 7 reports encoding and decoding times for the three main schemes. Unlike the publicly available official versions of JPEG-LS and JPEG2000, the code of RLPE was written in C++, but was not optimized. Unpublished results of experiments specifically carried out on hyperspectral data have demonstrated that coding time might be reduced by, say, $4 \div 5$ times, by optimizing the algorithm flow and the code, as well as by training off line. As it appears, a notable feature of RLPE is its processing asymmetry: due to training of predictors and block classification, encoding is more onerous than decoding, whose complexity is essentially dictated by context and arithmetic decoding. This feature may be valuable for remote access to archives, since an image is coded only once (when it is placed in the archive), but decoded as many times as it is retrieved by users.

4.3. Near-lossless compression performance comparisons

Two DPCM algorithms having L_∞-constrained coding capability will be compared first. Figure 5 shows performances of
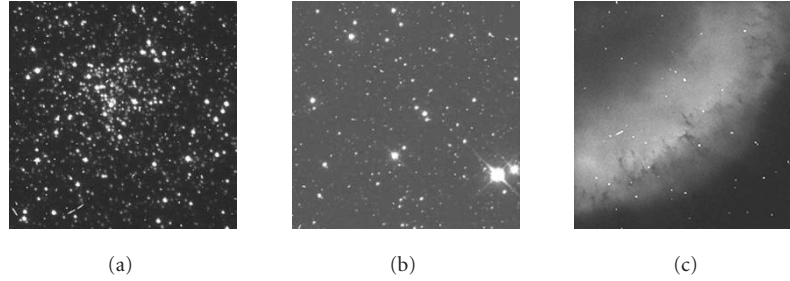
FIGURE 4: Details of size 256×256 taken from the three sample scenes. (a) Globular Cluster M30. (b) Irregular Galaxy SMC. (c) Ring Nebula.

TABLE 2: Subset of test images used for near-lossless compression experiments.

Name	u5fw0106r	u5wob405r	u531010er
Subject	Globular Cluster M30	Irregular Galaxy SMC	Ring Nebula
Acquisition date	31/05/1999	24/05/1999	16/10/1998
Center wavelength	334.4 nm	801.2 nm	501.2 nm
Bandwidth	37.4 nm	153.9 nm	2.7 nm
Exposure time	200 s	100 s	100 s

TABLE 3: Statistics of the three astrophysical images used for near-lossless compression experiments; all values are expressed as digital counts (squared, for variance), typical of uncalibrated data.

Name	Minimum	Maximum	Mean	Variance	Noise σ
u5fw0106r	304	4095	324.592	792.830	1.04
u5wob405r	304	4095	325.469	1502.920	1.29
u531010er	303	4095	326.403	517.191	0.80

RLPE+CTX+AC and JPEG-LS, carried out in terms of PSNR and MAD, on *Globular Cluster* and *SMC*. On both images RLPE attains a steady gain of about one dB on JPEG-LS, for bit rates greater than 1 bpp, slightly lower elsewhere. Since both methods are near lossless, errors of decoded values are likely to be uniformly distributed in $(-\delta, \delta)$, with the quantization step size $\Delta = 2\delta + 1$. Hence, MSE $= (\Delta^2 - 1)/12$ and from (3) the relationship between the MAD (δ) and PSNR will be

$$
\begin{aligned}
\text{PSNR}_{\text{(dB)}} = {} & 10 \log_{10} 12 + 20 \log_{10} g_{fs} \\
& - 20 \log_{10}(2\delta + 1)
\end{aligned}
\tag{4}
$$

if $g_{fs} = 4095$, (4) becomes

$$
\text{PSNR}_{\text{(dB)}} = 83 - 20 \log_{10}(2\delta + 1)
\tag{5}
$$

which is in accordance with the plots in Figure 5 , for bit rates greater than 0.5 bpp, that is, as long as quantization errors are independent of the data that is quantized.

Performance comparisons between RLPE and JPEG2000 have been carried out on the *Ring Nebula* test image and are shown in Figure 6. Unlike JPEG-LS, JPEG2000 is not L_∞-bounded, but L_2-bounded, which means that lossless compression is attainable —thanks to short 5/3 wavelet filters

[8]— but near-lossless compression is not. The consequence is that MADs larger and larger than those of RLPE are noticed as the bit rate decreases. The scale on ordinate was shrunk by a factor thirteen with respect to that of Figure 5, in order to accommodate the large range of MAD in the JPEG2000 plot. Besides being near-lossless, RLPE outperforms JPEG2000 also in PSNR. For rates higher than 1 bpp, the PSNR gain of RLPE over JPEG2000 is about 2 dB. Equivalently, RLPE saves 0.39 bpp in the reversible case, corresponding to 83 dB PSNR. As the bit rate decreases, this gain vanishes and the two plots cross each other at approximately 0.1 bpp. This effect is typical of all DPCM schemes and is due to the quantization noise feedback loop at the encoder.

The previously noticed error trends also reflect the visual quality of the decompressed images. Figure 7 shows *Ring Nebula* compressed at six different bit rates, including the lossless case, by RLPE (with CTX and AC) and JPEG2000. When the rate is high (0.913 bpp for RLPE and 0.912 for JPEG2000), the visual appearance of the two images is quite similar, notwithstanding the former exhibits MAD equal to one, the latter to 6, and the difference in PSNR is around 2 dB. Both compressed images are hardly distinguishable from the original (MAD = 0), even though JPEG2000 yields a perceivably smoother result. However, as the bit rate per pixel decreases, the JPEG2000 versions become smoother

TABLE 4: Bit rates on disk for lossless compression of different observations of the u5fw010 scene (Globular Cluster M30).

