A Lightweight Contextual Arithmetic Coder for On-Board Remote Sensing Data Compression

Joan Bartrina-Rapesta, Ian Blanes, *Senior Member, IEEE*, Francesc Aulí-Llinàs, *Senior Member, IEEE*, Joan Serra-Sagristà, *Senior Member, IEEE*, Victor Sanchez, and Michael W. Marcellin, *Fellow, IEEE*,

***Abstract*— The Consultative Committee for Space Data Systems (CCSDS) has issued several data compression standards devised to reduce the amount of data transmitted from satellites to ground stations. This paper introduces a contextual arithmetic encoder for on-board data compression. The proposed arithmetic encoder checks the causal adjacent neighbors, at most, to form the context and uses only bitwise operations to estimate the related probabilities. As a result, the encoder consumes few computational resources, making it suitable for on-board operation. Our coding approach is based on the prediction and mapping stages of CCSDS-123 lossless compression standard, an optional quantizer stage to yield lossless or near-lossless compression and our proposed arithmetic encoder. For both lossless and near-lossless compression, the achieved coding performance is superior to that of CCSDS-123, M-CALIC, and JPEG-LS. Taking into account only the entropy encoders, fixed-length codeword is slightly better than MQ and interleaved entropy coding.**

***Index Terms*— Arithmetic coding, Consultative Committee for Space Data Systems (CCSDS)-123, lossless and near-lossless coding, remote sensing data compression.**

I. Introduction

REMOTE sensing imagery is becoming an invaluable tool for governments, rescue teams, and aid organizations to manage infrastructure and natural resources, to appraise climate changes, or to give support when natural disasters strike. Since remote sensing images tend to be very large, high-performance compression techniques are of paramount importance.

Let I, J, and K be the number of columns, rows, and components of an image x and let $x_{i,j,k}$ denote a pixel at location (i, j, k) of the image. Such an image is commonly compressed employing one of three regimes: lossless compression, which allows perfect reconstruction of the original image x; lossy compression, which approximates x, introducing an error in the reconstructed image x' that enables a higher compression ratio than possible with lossless compression; or near-lossless compression, which is a particular case of lossy compression where the peak absolute error (PAE) of x' is controlled during the coding process with a tolerance value Δ. Specifically

$$\max_{i,j,k} \left\{ \left| x_{i,j,k} - x'_{i,j,k} \right| \right\} \leq \Delta. \quad (1)$$

Within the Consultative Committee for Space Data Systems (CCSDS) [1], the Multispectral and Hyperspectral Data Compression Working Group is in charge of proposing techniques for remote sensing data compression. Such techniques are mainly developed to be implemented on board, where limited resources are available and low complexity encoders are needed. In 1997, the CCSDS published CCSDS-121.0-B-1 [2], aimed at lossless data compression. In 2005, the CCSDS published CCSDS-122.0-B-1 [3], devised for lossless and lossy compression of monocomponent images based on wavelet transforms. In 2012, the CCSDS published its latest standard, CCSDS-123.0-B-1 [4], focused on lossless compression for multispectral and hyperspectral images based on prediction. Note that to date, there is no CCSDS standard proposal devised to multispectral and hyperspectral images for near-lossless coding. In what follows, we will refer to CCSDS-123.0-B-1 as CCSDS-123.

Lossless and near-lossless coding is an active research topic, as witnessed by the number of recent publications in the last decade [5]–[17]. Some of these contributions, such as [7], [11], and [15]–[17], yield better coding performance than CCSDS-123 for lossless compression but at the expense of an increased computational complexity. Among them, the results provided in [7] can be misleading, since they were obtained using images from the 1997 AVIRIS products, which are known to have undergone an inappropriate calibration [18]. Next three contributions [11], [15], and [16] yield better coding performance than CCSDS-123, but at the expense of an increased computational complexity due to the expensive algorithms applied to improve prediction estimation. The last contribution [17] yields competitive coding performance by including a light spectral regression in the spectral domain, which has a low computational cost.

It is worth noting that none of the previous techniques provides support for near-lossless coding, which is demanded if even better coding performance is requested. Near-lossless coding [5], [6], [8]–[10], [12]–[14] can yield higher compression ratios at a bounded distortion of $\Delta > 0$. Some of the most

Manuscript received December 8, 2016; revised March 30, 2017; accepted April 28, 2017. This work was supported in part by the Spanish Ministry of Economy and Competitiveness and in part by the European Regional Development Fund under Grant TIN2015-71126-R, in part by the Catalan Government under Grant 2014SGR-691, and in part by the Center National d'´Etudes Spatiales. *(Corresponding author: Joan Bartrina-Rapesta.)*

J. Bartrina-Rapesta, I. Blanes, F. Aulí-Llinàs, and J. Serra-Sagristà are with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, E-08193 Barcelona, Spain (e-mail: joan.bartrina@uab.cat).

V. Sanchez is with the University of Warwick, Coventry CV4 7AL, U.K.

M.W. Marcellin is with the University of Arizona, Tucson, AZ 85721-0104 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2701837

prominent recent contributions for near-lossless compression are [12], which presents an overview of the latest coding standards for remote sensing, including a near-lossless version of CCSDS-123; [8] and [9], which introduce a near-lossless coding based on wavelet transforms; [14], which goes one step further, proposing an embedded near-lossless coding system based on wavelet transform and prediction coding; and [13], which presents a rate control method for predictive image encoders using the CCSDS-123 predictor. Most of the latest contributions use the CCSDS-123 predictor, since it is suitable for being used on board thanks to its low complexity and high decorrelation efficiency.

After the predictor of CCSDS-123, one can choose between a sample- or a block-adaptive encoder. The sample-adaptive encoder achieves better performance than the block-adaptive encoder when the signal is encoded at more than 1 b/sample. However, because the minimum codeword length of the sample-adaptive encoder is 1 b, block-adaptive encoding yields superior performance for signals that can be encoded at less than 1 b/sample.

Although context-based arithmetic encoders typically obtain excellent coding performance at all rates, they are not included in CCSDS-123 because they can have a high computational demand owing to: 1) probability estimation; 2) the renormalization procedure; and 3) context formation, which are expensive operations and are executed intensively. Despite the computational demand of context-based arithmetic encoders, they are included in some remote sensing coding approaches [6], [19], [20]. Contributions aimed to reduce the computational load by estimating the probability using multiplication-free implementations can be found in the literature: the Q coder [21] approached the interval division by means of lookup tables and the M coder [22] uses a reduced range of possible subinterval sizes together with lookup tables. Some methods based on these approaches have been introduced in different standards [23]–[26]. The operations carried out by the renormalization procedure can be avoided if, instead of producing a single codeword, the coder produces short codewords of fixed length [27], [28]. In particular, [28] presents a context-adaptive binary arithmetic coder with fixed-length coderwords (FLWs) that outperforms the MQ [29] and M coders in terms of coding performance. FLW avoids the renormalization procedure but still estimates probabilities through the division.

It is worth noting that none of the previously mentioned contributions is devised to reduce the computation related to probability estimation and the renormalization simultaneously. In this paper, we propose an arithmetic encoder that: 1) utilizes inexpensive operations to estimate probabilities; 2) does not incorporate the renormalization procedure; and 3) employs a simple context model. It yields strong coding performance at low and high rates for remote sensing images. Our probability estimation procedure builds on that of FLW. Originally, FLW uses a sliding window to estimate the probability of the symbols coded using a division operation. Herein, the sliding window size of FLW is adapted to deal only with power of two sizes, which allows the use of low-complexity bitwise operations and spares the division.

Fig. 1. CCSDS-123 encoding scheme.

The proposed arithmetic coder is incorporated in a lossless and near-lossless coding scheme, providing improved compression performance over current remote sensing image compression techniques. Roughly described, the adopted coding scheme departs from the predictor and mapping included in CCSDS-123 and utilizes a near-lossless quantizer, employs a binary arithmetic coder that operates on a line-by-line and bitplane-by-bitplane basis, introduces a new context model that evaluates (at most) only causal adjacent samples, and uses only bitwise operations to estimate symbol probabilities. Extensive experimental results indicate that our proposed approach improves on CCSDS-123 in terms of lossless compression ratios and also outperforms a near-lossless version of the sample-adaptive and block-adaptive coders of CCSDS-123, JPEG-LS [30] and M-CALIC [6] in terms of lossless and near-lossless coding performances. Comparing only the entropy encoders, FLW is slightly better than MQ and interleaved entropy coder (IEC) [31].

The rest of this paper is structured as follows. Section II briefly reviews the CCSDS-123 coding system and a near-lossless technique for coding systems based on prediction. Section III describes our proposed context-based arithmetic coder with bitwise probability estimation. Section IV describes how our proposed arithmetic coder is incorporated in a coding scheme that uses the predictor of CCSDS-123. Section V presents the experimental results. Section VI concludes this paper.