Name	RLPE+CTX+AC	RLPE+AC	JPEG-LS	RLPE+Rice	JPEG2000	ZOP-Rice
u5fw0101r	2.56	2.70	2.76	2.93	2.98	3.19
u5fw0102r	2.06	2.09	2.22	2.38	2.42	2.67
u5fw0103r	2.56	2.71	2.76	2.93	2.99	3.19
u5fw0104r	2.63	2.79	2.81	2.98	3.03	3.23
u5fw0105r	2.61	2.76	2.82	2.96	3.01	3.24
u5fw0106r	2.41	2.50	2.63	2.80	2.84	3.04
u5fw0107r	2.55	2.68	2.76	2.93	2.98	3.19
u5fw0108r	2.38	2.47	2.59	2.77	2.82	2.99
u5fw0109r	3.36	3.61	3.49	3.78	2.83	4.00
u5fw010ar	2.00	2.03	2.17	2.33	2.37	2.63
u5fw010br	2.38	2.47	2.59	2.75	2.81	3.00
u5fw010cr	2.38	2.47	2.59	2.77	2.83	3.00
Average	2.49	2.61	2.68	2.86	2.91	3.11

TABLE 5: Bit rates on disk for lossless compression of observations of u5wob40 scene (SMC).

Name	RLPE+CTX+AC	RLPE+AC	JPEG-LS	RLPE+Rice	JPEG2000	ZOP-Rice
u5wob401r	2.32	2.38	2.51	2.66	2.72	2.93
u5wob402r	3.11	3.27	3.31	3.56	3.52	3.77
u5wob403r	2.09	2.12	2.25	2.42	2.46	2.70
u5wob404r	2.84	2.96	3.10	3.33	3.31	3.52
u5wob405r	2.41	2.48	2.61	2.76	2.82	3.02
u5wob406r	3.12	3.27	3.33	3.55	3.54	3.76
u5wob407r	1.91	1.91	2.04	2.24	2.25	2.53
u5wob408r	2.02	2.03	2.18	2.36	2.39	2.63
u5wob409r	2.77	2.87	3.07	3.28	3.28	3.42
u5wob40ar	2.79	2.88	3.09	3.33	3.30	3.45
u5wob40br	2.80	2.89	3.06	3.30	3.27	3.40
Average	2.56	2.64	2.77	2.98	2.98	3.19

TABLE 6: Bit rates on disk for lossless compression of observations of u531010 scene (Ring Nebula).

Name	RLPE+CTX+AC	RLPE+AC	JPEG-LS	JPEG2000	RLPE+Rice	ZOP-Rice
u5310109r	2.23	2.27	2.47	2.51	2.64	2.87
u531010am	2.34	2.39	2.63	2.68	2.79	2.99
u531010br	2.36	2.53	2.64	2.67	2.68	2.98
u531010cr	2.35	2.52	2.63	2.67	2.67	2.97
u531010dr	2.35	2.52	2.62	2.65	2.67	2.97
u531010er	2.24	2.36	2.49	2.56	2.57	2.87
u531010fr	2.24	2.36	2.49	2.56	2.56	2.87
u531010gr	2.24	2.36	2.49	2.56	2.57	2.87
Average	2.29	2.41	2.55	2.60	2.64	2.92

TABLE 7: Computing times (on 1.8 GHz Pentium PC) of RLPE+CTX+AC, JPEG-LS, and JPEG2000 for an 800 × 3200, 12 bit frame.

Processing	RLPE+CTX-AC	JPEG-LS	JPEG2000
Encoder	25 s	0.05 s	3.6 s
Decoder	1 s	0.03 s	4 s

and smoother, mainly because MAD increases from 6 to 53, since the difference in PSNR vanishes at 0.111 bpp.