II. CCSDS-123 and Near-Lossless Compression

A. CCSDS-123

The CCSDS-123 standard, which is limited to encoding samples of $N = 16$ b/pixel/band, can be structured in three stages: *predictor*, *mapper*, and *entropy encoder*. Fig. 1 illustrates the encoding pipeline of CCSDS-123.

In summary, the predictor estimates the value of the current sample $x_{i,j,k}$ using previously scanned samples. This predicted sample is denoted by $\widetilde{x}_{i,j,k}$. The prediction error Λ is computed as

$$\Lambda_{i,j,k} = x_{i,j,k} - \widetilde{x}_{i,j,k} \quad (2)$$

and then mapped to a non-negative integer $\lambda_{i,j,k}$ called the mapped prediction residual. The entropy encoder is in charge of encoding $\lambda_{i,j,k}$ without loss. For the entropy encoder in CCSDS-123, one can choose between a sample- and a block-adaptive encoder.

Further details of the CCSDS-123 stages can be found in [12] and [32].

B. Near-Lossless Compression

For the encoder described above, the decoder can reproduce $x_{i,j,k}$, without loss. In this section, we discuss the addition of a

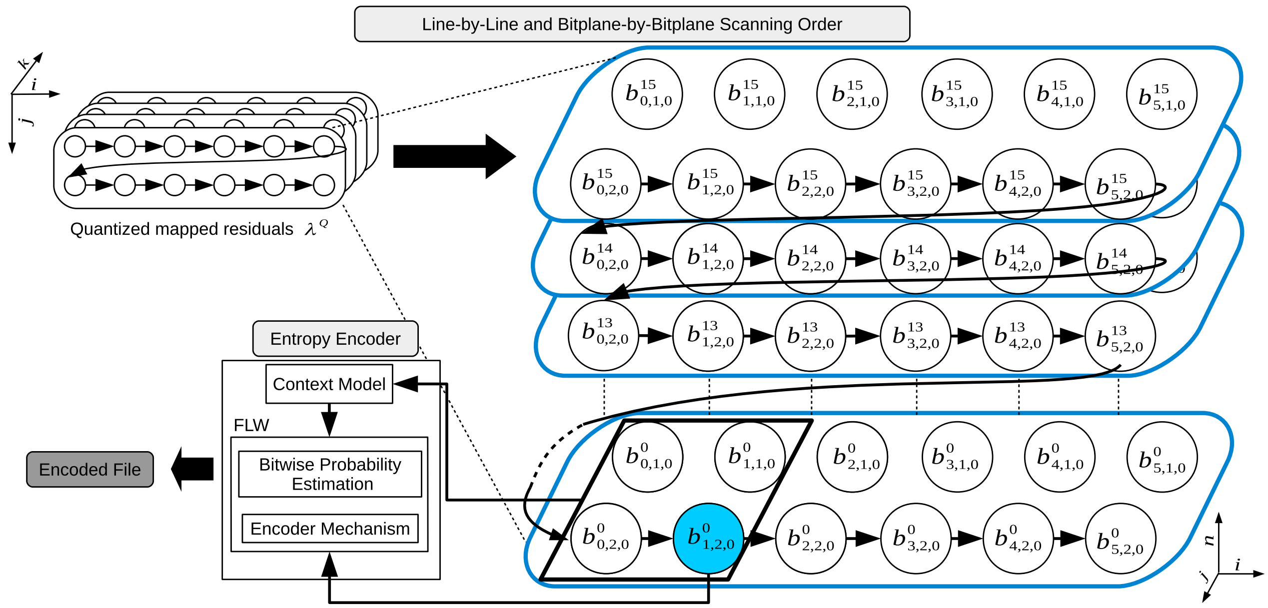


Fig. 2. Illustration of the scanning order and the entropy encoder.

quantizer, which results in higher compression ratios, but at the expense of some loss of fidelity in the decompressed image. The simplest and most effective way to design a prediction-based lossy compression algorithm is to quantize the prediction error $\Lambda_{i,j,k}$ with a quantizer Q, resulting in quantized-then-dequantized version $\widehat{\Lambda}_{i,j,k}$ (and, in consequence, $\widehat{\lambda}_{i,j,k}$). The resulting quantization index is referred to as $\Lambda^Q_{i,j,k}$ and its remapped version is denoted by $\lambda^Q_{i,j,k}$. Subsequent predictions $\widetilde{x}_{i,j,k}$ are calculated using previous reconstructed (lossy) samples $\widehat{x}_{i,j,k}$, which are obtained by implementing a decoder in the encoder [12], [33]. The decoder creates the reconstructed (lossy) image samples via

$$\widehat{x}_{i,j,k} = \widehat{\Lambda}_{i,j,k} + \widetilde{x}_{i,j,k}. \tag{3}$$

It is worth noting that the errors in the reconstructed pixels are identical to the errors introduced in the prediction errors by the quantizer. That is, $x_{i,j,k} - \widehat{x}_{i,j,k} = \Lambda_{i,j,k} - \widehat{\Lambda}_{i,j,k}$. Thus, the errors in reconstructed pixels can be precisely controlled by controlling the individual quantization errors. This is the basis of "near-lossless compression."

III. Lightweight Binary Arithmetic Coder With Context Model

The entropy encoder presented in this paper works with binary symbols. To this end, we denote the nth bit of the binary representation of $\lambda^Q_{i,j,k}$ by $b^n_{i,j,k}$, with $N-1 \geq n \geq 0$. Here, N is chosen to provide a sufficient number of bits to represent all the $\lambda^Q_{i,j,k}$, being $b^{N-1}_{i,j,k}$ the most significant bit.

To facilitate use with on-board sensors, our proposal processes data in a line-by-line fashion. Once a line is scanned, predicted, and mapped to positive values, it is entropy encoded on a bitplane-by-bitplane basis. The entropy encoder makes use of context model patterns obtained using a context window that contains symbols coded previously to the current symbol. The top left of Fig. 2 displays the quantized and remapped prediction residuals λ^Q. The binary representation of these samples is shown on the right, while the bottom left portrays the entropy encoder, which is fed by the current bit to be encoded and its context. The bit to be encoded is shaded in blue, while the context window is framed with a rectangle.

A. Context Model

Let $\mathbf{M}$ be the set of all possible patterns that can occur within the context window, with context $m \in \mathbf{M}$ being a particular realization, resulting in a context index $c \in \mathbf{C} = \{0, \ldots, C-1\}$. These context indices (loosely referred to as contexts in what follows) are determined by a modeling function $F : \mathbf{M} \rightarrow \mathbf{C}$. For each bit b to be coded, a probability model is used, corresponding to its context c. In particular, the probability model estimates the conditional probability $p(b|c) = p(b|F(m))$. After encoding, the probability model is updated with the latest coded bit b. That is, $p(b|c)$ is estimated on the fly. Specifically, our probability model estimates the probability $p(b = 0|c)$. A careful design of the context model is required to obtain high coding efficiency. This task is complicated by the goal of achieving low encoder complexity for the purpose of operating on onboard remote sensing scenarios.

A simple strategy for context modeling employs a context window that contains only the three nearest causal neighbors as depicted in Fig. 2. We consider several choices for the context modeling function F. The first ignores all samples within the context window except the one directly above the sample of interest. This is indicated in Fig. 3(a). Three other choices are shown in Fig. 3(b)–(d). The notations V, H, HV, and HVD are used in Fig. 3, where V (vertical) denotes the sample above the bit to be encoded, H (horizontal) denotes the sample to the left, and D (diagonal) denotes the sample to

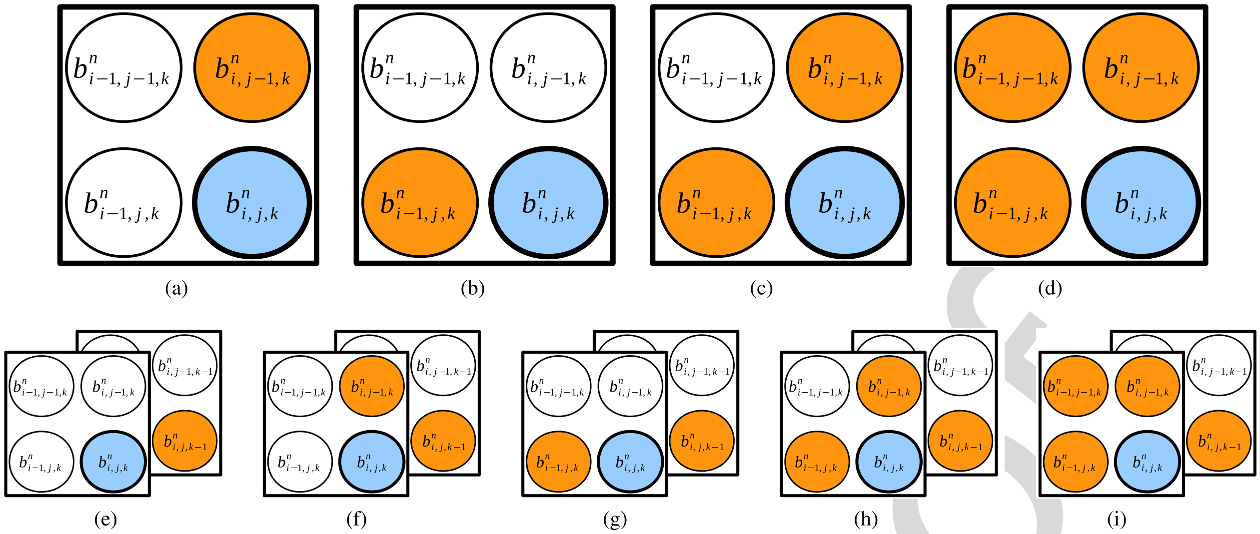


Fig. 3. Illustration of different context models to encode $b^n_{i,j,k}$. (a) V. (b) H. (c) HV. (d) HVD. (e) S. (f) VS. (g) HS. (h) HVS. (i) HVDS.

TABLE I
CONTEXT ASSIGNMENTS FOR THE V, H, HV, AND HVD MODELING FUNCTIONS

c	V	H	HV		HVD		
	$s^n_{i,j-1,k}$	$s^n_{i-1,j,k}$	$s^n_{i,j-1,k}$	$s^n_{i-1,j,k}$	$s^n_{i,j-1,k}$	$s^n_{i-1,j,k}$	$s^n_{i-1,j-1,k}$
0	0	0	0	0	0	0	0
1	1	1	0	1	0	0	1
2			1	0	0	1	0
3			1	1	0	1	1
4					1	0	0
5					1	0	1
6					1	1	0
7					1	1	1

the left and above. To take advantage of dependencies between spectral components, the preceding spectral component $k-1$ can be included in the context window. In this case, S (spectral) denotes the coregistered sample in the previous spectral component. The inclusion of this sample gives rise to five additional modeling functions as shown in Fig. 3(e)–(i). Note that if only samples H and S are employed by the modeling function, only the current scanned line must be stored in memory. For all other modeling functions, the previous and the current lines are necessary.

Rather than the actual bit (from bitplane n) of each neighboring sample, the so-called "significance state" is employed to compute the context c. To this end, let $s^n_{i,j,k}$ denote the significance state of the sample at location i, j, k at bitplane n. A value of 1 indicates that the sample contains a 1 at bitplane n or higher. Table I shows how c is derived from the significance states of the neighbors for the V, H, HV, and HVD modeling functions. The S modeling function results in two states, i.e., $c \in \{0, 1\}$. The VS, HS, HVS, and HVDS modeling functions result in twice the number of states than their counterparts that do not employ S. They are not shown in Table I for the sake of space. The experimental results for all context modeling functions are presented in a subsequent section.

Before finishing this section, we note that the entropy coder and its associated probability models are initialized at the beginning of each bit plane of each component. In particular, the initial probability model for each context is set to a value of $p(b=0|c) = 0.66$. The probability is biased toward 0 since, as found empirically, bits of higher bitplanes have higher probability of being 0, thus allowing FLW to adapt faster. This, together with the fact that all bitplane data from the current line (and its predecessor, when relevant) are available in the encoder, leads to the conclusion that the bitplanes of the current line can be encoded in parallel. This parallel strategy is not possible in the decoder. The use of significant states in the context formation process requires that bitplanes be decoded sequentially. We note that the probabilities are reset ($p(b=0|c) = 0.66$) at the beginning of each component without penalizing the coding performance. This is because only 2^{12} symbols are encoded with the default probability value, which on average for the image corpora used, corresponds to the 0.06% of the total symbols per band to be encoded.

B. Bitwise Probability Estimation

As mentioned before, FLW was devised to reduce computational costs through the use of FLWs, which avoids a renormalization operation, but is not aimed to reduce the computational load derived from probability estimation [28].

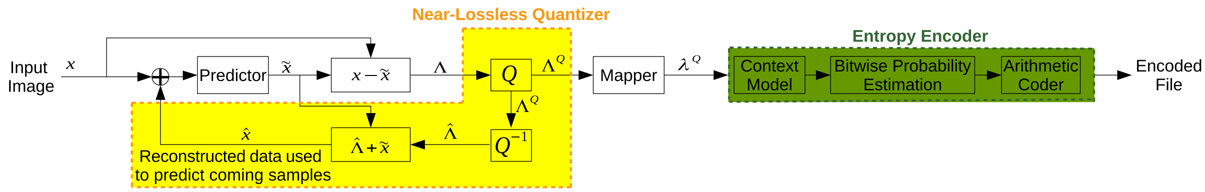


Fig. 4. Adopted coding approach.

TABLE II
MAIN DIFFERENCES BETWEEN CCSDS-123 AND THE ADOPTED APPROACH

	Predictor	Quantizer	Mapper	Data Scanned Supported	Context Model	Entropy Encoder
CCSDS-123	✔	✘	✔	Line or Block	✘	Sample-Adaptive or Block-Adaptive
Proposal	✔	✔	✔	Line	✔	Contextual Binary Arithmetic Coder

For each context c, FLW uses a sliding window of symbols coded with that context. The length of this window varies between $\mathcal{T}$ and $2\mathcal{T}-1$ symbols. The probability estimate is updated once every $\mathcal{V}$ symbols coded, according to

$$p(b=0|c) = \frac{Z \ll \mathcal{B}}{W} \quad (4)$$

with W representing the number of symbols within the window, Z the number of zeroes within the window, and $\mathcal{B}$ the number of bits used to express symbol probabilities. The numerator of the expression is computed by left shifting the binary representation of Z by $\mathcal{B}$ bits. The size of the window is incremented each time a symbol is encoded using context c until $W = 2\mathcal{T}$, at which time the window size is immediately reduced to $\mathcal{T}$ and the number of zeroes within the window is updated according to $Z \leftarrow Z - Z'$ and $Z' \leftarrow Z$, with Z' being the number of zeroes coded during the most recent $\mathcal{T}$ symbols.

In the original approach of FLW as formulated above, $p(b|c)$ is computed via a division operation to achieve maximum accuracy. Such a division may tax the on-board computational resources in a remote sensing scenario. To reduce computational complexity, we propose to estimate the probability through bitwise operations. The substitution of the division by bitwise operations requires that $\mathcal{V} = \mathcal{T}$ and that both are a power of two. This forces the sliding window to contain a power of two symbols, so the probability can be updated using only bit shift operations according to

$$p(b|c) = (Z \ll \mathcal{B}) \gg \log_2(W) \quad (5)$$

where W and Z are then updated through $W \leftarrow W \gg 1$ and $Z \leftarrow Z \gg 1$. Note that this update rule for Z approximates only the number of zeroes in the most recent $\mathcal{T}$ coded samples. Nevertheless, the update can be carried out in the decoder using the same approximation. At the beginning of encoding, the probability is first updated when $\mathcal{V}$ symbols are coded. Subsequently, it is updated every $\mathcal{V}/2$ symbols. The strategy proposed here can be seen as a special case of (4), which was not explored in [28].

Using (5) instead of (4) reduces the flexibility of the arithmetic coder since the updating of the probability estimates and the window size are tied together. The maximum performance achieved with the original formulation of the arithmetic coder proposed in [28] is achieved when the probability estimate is updated every symbol, i.e., $\mathcal{V} = 1$, regardless of the window size. The strategy proposed here provides a significant reduction in complexity with a minor reduction in compression performance. The experimental results provided in Section V indicate that our approach yields highly competitive compression performance.

IV. ADOPTED CODING APPROACH

Although the novel entropy encoder presented here may be incorporated in any coding system, we employ it in the CCSDS-123 coding pipeline. Fig. 4 illustrates the adopted coding approach, which employs the predictor and mapper of CCSDS-123, but adds a near-lossless quantizer (see the yellow block), and substitutes the usual CCSDS-123 encoder by our entropy encoder (see the green block). The circle containing a cross at the left side of Fig. 4 indicates that the input to the predictor is either the original pixel x (when the optional quantization is not present) or the reconstructed pixel $\hat{x}$ (when quantization is present).