The grainy appearance of the nebula completely disappears, replaced by an artificially uniform smoothness. At the lowest bit rate (0.111 bpp), JPEG2000 yields a result definitely unacceptable: all fine details have been removed and ringing artifacts appear around stars. Conversely, in the near-lossless RLPE-compressed versions, the grainy appearance of the nebula becomes coarser and coarser as the bit rate decreases. Also at the lowest bit rate (0.111 bpp), even if a striping distortion markedly appears in the dark background, the image has still a certain fidelity to the original.
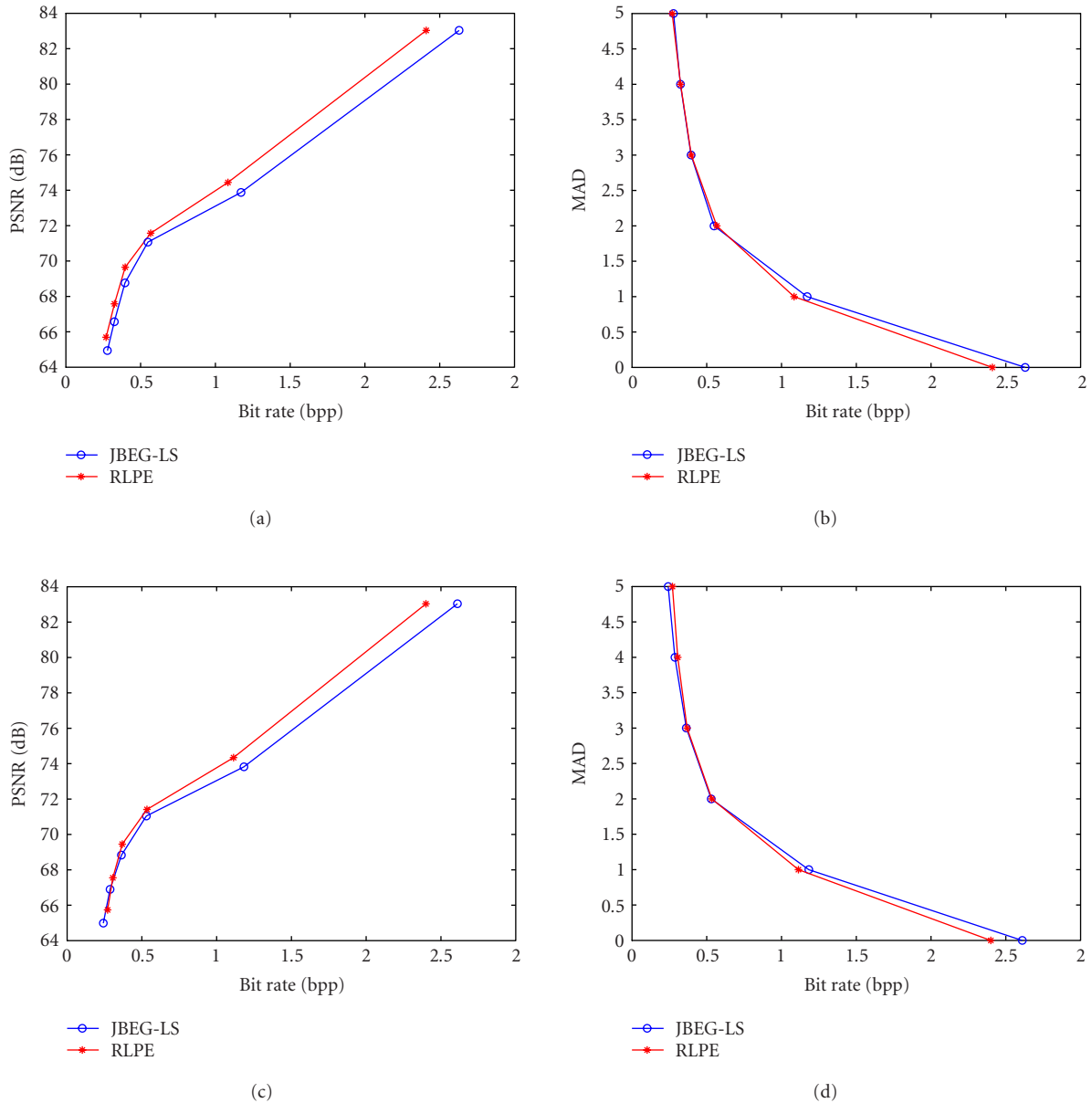
FIGURE 5: Lossy compression performances of RLPE+CTX+AC and JPEG-LS. Globular Cluster: (a) PSNR versus bit rate, (b) MAD versus bit rate. SMC: (c) PSNR versus bit rate, (d) MAD versus bit rate.

To provide a deeper insight into the difference between near-lossless and lossy compressions, or better between L_∞-bounded and L_2-bounded compressions, the amplitude distributions of compression-induced errors have been plotted in Figure 8 for RLPE and JPEG2000 at high and low bit rates. The distribution of errors introduced by RLPE is practically uniform at high rate, slightly decaying at low rates, because quantization errors are no longer independent of prediction errors that are quantized. In both cases, however, their distribution has no heavy tails. Instead, JPEG2000 exhibits tails more pronounced than those of a Gaussian function. Logarithmic scale on the y-axis is used throughout, for displaying convenience.

Second-order statistics of the compression-induced errors have been investigated as well. Figure 9 shows the original *Ring Nebula* and the pixel map of errors introduced by RLPE and by JPEG2000 at the same bit rate of 0.111 bpp, corresponding to approximately identical 66 dB PSNR. Displayed errors have been linearly stretched and biased to avoid negative values. While the distortion introduced by RLPE is substantially similar to pure noise, especially in the body of the nebula, the error map produced by JPEG2000 contains plenty of fine spatial details (including edges of stars) that have been destroyed by compression. An analysis of the spatial correlation coefficient (CC) of each error map reveals that RLPE yields CC equal to 0.19 (average between row
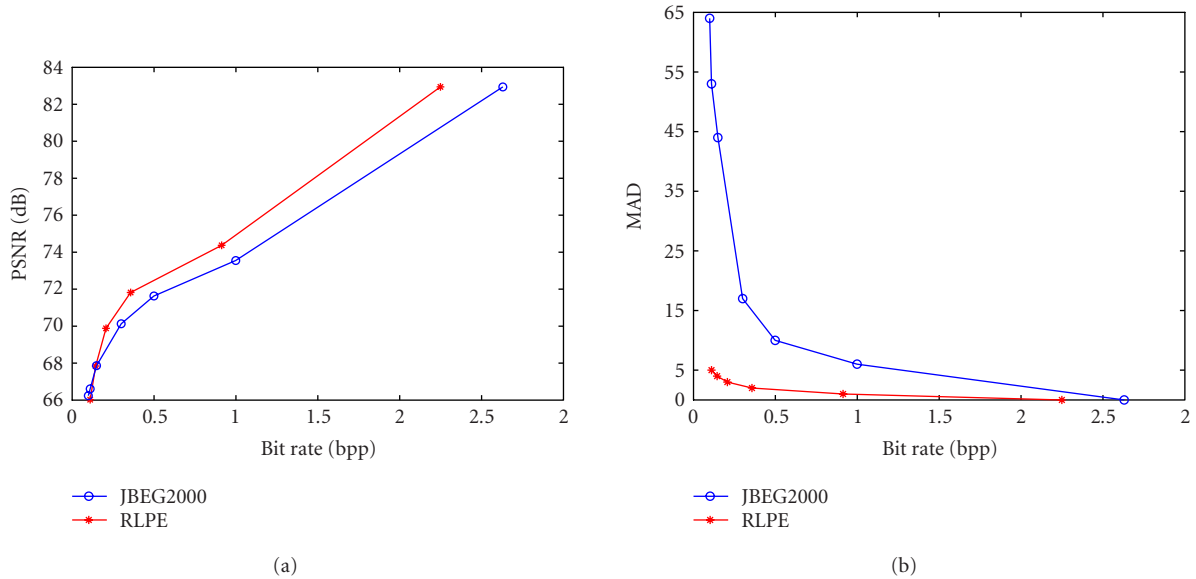
FIGURE 6: Lossy compression performances of RLPE+CTX+AC and JPEG2000. Ring Nebula: (a) PSNR versus bit rate. (b) MAD versus bit rate.

and column directions); conversely, the average CC of the JPEG2000 error map is 0.28, thereby revealing that what has been removed by compression is more likely to be a correlated *signal*.

4.4. *Virtually lossless compression*

The analysis reported has pointed out that quality evaluation of compressed astrophysical images cannot rely on PSNR distortion measurements only. We notice that the wavelet-based JPEG2000 algorithm achieves the effect of progressively "denoising" the image as the target compression ratio increases. This fact is not surprising, since it has been demonstrated that suppression of small wavelet coefficients, which happens because of quantization, yields a powerful method for image denoising, established also in the field of astrophysical image processing [32]. Image denoising may also become the key to compression of astronomical images [4], when the bottleneck of a very low bit rate imposes a reduction in image entropy, selectively obtained by denoising the background only. However, what may appear as "noise" is likely to be informative to an astrophysicist. Therefore, the data may become little useful once they have been compressed by means of an otherwise advanced L_2-bounded method like JPEG2000.

On the contrary, near-lossless methods, like JPEG-LS and RLPE seem to be more suitable than JPEG2000 for locally preserving even subtle objects of variable coarseness. The main reason of that is the *quantization noise-shaping* effect achieved by L_∞-bounded image encoders, like those based on DPCM. Indeed, noise modeling was found to be the key to compression of astrophysical images [33].

The term *virtually lossless* compression, which motivates the present paper, is now discussed in greater detail. It indicates that the distortion introduced by compression should

appear as an additional amount of noise, being uncorrelated and having space-invariant first-order statistics such that the overall probability density function (PDF) of the noise corrupting the decompressed data, that is, intrinsic noise plus compression-induced noise, closely matches the noise PDF of the original data. This requirement is trivially fulfilled if compression is lossless, but may also hold if the difference between uncompressed and decompressed data exhibits a peaked and narrow PDF without tails, as it happens for near-lossless techniques, whenever the user defined MAD is sufficiently smaller than the standard deviation σ_n of the background noise. Both MAD and σ_n are intended to be expressed either in physical units, for calibrated data, or as digital counts otherwise. Therefore, noise modeling and estimation from the uncompressed data becomes a major task to accomplish a virtually lossless compression [11]. The underlying assumption is that the dependence of the noise on the signal is null, or weak. However, signal independence of the noise may not strictly hold for astronomical images, especially for weak signals, dominated by shot noise. This further uncertainty in the noise model may be encompassed by imposing a margin on the relationship between target MAD and RMS value of background noise.