The adopted coding scheme is evaluated with a uniform quantizer (UQ) and a uniform scalar deadzone quantizater (USDQ) [29]. The UQ operates over $\Lambda_{i,j,k}$ to obtain a quantization index according to

$$\Lambda^Q_{i,j,k} = \mathrm{sign}(\Lambda_{i,j,k}) \left\lfloor \frac{|\Lambda_{i,j,k}| + \Delta}{2\Delta + 1} \right\rfloor \quad (6)$$

where $2\Delta + 1$ is the quantization step size. The operation to reconstruct $\hat{\Lambda}_{i,j,k}$ from its quantization index is given by

$$\hat{\Lambda}_{i,j,k} = \mathrm{sign}\big(\Lambda^Q_{i,j,k}\big)(2\Delta+1)\Lambda^Q_{i,j,k}. \quad (7)$$

The UQ is employed in lossless compression techniques such as JPEG-LS, M-CALIC, and 3-D-CALIC [34]. On the other hand, the USDQ quantizes $\Lambda_{i,j,k}$ to obtain a quantization index according to

$$\Lambda^Q_{i,j,k} = \mathrm{sign}(\Lambda_{i,j,k}) \left\lfloor \frac{|\Lambda_{i,j,k}|}{\Delta + 1} \right\rfloor \quad (8)$$

TABLE III

SUMMARY OF DATA USED IN THE EXPERIMENTAL RESULTS. SENSOR NAME, ITS ABBREVIATION, THE NUMBER OF IMAGES FROM EACH SENSOR, AND FIRST-ORDER ENTROPIES (IN BITS PER SAMPLE) ON AVERAGE PER SENSOR ARE PROVIDED. THE LAST TWO COLUMNS INDICATE THE PREDICTOR MODE AND THE LOCAL SUM USED FOR EACH SENSOR

Sensor	Abbreviation	Number of images	Entropy	Predictor Mode	Local Sum
Aviris Calibrated	AC	5	9.77	Neighbor Oriented	Full Mode
Aviris Uncalibrated	AU	7	11.21	Neighbor Oriented	Full Mode
Airs	A	9	11.34	Neighbor Oriented	Reduced Mode
Casi	C	2	10.52	Neighbor Oriented	Reduced Mode
Crism	Cr	20	10.69	Column Oriented	Reduced Mode
Hyperion	H	4	9.53	Column Oriented	Reduced Mode
M3	M3	5	9.19	Column Oriented	Reduced Mode
Total / average	——	52	10.41	——	——

where the quantization step is $\Delta + 1$. The operation to reconstruct $\hat{\Lambda}_{i,j,k}$ from its quantization index is expressed as

$$\hat{\Lambda}_{i,j,k} = \text{sign}\big(\Lambda^{Q}_{i,j,k}\big)(\Delta + 1)\Lambda^{Q}_{i,j,k}. \tag{9}$$

Due to its straightforward implementation and excellent performance, the USDQ has been selected for the JPEG 2000 standard [24]. The USDQ partitions the range of input values into intervals all of size Δ, except for the interval that contains zero, which is of size 2Δ. This results in all absolute pixel errors $|x_{i,j,k} - \widehat{x}_{i,j,k}|$ being bounded above Δ for both quantizers.

Table II summarizes the main differences between CCSDS-123 and the adopted coding scheme.

V. EXPERIMENTAL RESULTS

This section presents a set of experiments aimed at the analysis and evaluation of the adopted coding scheme. First, the proposed context modeling functions are evaluated in terms of the conditional entropy of the prediction residual. The bitwise probability estimator is then evaluated via the same performance metric to determine its proper configuration. A variety of binary encoder mechanisms such as IEC, MQ, and FLW are evaluated in terms of their lossless compression performance in conjunction with the proposed context modeling and probability estimation. Finally, the resulting proposed overall approach is compared in terms of lossless and near-lossless compression performances with CCSDS-123, JPEG-LS, and M-CALIC.

For the experiments conducted in this paper, we have selected a set of images[1] collected with different sensors that are included in CCSDS MHDC-WG corpus. The sensor names and their main features are listed in Table III. The average entropy is reported for each image type. The reported values are first-order entropy; they represent the entropy of individual pixels, without accounting for any dependencies among pixels within or between components.

In [35], the impact of different CCSDS-123 parameters that control the operation of the prediction and the entropy encoder was evaluated, suggesting that a correct parameter selection had more impact on the predictor stage than in the entropy encoder stage. Concerning the prediction, the parameters local sum type, prediction mode, the number of prediction bands, and predictor adaption rate were the most critical. Extensive experimental evaluations were conducted to find suitable configurations.

In this paper, leaning on the results in [35] and after conducting an extensive evaluation also, experimental results are produced for the following parameter configuration: the local sum type and predictor mode depend on the acquisition sensor (as indicated in the last two columns of Table III); the number of prediction bands P is set to 3, since it is a good tradeoff between the computational load and the coding performance; and the predictor adaptation rate $\nu_{\max}$ is set to 3, since, in general, it yields the best performance.

For evaluating the performance of context modeling and probability estimation, we employ the conditional entropy of the prediction residuals, as mentioned above. For the work proposed here, binary entropy coding is employed. To yield results with units in bits per pixel, the binary entropies of all bitplanes are added. Since our context model estimates the probability of $p(b = 0|c)$, the conditional entropy of an image (in bits) is computed as

$$H(\lambda^Q) = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1}\sum_{k=0}^{K-1}\sum_{n=0}^{15} \times \begin{cases} \log_2\left(p\left(b^n_{i,j,k} = 0|c\right)\right) & \text{if } b^n_{i,j,k} = 0 \\ \log_2\left(1 - p\left(b^n_{i,j,k} = 0|c\right)\right) & \text{if } b^n_{i,j,k} = 1 \end{cases} \tag{10}$$

where λ^Q denotes the symbols to be entropy coded.

A. Context Modeling Function

The context model is used to select the probability model that is employed to encode the current symbol. In this first experiment, each of the probability models themselves is estimated using the high-performance method given by (4) employing $\mathcal{V} = 1$ and $\mathcal{T} = 2^{12}$, without regard to complexity.

Table IV provides the conditional entropy obtained (in bits per sample) for the different context formations defined in Section III-A, i.e., V, H, HV, HVD, S, VS, HS, HVS, and HVDS. The results from Table III suggest the following.

[1]The images used are available at http://cwe.ccsds.org/sls/docs/sls-dc/123.0-B-Info/TestData

TABLE IV

CONDITIONAL ENTROPY OF THE PREDICTION RESIDUALS (IN BITS PER SAMPLE) FOR THE CONTEXT MODELING FUNCTIONS DENOTED BY V, H, HV, HVD, S, VS, HS, HVS, AND HVDS. RESULTS ARE REPORTED ON AVERAGE FOR DIFFERENT SENSORS AND $\Delta = 0$

	Context Formation								
Sensor	V	H	HV	HVD	S	VS	HS	HVS	HVDS
AC	3.69	3.69	3.68	3.68	3.69	3.68	3.68	3.68	3.67
AU	5.01	5.00	5.00	4.99	5.00	4.99	4.99	4.99	4.99
A	4.24	4.23	4.24	4.24	4.24	4.24	4.24	4.24	4.23
C	4.85	4.85	4.84	4.84	4.85	4.84	4.84	4.83	4.83
Cr	4.15	4.21	4.14	4.14	4.20	4.12	4.19	4.12	4.12
H	4.26	4.26	4.25	4.25	4.29	4.26	4.26	4.25	4.25
M3	2.66	2.70	2.65	2.65	2.70	2.64	2.68	2.63	2.63
Average	4.12	4.14	4.11	4.11	4.14	4.11	4.13	4.10	4.10

TABLE V

CONDITIONAL ENTROPY OF THE PREDICTION RESIDUALS (IN BITS PER SAMPLE) FOR $\Delta = 0$ RESULTING FROM THE MAXIMUM PRECISION AND THE BITWISE PROBABILITY ESTIMATORS. THE V CONTEXT MODEL IS EMPLOYED IN EACH CASE. THE BEST RESULTS FOR EACH STRATEGY ARE REPRESENTED IN BOLD

	division $\mathcal{V} = 1$					bitwise operations $\mathcal{V} = \mathcal{T}$				
$\mathcal{T} =$	2^{16}	2^{14}	2^{12}	2^{10}	2^{8}	2^{16}	2^{14}	2^{12}	2^{10}	2^{8}
AC	3.70	3.69	**3.69**	3.70	3.75	3.72	3.70	**3.69**	3.70	3.77
AU	5.02	5.01	**5.01**	5.01	5.06	5.04	5.02	**5.01**	5.02	5.08
A	4.27	4.26	**4.23**	4.24	4.26	4.32	4.29	**4.24**	4.26	4.28
C	4.94	4.86	**4.85**	4.86	4.89	4.90	4.86	**4.85**	4.86	4.91
Cr	4.16	4.15	**4.15**	4.15	4.18	4.21	4.17	**4.16**	4.16	4.20
H	4.27	4.26	**4.26**	4.27	4.31	4.29	4.27	**4.26**	4.27	4.32
M3	2.69	2.66	**2.66**	2.67	2.70	2.71	2.68	**2.67**	2.68	2.72
Average	4.15	4.13	**4.12**	4.13	4.16	4.17	4.14	**4.13**	4.13	4.18

1) All of the modeling functions provide significant improvements over the pixel entropy reported in Table III.
2) The differences in performance between the modeling functions are generally small.
3) Although the context models H and S yield the worst performance on average, they are the best option when memory resources are severely limited since they need only to store samples from the current line to be encoded.
4) Adding the S sample to a context results in an improvement of only about 0.01 b/sample.
5) The V context obtains a coding benefit of 0.02 b/sample on average with respect to the H context and only adds the previous processed line to its storage requirements.