In the present case, the noise standard deviation σ_n of the three test images, whose statistics are reported in Table 3, was measured by means of the scatterplot-based method described in [34, 35], and found to be $\sigma_n = 1.04$, $\sigma_n = 1.29$, and $\sigma_n = 0.80$, for *Globular Cluster*, *SMC*, and *Ring Nebula*, respectively. Near-lossless compression is crucial for *Ring Nebula*, as it visually appears from Figure 7. In fact, near-lossless compression with MAD $= \delta = 1$ (i.e., quantization step size $\Delta = 2\delta + 1 = 3$) would yield an RMS distortion $\epsilon = \sqrt{2/3} \approx 0.82$, slightly greater than the noise RMS value $\sigma_n = 0.80$, which would have the effect of increasing
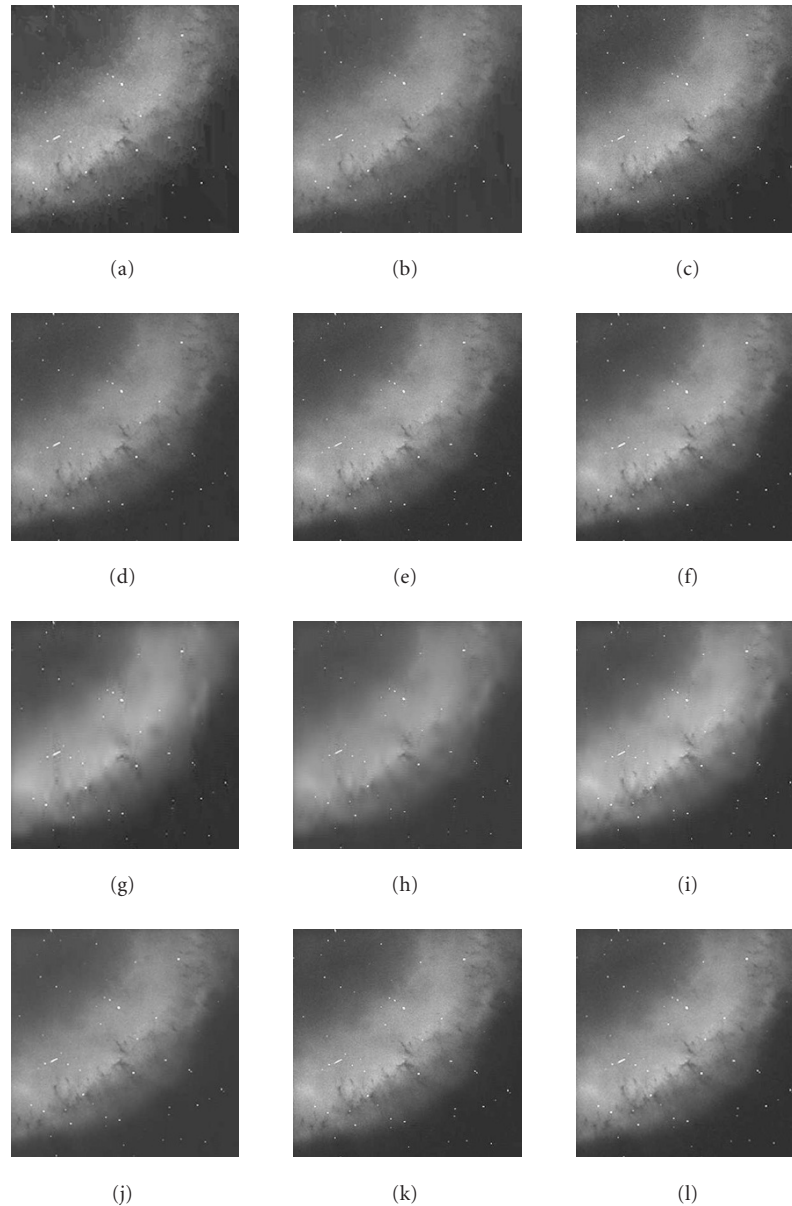
Figure 7: Compressed 256×256 details taken from Ring Nebula obtained at the same bit rates per pixel bpp, with exception of lossless versions included for both RLPE and JPEG2000 (JP2K). RLPE for (a) 0.111 bpp, MAD=5; (b) 0.179 bpp, MAD=4; (c) 0.207 bpp, MAD=3; (d) 0.447 bpp, MAD=2; (e) 0.913 bpp, MAD=1; (f) 2.24 bpp, MAD=0. JP2K for (g) 0.111 bpp, MAD=53; (h) 0.179 bpp, MAD 32; (i) 0.207 bpp, MAD =20; (j) 0.447 bpp, MAD =10; (k) 0.912 bpp, MAD =6; (l) 2.63 bpp, MAD =0.

by a factor greater than $\sqrt{2}$, after decompression. Equivalently, the intrinsic SNR of the uncompressed image would be decremented by 3 dB after compression/decompression. In this specific case, *virtually lossless* compression should better coincide with *lossless* compression. Near-lossless compression of *Ring Nebula* with MAD equal to one is unable to retain the quality of the data, because the compression-induced MSE is not one order of magnitude lower than σ_n^2, as it would be recommended for virtually lossless compression. However, when the extremely concentrated error PDF, produced by RLPE when $\delta = 1$ and shown as first entry in

Figure 8, is convolved with the intrinsic noise PDF, assumed to be tailed, the overall PDF will be approximately unchanged in shape, even if doubled in variance. This behavior explains why some of the RLPE-compressed versions of *Ring Nebula* are more similar to the original than to the corresponding JPEG2000 versions. The reason is that *tails* in the error PDF may give rise to, or suppress, local "noise" patterns, whose presence, or absence, is unlikely to be found in the uncompressed image.

The rationale of virtually lossless compression can be summarized by the following protocol. Measure the noise
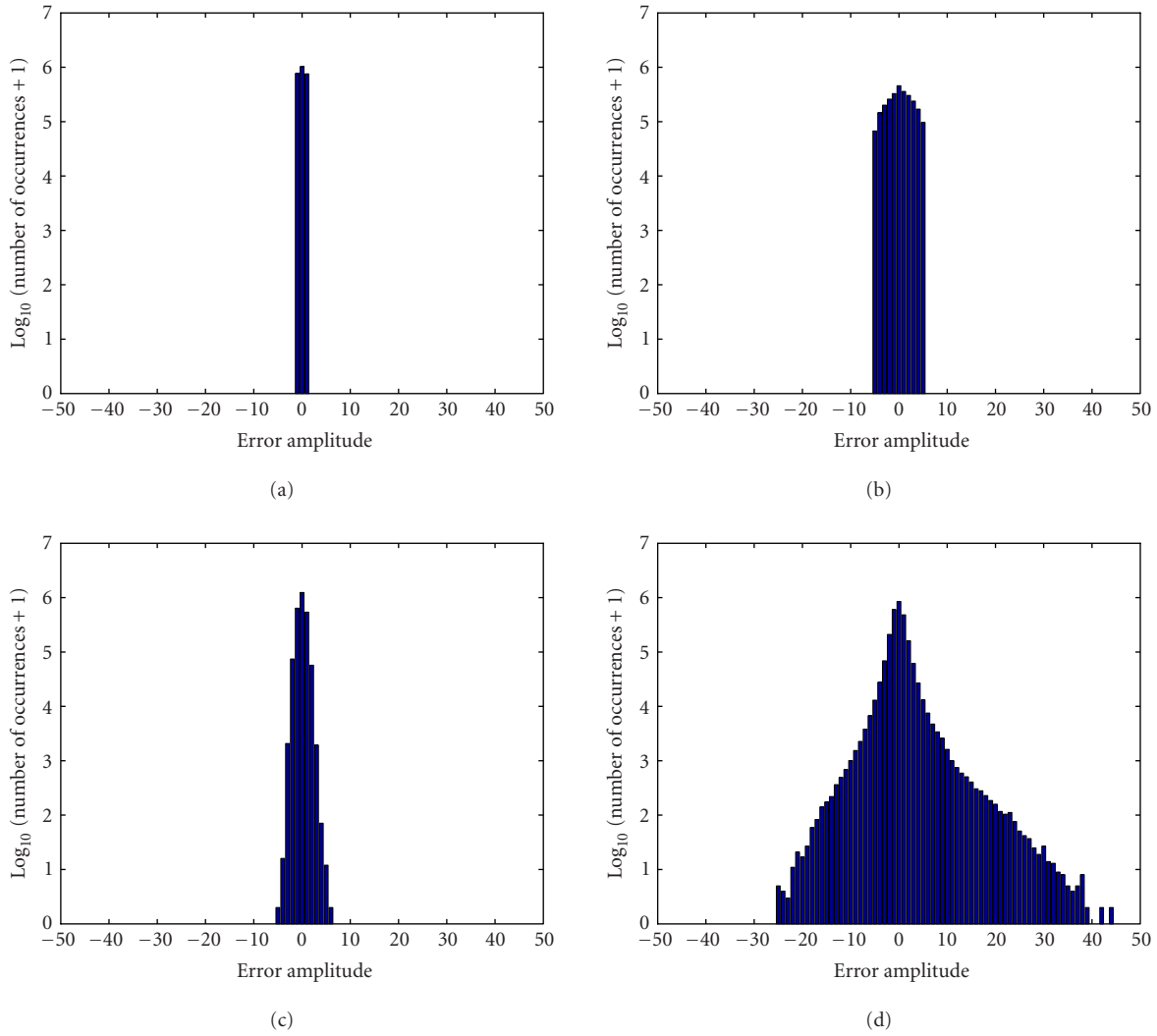
FIGURE 8: Reconstruction error distributions for near-lossless/lossy coding of *Ring Nebula*. RLPE for bit rate of (a) 0.913 bpp, (b) 0.111 bpp. JPEG2000 for bit rate of (c) 0.912 bpp, (d) 0.111 bpp.
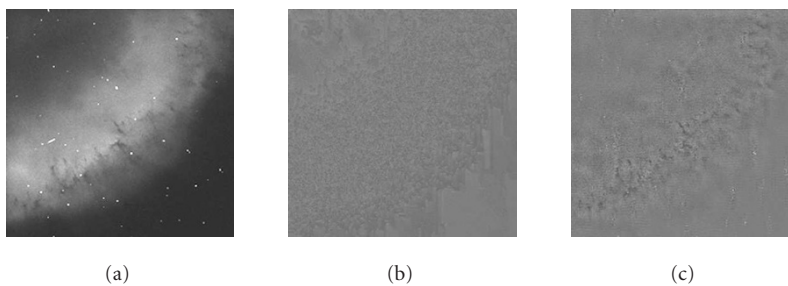

FIGURE 9: Original 256×256 details of Ring Nebula: (a) pixel differences with 0.111 bpp, (b) RLPE decompressed version, and (c) JPEG2000 decompressed version.

RMS, σ_n; if $\sigma_n < 1$, lossless compression is mandatory. Otherwise, if $1 \leq \sigma_n < 3$, near-lossless compression with MAD = 1 (hence, $\Delta = 3$) might be attempted. For $3 \leq \sigma_n < 5$, com-pression with MAD = 1 is recommended, to avoid wasting bits encoding the noise. In the general case, the relationship between MAD and σ_n, also including a margin of

approximately one dB, is

$$\text{MAD} = \left\lfloor \max\left\{ 0, \frac{\sigma_n - 1}{2} \right\} \right\rfloor. \tag{6}$$

This protocol is substantially in accordance with the results reported by Maris et al. [5]. The main difference is operative. In [5] the step size of the UTQ to quantize the analog signal is designed in such a way that compression *must* be lossless thereafter. In the present case, the signal may have been previously quantized based on different requirement; afterwards a check on the noise is made to decide whether lossless compression is really necessary, or near-lossless compression could be used instead without penalty, being de facto *virtually lossless*. Depending on the application context and the type of data, the relationship (6) may also be relaxed, for example, by imposing that the ratio MSE(noise)/MSE(compression) is greater than, say, 3 dB, instead of the $10 \div 11$ dB, given by (6).