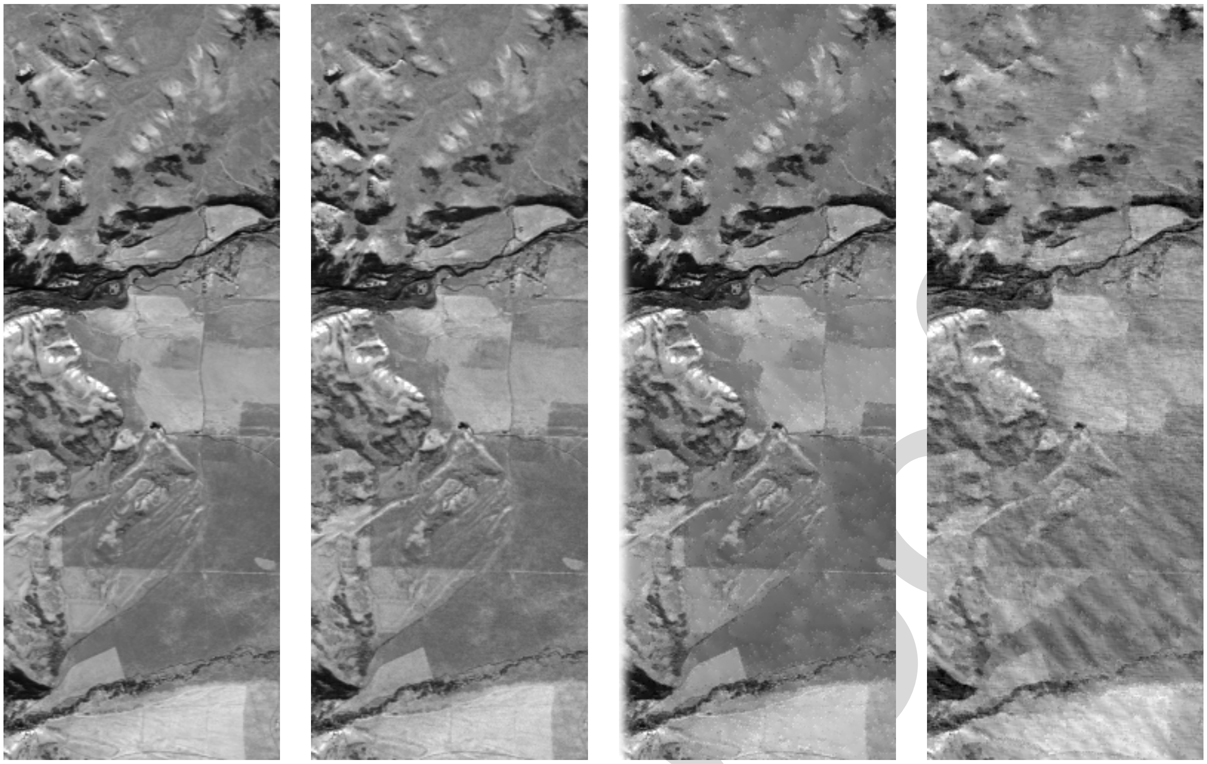
In what follows, we select context model V for further evaluation due to its favorable tradeoff among the performance, memory resources, and computational load.

B. Probability Estimation

This section reports the results obtained by the two different probability estimation strategies discussed in Section III. In particular, Table V reports the conditional entropy of the prediction residuals resulting from the two different probability estimation strategies. In both cases, the V context model is employed. The left of Table V presents results for the maximum precision technique (using division), as defined by (4). These results are shown for different values of $\mathcal{T}$, but $\mathcal{V} = 1$. The right side of Table V presents results for the bitwise strategy, as defined by (5). The same values of $\mathcal{T}$ are explored, but always with $\mathcal{V} = \mathcal{T}$, as required to avoid division. The results suggest that $\mathcal{T} = 2^{12}$ attains the highest performance for both strategies. A larger $\mathcal{T}$ degrades the coding performance because the window may contain symbols that are not correlated with the current one. A smaller $\mathcal{T}$ degrades the coding performance because there are insufficient symbols to reliably estimate the probabilities $p(b|C)$. The results of Table V also indicate that the low-complexity strategy that employs bitwise operations is as competitive as that employing division. Although not tabulated here for the sake of space, these results hold for the other context modeling functions considered in the previous sections.

C. Entropy Coding

We note that the context model and probability estimator proposed here can be used with any entropy encoder that codes binary symbols according to a given probability model, such as MQ, IEC or the adopted FLW. Table VI provides the actual



(a) (b) (c) (d)

Fig. 5. Visual comparison for the "Aviris Calibrated Yellowstone sc00" image. (a) Original. (b) Proposed approach at 0.43 b/sample ($\Delta = 20$). (c) M-CALIC at 0.42 b/sample ($\Delta = 30$). (d) CCSDS-123 at 0.50 b/sample ($\Delta = 80$).

TABLE VI

CODING PERFORMANCE (IN Bits per Sample) OF THE PROPOSED APPROACH USING MQ, IEC, AND FLW ENTROPY ENCODING. ALL RESULTS EMPLOY CONTEXT MODEL V AND BITWISE PROBABILITY ESTIMATION WITH $\mathcal{T} = \mathcal{V} = 2^{12}$

Sensor	MQ	IEC	FLW
AC	3.74	3.72	3.71
AU	5.06	5.04	5.03
A	4.30	4.29	4.28
C	4.90	4.88	4.87
Cr	4.20	4.18	4.18
H	4.31	4.29	4.28
M3	2.71	2.70	2.69
Average	4.17	4.16	4.15

compression results (in bits per sample) obtained using the MQ, IEC, and FLW entropy coders. In each case, the results are obtained with context model V and the bitwise estimator with $\mathcal{T} = \mathcal{V} = 2^{12}$. From these results, we can see that, on average, FLW yields slightly better results than IEC and MQ.

D. Lossless and Near-Lossless Compression

The results reported in this section compare the lossless performance of the proposed approach with those of JPEG-LS, M-CALIC, and CCSDS-123. Additionally, we compare its *near-lossless* performance with those of JPEG-LS and M-CALIC and the implementation of CCSDS-123. Different quantizers have been combined with our proposal and CCSDS-123, to obtain an as fair as possible comparison. In particular, the UQ and the USDQ discussed in Section IV are compared.

M-CALIC and the near-lossless version of CCSDS-123 are considered to be state of the art in terms of compression performance and computational complexity, and JPEG-LS is a standard technique with near-lossless features. All results for the proposed scheme are produced using the FLW arithmetic coder, context model V, and the bitwise probability estimator having $\mathcal{V} = \mathcal{T} = 2^{12}$. The results reported in Table VII indicate that our method outperforms both M-CALIC and CCSDS-123 in terms of lossless coding ($\Delta = 0$) for all sensors. In the near-lossless regime ($\Delta > 0$), the proposed approach outperforms M-CALIC when the USDQ is used and in most cases for the UQ. In particular, M-CALIC obtains slightly better results than our proposal only for images acquired with sensors AIRS and Hyperion when the UQ is used. On the other hand, the proposed system always outperforms the near-lossless extension of CCSDS-123 for both quantizers. In addition, in general, for the same Δ value, the coding performance is better for the USDQ than for UQ. Although achieved bit rates vary widely from image to image, low bit rates can be obtained for all images with a modest value of PAEs (maximum absolute pixel error).

E. Visual Comparison

To evaluate visual performance, we show a region cropped from an image encoded at the "same" bit rate by the proposed approach with the UQ, M-CALIC, and CCSDS-123. For CCSDS-123, we employ the block-adaptive coder since we want to compare the images at a bit rate lower than 1 b/sample.