5. CONCLUDING REMARKS

The key to achieve a compression preserving the scientific quality of the data, for either astrophysical or remote-sensing applications, is represented by the following twofold recommendation: (1) absence of *tails* in the PDF of the error between uncompressed and decompressed image, in order to maximize the ratio $\sqrt{\text{MSE}}$/MAD, that is, RMSE/MAD, or equivalently to minimize MAD for a given RMSE; (2) MSE lower by one order of magnitude than the variance of background noise σ_n^2. Near-lossless methods are capable of fulfilling such requirements, provided that the quantization step size Δ is chosen as an odd integer such that $\Delta \approx \sigma_n$. If the data is intrinsically little noisy, the protocol may lead to the direct use of lossless compression, that is, $\Delta = 1$, to obtain what has been denoted as *virtually lossless* compression.

ACKNOWLEDGMENT

REFERENCES

[1] F. Murtagh, J.-L. Starck, and M. Louys, "Distributed visual information management in astronomy," *IEEE Computing in Science & Engineering*, vol. 4, no. 6, pp. 14–23, 2002.

[2] M. Louys, J.-L. Starck, S. Mei, F. Bonnarel, and F. Murtagh, "Astronomical image compression," *Astronomy and Astrophysics Supplement Series*, vol. 136, no. 3, pp. 579–590, 1999.

[3] 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, 1994.

[4] J.-L. Starck, F. Murtagh, B. Pirenne, and M. Albrecht, "Astronomical image compression based on noise suppression," *Publications of the Astronomical Society of the Pacific*, vol. 108, pp. 446–455, 1996.

[5] M. Maris, D. Maino, C. Burigana, A. Mennella, M. Bersanelli, and F. Pasian, "The effect of signal digitisation in CMB experiments," *Astronomy & Astrophysics*, vol. 414, no. 2, pp. 777–794, 2004.

[6] K. K. Rao and J. J. Hwang, *Techniques and Standards for Image, Video, and Audio Coding*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1996.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 1998.

[8] ISO/IEC JTC 1/SC 29/WG1, ISO/IEC FCD 15444-1: Information technology-JPEG 2000 image coding system: Core coding system [WG 1 N 1646], March 2000.

[9] J.-L. Starck and F. Murtagh, "Astronomical image and signal processing: looking at noise, information and scale," *IEEE Signal Processing Mag.*, vol. 18, no. 2, pp. 30–40, 2001.

[10] K. Chen and T. V. Ramabadran, "Near-lossless compression of medical images through entropy-coded DPCM," *IEEE Trans. Med. Imag.*, vol. 13, no. 3, pp. 538–548, 1994.

[11] 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, 2001.

[12] B. Aiazzi, L. Alparone, S. Baronti, and F. Lotti, "Lossless image compression by quantization feedback in a content-driven enhanced Laplacian pyramid," *IEEE Trans. Image Processing*, vol. 6, no. 6, pp. 831–843, 1997.

[13] F. Golchin and K. K. Paliwal, "Classified adaptive prediction and entropy coding for lossless coding of images," in *Proc. IEEE International Conference on Image Processing (ICIP '97)*, vol. 3, pp. 110–113, Santa Barbara, Calif, USA, October 1997.

[14] B. Aiazzi, L. Alparone, and S. Baronti, "Near-lossless image compression by relaxation-labelled prediction," *Signal Processing*, vol. 82, no. 11, pp. 1619–1631, 2002.

[15] G. Dong, H. Ye, and L. W. Cahill, "Adaptive combination of linear predictors for lossless image compression," *IEE Proceedings-Science, Measurement and Technology*, vol. 147, no. 6, pp. 414–419, 2000.

[16] B. Aiazzi, L. Alparone, and S. Baronti, "Fuzzy logic-based matching pursuits for lossless predictive coding of still images," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 4, pp. 473–483, 2002.

[17] 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, 1999.

[18] 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, 2003.

[19] B. Carpentieri, M. J. Weinberger, and G. Seroussi, "Lossless compression of continuous-tone images," *Proc. IEEE*, vol. 88, no. 11, pp. 1797–1809, 2000.

[20] T. V. Ramabadran and K. Chen, "The use of contextual information in the reversible compression of medical images," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 185–195, 1992.

[21] X. Wu and N. D. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Commun.*, vol. 45, no. 4, pp. 437–444, 1997.

[22] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS," *IEEE Trans. Image Processing*, vol. 9, no. 8, pp. 1309–1324, 2000.

[23] B. Aiazzi, L. Alparone, and S. Baronti, "Context modeling for near-lossless image coding," *IEEE Signal Processing Lett.*, vol. 9, no. 3, pp. 77–80, 2002.

[24] X. Wu and P. Bao, "L_∞ constrained high-fidelity image compression via adaptive context modeling," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 536–542, 2000.

[25] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Communications of the ACM*, vol. 30, no. 6, pp. 520–540, 1987.

[26] R. F. Rice and J. R. Plaunt, "Adaptive variable-length coding for efficient compression of spacecraft television data," *IEEE Trans. Commun. Technol.*, vol. 19, no. 6, pp. 889–897, 1971.

[27] M. J. Weinberger, J. J. Rissanen, and R. B. Arps, "Applications of universal context modeling to lossless compression of gray-scale images," *IEEE Trans. Image Processing*, vol. 5, no. 4, pp. 575–586, 1996.

[28] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*, Plenum Press, New York, NY, USA, 1981.

[29] Consultative Committee for Space Data Systems, Washington, DC, USA: CCSDS, Lossless Data Compression: Report Concerning Space Data Systems Standards, (Green Book), May 1997.

[30] ISO TC 20/SC 13/ICS 49.140, 15887-2000, Space data and information transfer systems—Data systems—Lossless data compression, 2000.

[31] S. W. Golomb, "Run-length encodings," *IEEE Trans. Inform. Theory*, vol. 12, no. 3, pp. 399–401, 1966.

[32] J.-L. Starck and F. Murtagh, "Image restoration with noise suppression using the wavelet transform," *Astronomy and Astrophysics*, vol. 288, no. 1, pp. 342–348, 1994.

[33] F. Murtagh, J.-L. Starck, and M. Louys, "Very-high-quality image compression based on noise modeling," *International Journal of Imaging Systems and Technology*, vol. 9, no. 1, pp. 38–45, 1998.

[34] 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, 2001.

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

Cinzia Lastri was born in Florence, Italy in 1976. She received the Laurea degree in telecommunication engineering from the University of Florence, Florence, Italy, in 2002 with a thesis on "Reversible compression of remote sensing hyperspectral data acquired from satellite." Since 2002, she has been working with the support of CNR at the Istituto di Fisica Applicata "Nello Carrara" (IFAC-CNR) in Florence, in the framework of the ESA contract ESTEC 15662/01/NL/MM "Advanced methods for lossless compression of hyperspectral data." In 2004 she applied for the Ph.D. degree at the University of Florence on the theme "Digital image processing and compression of multispectral and hyperspectral data" for transmission and archiving of remote sensing data. She has coauthored three chapters in international books on data compression and more than ten works published in proceedings of international conferences.

Bruno Aiazzi was born in Borgo San Lorenzo, Florence, Italy, in 1961. He received the Laurea degree in electronic engineering from the University of Florence, Florence, Italy, in 1991. In 1992, he won a fellowship on *digital image compression for broadband communications networks*, supported by the National Research Council (CNR), Italy. After several years as a Research Fellow and Research Assistant, since 2001, he has been a Researcher with IFAC-CNR (formerly IROE-CNR), Florence, Italy, where he currently participates in research activities concerning image quality definition and measurement, advanced methods for lossless and near-lossless remote sensing data compression, multiresolution image analysis and data fusion, and SAR image analysis and classification. He has been involved in several international research projects with ESA, CNES, and ASI on advanced data compression and image fusion algorithms for environmental applications. He is responsible for SAR image analysis and classification in a project funded by the Italian Space Agency. He has coauthored over 20 papers published in international peer-reviewed journals.

Luciano Alparone obtained the Laurea degree in electronic engineering "summa cum laude" and the Ph.D. degree from the University of Florence, Florence, Italy, in 1985 and 1990, respectively. Since 1992, he has been with the Department of Electronics and Telecommunications of the University of Florence, first as an Assistant Professor, and since 2002 as an Associate Professor of Electrical Communications. In 1989, he was a Research Fellow with the Signal Processing Division at the University of Strathclyde, Glasgow, UK. During spring 2000 and summer 2001, he was a Visiting Professor at the Tampere International Center for Signal Processing (TICSP), Tampere, Finland. His main research interests are lossless and near-lossless compression of remote sensing and medical imagery, multiresolution image analysis and processing, nonlinear filtering, multisensor image fusion, and analysis and processing of hyperspectral and synthetic aperture radar images. He coauthored 50 papers published on international peer-reviewed journals and holds a patent on a procedure for progressive image transmission.

Stefano Baronti was born in Florence, Italy, in 1954. He received the Laurea degree in electronic engineering from the University of Florence, Florence, Italy, in 1980. After a period spent with the Italian Highway Company working on data collection and analysis, he joined the National Research Council of Italy (CNR) in 1985 as a Researcher of the "Nello Carrara" IFAC-CNR (formerly IROE-CNR), Florence, Italy. From 1985 to 1989, he was involved in an ESPRIT Project of the European Union aimed at the development of an automated system for quality control of composite materials through analysis of infrared image sequences. Later, he moved toward remote sensing image processing by participating in and as the head of several projects funded by the Italian, French, and European Space Agencies. His research topics are in digital image processing and analysis aimed at computer vision and cultural heritage applications, data compression, and image communication (including medical imaging), and optical and microwave remote sensing by synthetic aperture radar. He has coauthored about 40 papers published in international peer-reviewed journals. He is a Member of IEEE Signal Processing Society and the IEEE Geoscience and Remote Sensing Society's Data Fusion Committee.