TABLE VII
LOSSLESS ($\Delta = 0$) AND NEAR-LOSSLESS ($\Delta > 0$) COMPRESSION RESULTS FOR THE PROPOSED APPROACH. FOR COMPARISON, THE RESULTS FOR JPEG-LS, M-CALIC, AND CCSDS-123 ARE INCLUDED. BOTH A UQ AND A USDQ HAVE BEEN USED IN OUR PROPOSED APPROACH AND IN OUR NEAR-LOSSLESS EXTENSION TO CCSDS-123 TO PRODUCE RESULTS FOR $\Delta > 0$. THE RESULTS ARE REPORTED IN BITS PER SAMPLE (LOWER IS BETTER)

		CCSDS-123				JPEG-LS	M-CALIC	Our Proposal	
		with UQ		with USDQ				with UQ	with USDQ
Sensor	Δ values	Sample adaptive	Block adaptive	Sample adaptive	Block adaptive				
AC	$\Delta = 0$	3.73	3.91	3.73	3.91	6.41	3.87	**3.71**	**3.71**
	$\Delta = 10$	1.20	0.97	1.20	0.94	2.45	0.76	0.62	**0.60**
	$\Delta = 20$	1.10	0.74	1.10	0.72	1.81	0.50	0.38	**0.36**
	$\Delta = 30$	1.07	0.64	1.07	0.63	1.48	0.40	0.28	**0.27**
AU	$\Delta = 0$	5.06	5.23	5.06	5.23	7.47	5.13	**5.03**	**5.03**
	$\Delta = 10$	1.69	1.68	1.85	1.78	3.41	1.46	**1.39**	1.52
	$\Delta = 20$	1.35	1.19	1.40	1.21	2.68	0.95	**0.87**	0.90
	$\Delta = 30$	1.24	0.98	1.26	0.99	2.30	0.73	**0.65**	0.67
A	$\Delta = 0$	4.29	4.48	4.29	4.48	6.85	4.28	**4.27**	**4.27**
	$\Delta = 10$	1.23	1.12	1.18	0.96	2.62	0.73	0.76	**0.62**
	$\Delta = 20$	1.10	0.75	1.07	0.65	1.86	0.41	0.39	**0.30**
	$\Delta = 30$	1.06	0.63	1.05	0.56	1.50	0.31	0.28	**0.22**
C	$\Delta = 0$	4.97	5.15	4.97	5.15	6.79	4.91	**4.87**	**4.87**
	$\Delta = 10$	1.47	1.48	1.51	1.47	2.64	1.12	**1.10**	**1.10**
	$\Delta = 20$	1.25	1.06	1.25	1.03	1.94	0.68	0.67	**0.64**
	$\Delta = 30$	1.17	0.88	1.18	0.87	1.58	0.52	0.49	**0.48**
Cr	$\Delta = 0$	4.40	4.50	4.40	4.50	5.10	6.91	**4.18**	**4.18**
	$\Delta = 10$	1.64	1.66	1.63	1.42	1.83	2.75	1.26	**0.99**
	$\Delta = 20$	1.43	1.34	1.39	1.05	1.47	2.02	0.92	**0.64**
	$\Delta = 30$	1.34	1.17	1.30	0.89	1.29	1.64	0.76	**0.50**
H	$\Delta = 0$	4.37	4.57	4.37	4.57	6.24	4.80	**4.28**	**4.28**
	$\Delta = 10$	1.38	1.45	1.21	1.08	2.74	1.02	1.06	**0.68**
	$\Delta = 20$	1.24	1.19	1.09	0.74	2.06	0.52	0.77	**0.35**
	$\Delta = 30$	1.18	1.04	1.06	0.63	1.68	0.33	0.62	**0.25**
M	$\Delta = 0$	2.81	2.97	2.81	2.97	4.24	5.18	**2.69**	**2.69**
	$\Delta = 10$	1.26	1.21	1.11	0.74	1.38	1.32	0.76	**0.33**
	$\Delta = 20$	1.17	1.01	1.07	0.60	1.20	0.78	0.56	**0.20**
	$\Delta = 30$	1.14	0.89	1.06	0.55	1.00	0.53	0.46	**0.15**
Average	$\Delta = 0$	4.23	4.40	4.23	4.40	6.16	5.01	**4.15**	**4.15**
	$\Delta = 10$	1.41	1.37	1.39	1.20	2.44	1.31	1.00	**0.83**
	$\Delta = 20$	1.23	1.04	1.20	0.86	1.86	0.84	0.65	**0.48**
	$\Delta = 30$	1.17	0.89	1.14	0.73	1.55	0.64	0.51	**0.36**

We note that none of the schemes compared here includes precise rate control. For this reason, we have employed the following methodology: 1) encode an image using a variety of different quantization step sizes for each compression scheme and 2) choose those encoded images that yield bit rates as close as possible for the three algorithms. We note that a close match was not obtained in the case of CCSDS-123, so a step size was chosen to afford a higher bit rate than that of the proposed approach, thus giving an advantage to CCSDS-123 in terms of visual performance.

The results of this process are shown in Fig. 5 for a crop from component 122 of the image "Aviris Calibrated Yellowstone sc00." The bit rates obtained are 0.43, 0.42, and 0.50 for the proposed approach, M-CALIC, and CCSDS-123, respectively. The reader is invited to zoom in to see the specific visual artifacts arising from the different compression schemes. Fig. 5 indicates that the image obtained by the proposed approach has higher visual quality than those by (near lossless) CCSDS-123 and M-CALIC. In particular, the proposed approach preserves edges and textures very well, while M-CALIC results in smoothness and loss of texture. CCSDS-123 also removes texture, but also introduces an annoying "banding" effect, due to the high step size required to reach 0.50 b/sample.

VI. CONCLUSION

This paper proposes an entropy encoder based on an efficient definition for a context model and the associated strategy to estimate probabilities for use in a fixed-length arithmetic encoder using low-cost bitwise operations. These contributions are incorporated in a coding approach that employs the predictor included in CCSDS-123. A near-lossless quantizer has also been deployed. The entropy encoder works on a line-by-line and bitplane-by-bitplane scanning order. The experimental results indicate that the use of a single neighbor for the context formation is enough to properly exploit the contextual information in the arithmetic encoder and that it is possible to estimate the probability employing bitwise operations without penalizing the coding efficiency. Further results indicate that, on average, our proposal improves the current standard version of CCSDS-123 for lossless coding by more than 0.1 b/sample. Compared with M-CALIC, our proposal provides an average improvement of 0.86 b/sample for lossless, whereas for near-lossless, the benefit ranges from 0.13 to 0.31 b/sample, depending on the allowed PAE.

REFERENCES

[1] *Consultative Committee for Space Data Systems (CCSDS)*. [Online]. Available: http://www.ccsds.org AQ:1
[2] *Lossless Data Compression*, document CCSDS-121.0-B-1, Blue Book, May 1997. AQ:2
[3] *Image Data Compression*, document CCSDS-122.0-B-1, Blue Book, Nov. 2005.
[4] *Lossless Multispectral & Hyperspectral Image Compression*, document CCSDS-123.0-B-1, Blue Book, May 2012.
[5] *JPEG-LS Lossless and Near-Lossless Compression for Continuous-Tone Still Images*, document, ISO/IEC, 1999. AQ:3
[6] E. Magli, G. Olmo, and E. Quacchio, "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 1, pp. 21–25, Jan. 2004.
[7] F. Rizzo, B. Carpentieri, G. Motta, and J. A. Storer, "Low-complexity lossless compression of hyperspectral imagery via linear prediction," *IEEE Signal Process. Lett.*, vol. 12, no. 2, pp. 138–141, Feb. 2005.
[8] G. Carvajal, B. Penna, and E. Magli, "Unified lossy and near-lossless hyperspectral image compression based on JPEG 2000," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 593–597, Oct. 2008.
[9] C.-W. Chen, T.-C. Lin, S.-H. Chen, and T.-K. Truong, "A near lossless wavelet-based compression scheme for satellite images," in *Proc. WRI World Congr. Comput. Sci. Inf. Eng.*, vol. 6. Mar. 2009, pp. 528–532.
[10] S.-C. Tai, T.-M. Kuo, C.-H. Ho, and T.-W. Liao, "A near-lossless compression method based on CCSDS for satellite images," in *Proc. Int. Symp. Comput., Consum. Control*, Jun. 2012, pp. 706–709.
[11] J. Song, Z. Zhang, and X. Chen, "Lossless compression of hyperspectral imagery via RLS filter," *Electron. Lett.*, vol. 49, no. 16, pp. 992–994, Aug. 2013.
[12] I. Blanes, E. Magli, and J. Serra-Sagrista, "A tutorial on image compression for optical space imaging systems," *IEEE Geosci. Remote Sens. Mag.*, vol. 2, no. 3, pp. 8–26, Sep. 2014.
[13] D. Valsesia and E. Magli, "A novel rate control algorithm for onboard predictive coding of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6341–6355, Oct. 2014.
[14] J. Beerten, I. Blanes, and J. Serra-Sagristà, "A fully embedded two-stage coder for hyperspectral near-lossless compression," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1775–1779, Aug. 2015.
[15] W. Jiaji, K. Wanqiu, M. Jarno, and H. Bormin, "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2194–2198, Dec. 2015.
[16] G. Fang and G. Shuxu, "Lossless compression of hyperspectral images using conventional recursive least-squares predictor with adaptive prediction bands," *J. Appl. Remote Sens.*, vol. 10, no. 1, p. 015010, Feb. 2016.
[17] N. Amrani, J. Serra-Sagristà, V. Laparra, M. W. Marcellin, and J. Malo, "Regression wavelet analysis for lossless coding of remote-sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5616–5627, Sep. 2016.
[18] A. B. Kiely and M. A. Klimesh, "Exploiting calibration-induced artifacts in lossless compression of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2672–2678, Aug. 2009.
[19] X. Wu and P. Bao, "L_∞ constrained high-fidelity image compression via adaptive context modeling," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 536–542, Apr. 2000.
[20] B. Aiazzi, L. Alparone, and S. Baronti, "Context modeling for near-lossless image coding," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 77–80, Mar. 2002.
[21] M. J. Slattery and J. L. Mitchell, "The Qx-coder," *IBM J. Res. Develop.*, vol. 42, no. 6, pp. 767–784, Nov. 1998.
[22] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 620–636, Jul. 2003.
[23] *Information Technology—Lossy/Lossless Coding of Bi-Level Images*, ISO/IEC Standard 14 492:2001, Dec. 2001.
[24] *Information Technology—JPEG 2000 Image Coding System—Part 1: Core Coding System*, ISO/IEC, Standard 15 444-1, Dec. 2000. AQ:4
[25] *H.264—Advanced Video Coding for Generic Audiovisual Services*, ITU-T, Standard, May 2003.
[26] *H.265—High Efficiency Video Coding*, ITU-T, Standard, Apr. 2015. AQ:5
[27] M. D. Reavy and C. G. Boncelet, "An algorithm for compression of bilevel images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 669–676, May 2001.
[28] F. Aulí-Llinàs, "Context-adaptive binary arithmetic coding with fixed-length codewords," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1385–1390, Aug. 2015.
[29] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*. Norwell, MA, USA: Kluwer, 2002.
[30] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000.
[31] A. Kiely and M. Klimesh, "A new entropy coding technique for data compression," Jet Propuls. Lab., California Inst. Technol., Pasadena, CA, USA, Tech. Rep., Jun. 2001. AQ:6
[32] *Lossless Multispectral & Hyperspectral Image Compression*, document CCSDS-120.2-G-1, Green Book, Dec. 2015.
[33] N. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.
[34] X. Wu and N. Memon, "Context-based lossless interband compression-extending CALIC," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 994–1001, Jun. 2000.
[35] E. Augé, J. E. Sánchez, A. Kiely, I. Blanes, and J. Serra-Sagristà, "Performance impact of parameter tuning on the CCSDS-123 lossless multi- and hyperspectral image compression standard," *J. Appl. Remote Sens.*, vol. 7, no. 1, p. 074594, 2013.

Joan Bartrina-Rapesta received the B.Sc., B.E., M.S. (Hons.), and Ph.D. (Hons.) degrees in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2002, 2004, 2006, and 2009, respectively.

Since 2012, he has been an Associate Professor with the Department of Information and Communications Engineering, UAB. He has co-authored numerous papers in journals and conferences and has guided Ph.D. students. His research interests include image coding topics, computing, and transmission.

Dr. Bartrina-Rapesta was a recipient of a Doctoral Fellowship from UAB. He is a Reviewer for magazines and symposiums.

Ian Blanes (S'05–M'11–SM'17) received the B.S., M.S., and Ph.D. degrees in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2007, 2008, and 2010, respectively.

In 2010, he was a Visiting Post-Doctoral Researcher with the Centre National d'Etudes Spatiales, Paris, France. Since 2011, he has been actively involved in the creation of new on-board data compression standards within the framework of the CCSDS Multispectral Hyperspectral Data Compression Working Group. Since 2003, he has been with the Group on Interactive Coding of Images, UAB, where he is currently an Associate Professor. His research interests include data compression in space-borne instruments.

Dr. Blanes was the second-place finisher in the 2007 Best Computer-Science Student Awards by the Spanish Ministry of Education.

Francesc Aulí-Llinàs (S'06–M'08–SM'14) received the B.E. and Ph.D. (*cum laude*) degrees in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2002 and 2006, respectively.

From 2002 to 2015, he was consecutively funded in competitive fellowships, including a Ramón y Cajal Grant awarded with the i3 certificate, when he carried out two post-doctoral research stages with Prof. D. Taubman and M. Marcellin. Since 2016, he has been an Associate Professor (with the Full Professor certificate) with the Department of Information and Communications Engineering, UAB. He has participated and supervised various projects funded by the Spanish government and the European Union. His research interests include image and video coding, computing, and transmission.

Dr. Aulí-Llinàs is a Reviewer for magazines and symposiums.

Joan Serra-Sagristà (S'97–M'05–SM'11) received the Ph.D. degree in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 1999.

From 1997 to 1998, he was with the University of Bonn, Bonn, Germany, funded by DAAD. He has co-authored over 100 publications. He is currently an Associate Professor (hab. Full professor) with the Department of Information and Communications Engineering, UAB. His research interests include data compression, with special attention to image coding for remote sensing applications.

Dr. Serra-Sagristà was a recipient of the Spanish Intensification Young Investigator Award in 2006. He serves or has served as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON IMAGE PROCESSING, and as the Committee Chair for Data Compression Conference.

Victor Sanchez received the M.Sc. degree from the University of Alberta, Edmonton, AB, Canada, in 2003, and the Ph.D. degree from the University of British Columbia, Vancouver, BC, Canada, in 2010.

From 2011 to 2012, he was a Post-Doctoral Researcher with the Video and Image Processing Laboratory, University of California, Berkeley, CA, USA. In 2012, he was a Visiting Lecturer with the Group on Interactive Coding of Images, Universitat Autònoma de Barcelona, Barcelona, Spain. He is currently an Associate Professor with the Department of Computer Science, University of Warwick, Coventry, U.K. His research has been funded by the Consejo Nacional de Ciencia y Tecnologia, Mexico, the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, the FP7 and H2020 programs of the European Union, and the Engineering and Physical Sciences Research Council, U.K. He has authored several technical papers in these areas and co-authored a book (Springer, 2012). His research interests include signal and information processing with applications to multimedia analysis, security, image and video coding, and communications.

Michael W. Marcellin (S'81–M'87–SM'93–F'02) received the B.S. degree in electrical engineering from San Diego State University, San Diego, CA, USA, in 1983, and the M.S. and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, TX, USA, in 1985 and 1987, respectively.

Since 1988, he has been with the University of Arizona, Tucson, AZ, USA, where he currently holds the title of Regents' Professor, and is the International Foundation for Telemetering Chaired Professor. He has authored or co-authored more than 200 publications in these areas. His research interests include digital communication and data storage systems, data compression, and signal processing.

Dr. Marcellin is a recipient of numerous honors, including six teaching awards.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

PLEASE NOTE: We cannot accept new source files as corrections for your paper. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.

Please be aware that authors are required to pay overlength page charges ($200 per page) if the paper is longer than 6 pages. If you cannot pay any or all of these charges please let us know.

This pdf contains 2 proofs. The first half is the version that will appear on Xplore. The second half is the version that will appear in print. If you have any figures to print in color, they will be in color in both proofs.

The "Open Access" option for your paper expires when the paper is published on Xplore in an issue with page numbers. Papers in "Early Access" may be changed to Open Access.

If you have not completed your electronic copyright form (ECF) and payment option please return to Scholar One "Transfer Center." In the Transfer Center you will click on "Manuscripts with Decisions" link. You will see your article details and under the "Actions" column click "Transfer Copyright." From the ECF it will direct you to the payment portal to select your payment options and then return to ECF for copyright submission.

AQ:1 = Please confirm the title and also provide the accessed date for ref. [1].
AQ:2 = Please confirm the title, document no., and organization name for refs. [2]–[4], and [32].
AQ:3 = Please confirm the title, organization name, and year for ref. [5]. Also provide the document no.
AQ:4 = Please confirm the title, organization name, standard no., and year for ref. [24].
AQ:5 = Please confirm the title, organization name, and year for refs. [25] and [26]. Also provide the standard no.
AQ:6 = Please confirm the author name, title, department name, organization name, organization location, and year for ref. [31]. Also provide the report no.

A Lightweight Contextual Arithmetic Coder for On-Board Remote Sensing Data Compression

Joan Bartrina-Rapesta, Ian Blanes, *Senior Member, IEEE*, Francesc Aulí-Llinàs, *Senior Member, IEEE*, Joan Serra-Sagristà, *Senior Member, IEEE*, Victor Sanchez, and Michael W. Marcellin, *Fellow, IEEE*,

***Abstract*— The Consultative Committee for Space Data Systems (CCSDS) has issued several data compression standards devised to reduce the amount of data transmitted from satellites to ground stations. This paper introduces a contextual arithmetic encoder for on-board data compression. The proposed arithmetic encoder checks the causal adjacent neighbors, at most, to form the context and uses only bitwise operations to estimate the related probabilities. As a result, the encoder consumes few computational resources, making it suitable for on-board operation. Our coding approach is based on the prediction and mapping stages of CCSDS-123 lossless compression standard, an optional quantizer stage to yield lossless or near-lossless compression and our proposed arithmetic encoder. For both lossless and near-lossless compression, the achieved coding performance is superior to that of CCSDS-123, M-CALIC, and JPEG-LS. Taking into account only the entropy encoders, fixed-length codeword is slightly better than MQ and interleaved entropy coding.**

***Index Terms*— Arithmetic coding, Consultative Committee for Space Data Systems (CCSDS)-123, lossless and near-lossless coding, remote sensing data compression.**

I. Introduction

REMOTE sensing imagery is becoming an invaluable tool for governments, rescue teams, and aid organizations to manage infrastructure and natural resources, to appraise climate changes, or to give support when natural disasters strike. Since remote sensing images tend to be very large, high-performance compression techniques are of paramount importance.

Let I, J, and K be the number of columns, rows, and components of an image x and let $x_{i,j,k}$ denote a pixel at location (i, j, k) of the image. Such an image is commonly compressed employing one of three regimes: lossless compression, which allows perfect reconstruction of the original image x; lossy compression, which approximates x, introducing an error in the reconstructed image x' that enables a higher compression ratio than possible with lossless compression; or near-lossless compression, which is a particular case of lossy compression where the peak absolute error (PAE) of x' is controlled during the coding process with a tolerance value Δ. Specifically

$$\max_{i,j,k} \left\{ \left| x_{i,j,k} - x'_{i,j,k} \right| \right\} \leq \Delta. \quad (1)$$

Within the Consultative Committee for Space Data Systems (CCSDS) [1], the Multispectral and Hyperspectral Data Compression Working Group is in charge of proposing techniques for remote sensing data compression. Such techniques are mainly developed to be implemented on board, where limited resources are available and low complexity encoders are needed. In 1997, the CCSDS published CCSDS-121.0-B-1 [2], aimed at lossless data compression. In 2005, the CCSDS published CCSDS-122.0-B-1 [3], devised for lossless and lossy compression of monocomponent images based on wavelet transforms. In 2012, the CCSDS published its latest standard, CCSDS-123.0-B-1 [4], focused on lossless compression for multispectral and hyperspectral images based on prediction. Note that to date, there is no CCSDS standard proposal devised to multispectral and hyperspectral images for near-lossless coding. In what follows, we will refer to CCSDS-123.0-B-1 as CCSDS-123.

Lossless and near-lossless coding is an active research topic, as witnessed by the number of recent publications in the last decade [5]–[17]. Some of these contributions, such as [7], [11], and [15]–[17], yield better coding performance than CCSDS-123 for lossless compression but at the expense of an increased computational complexity. Among them, the results provided in [7] can be misleading, since they were obtained using images from the 1997 AVIRIS products, which are known to have undergone an inappropriate calibration [18]. Next three contributions [11], [15], and [16] yield better coding performance than CCSDS-123, but at the expense of an increased computational complexity due to the expensive algorithms applied to improve prediction estimation. The last contribution [17] yields competitive coding performance by including a light spectral regression in the spectral domain, which has a low computational cost.

It is worth noting that none of the previous techniques provides support for near-lossless coding, which is demanded if even better coding performance is requested. Near-lossless coding [5], [6], [8]–[10], [12]–[14] can yield higher compression ratios at a bounded distortion of $\Delta > 0$. Some of the most

Manuscript received December 8, 2016; revised March 30, 2017; accepted April 28, 2017. This work was supported in part by the Spanish Ministry of Economy and Competitiveness and in part by the European Regional Development Fund under Grant TIN2015-71126-R, in part by the Catalan Government under Grant 2014SGR-691, and in part by the Center National d'Etudes Spatiales. *(Corresponding author: Joan Bartrina-Rapesta.)*

J. Bartrina-Rapesta, I. Blanes, F. Aulí-Llinàs, and J. Serra-Sagristà are with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, E-08193 Barcelona, Spain (e-mail: joan.bartrina@uab.cat).

V. Sanchez is with the University of Warwick, Coventry CV4 7AL, U.K.

M.W. Marcellin is with the University of Arizona, Tucson, AZ 85721-0104 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2701837

prominent recent contributions for near-lossless compression are [12], which presents an overview of the latest coding standards for remote sensing, including a near-lossless version of CCSDS-123; [8] and [9], which introduce a near-lossless coding based on wavelet transforms; [14], which goes one step further, proposing an embedded near-lossless coding system based on wavelet transform and prediction coding; and [13], which presents a rate control method for predictive image encoders using the CCSDS-123 predictor. Most of the latest contributions use the CCSDS-123 predictor, since it is suitable for being used on board thanks to its low complexity and high decorrelation efficiency.

After the predictor of CCSDS-123, one can choose between a sample- or a block-adaptive encoder. The sample-adaptive encoder achieves better performance than the block-adaptive encoder when the signal is encoded at more than 1 b/sample. However, because the minimum codeword length of the sample-adaptive encoder is 1 b, block-adaptive encoding yields superior performance for signals that can be encoded at less than 1 b/sample.

Although context-based arithmetic encoders typically obtain excellent coding performance at all rates, they are not included in CCSDS-123 because they can have a high computational demand owing to: 1) probability estimation; 2) the renormalization procedure; and 3) context formation, which are expensive operations and are executed intensively. Despite the computational demand of context-based arithmetic encoders, they are included in some remote sensing coding approaches [6], [19], [20]. Contributions aimed to reduce the computational load by estimating the probability using multiplication-free implementations can be found in the literature: the Q coder [21] approached the interval division by means of lookup tables and the M coder [22] uses a reduced range of possible subinterval sizes together with lookup tables. Some methods based on these approaches have been introduced in different standards [23]–[26]. The operations carried out by the renormalization procedure can be avoided if, instead of producing a single codeword, the coder produces short codewords of fixed length [27], [28]. In particular, [28] presents a context-adaptive binary arithmetic coder with fixed-length coderwords (FLWs) that outperforms the MQ [29] and M coders in terms of coding performance. FLW avoids the renormalization procedure but still estimates probabilities through the division.

It is worth noting that none of the previously mentioned contributions is devised to reduce the computation related to probability estimation and the renormalization simultaneously. In this paper, we propose an arithmetic encoder that: 1) utilizes inexpensive operations to estimate probabilities; 2) does not incorporate the renormalization procedure; and 3) employs a simple context model. It yields strong coding performance at low and high rates for remote sensing images. Our probability estimation procedure builds on that of FLW. Originally, FLW uses a sliding window to estimate the probability of the symbols coded using a division operation. Herein, the sliding window size of FLW is adapted to deal only with power of two sizes, which allows the use of low-complexity bitwise operations and spares the division.

Fig. 1. CCSDS-123 encoding scheme.

The proposed arithmetic coder is incorporated in a lossless and near-lossless coding scheme, providing improved compression performance over current remote sensing image compression techniques. Roughly described, the adopted coding scheme departs from the predictor and mapping included in CCSDS-123 and utilizes a near-lossless quantizer, employs a binary arithmetic coder that operates on a line-by-line and bitplane-by-bitplane basis, introduces a new context model that evaluates (at most) only causal adjacent samples, and uses only bitwise operations to estimate symbol probabilities. Extensive experimental results indicate that our proposed approach improves on CCSDS-123 in terms of lossless compression ratios and also outperforms a near-lossless version of the sample-adaptive and block-adaptive coders of CCSDS-123, JPEG-LS [30] and M-CALIC [6] in terms of lossless and near-lossless coding performances. Comparing only the entropy encoders, FLW is slightly better than MQ and interleaved entropy coder (IEC) [31].

The rest of this paper is structured as follows. Section II briefly reviews the CCSDS-123 coding system and a near-lossless technique for coding systems based on prediction. Section III describes our proposed context-based arithmetic coder with bitwise probability estimation. Section IV describes how our proposed arithmetic coder is incorporated in a coding scheme that uses the predictor of CCSDS-123. Section V presents the experimental results. Section VI concludes this paper.

II. CCSDS-123 and Near-Lossless Compression

A. CCSDS-123

The CCSDS-123 standard, which is limited to encoding samples of $N = 16$ b/pixel/band, can be structured in three stages: *predictor*, *mapper*, and *entropy encoder*. Fig. 1 illustrates the encoding pipeline of CCSDS-123.

In summary, the predictor estimates the value of the current sample $x_{i,j,k}$ using previously scanned samples. This predicted sample is denoted by $\widetilde{x}_{i,j,k}$. The prediction error Λ is computed as

$$\Lambda_{i,j,k} = x_{i,j,k} - \widetilde{x}_{i,j,k} \tag{2}$$

and then mapped to a non-negative integer $\lambda_{i,j,k}$ called the mapped prediction residual. The entropy encoder is in charge of encoding $\lambda_{i,j,k}$ without loss. For the entropy encoder in CCSDS-123, one can choose between a sample- and a block-adaptive encoder.

Further details of the CCSDS-123 stages can be found in [12] and [32].

B. Near-Lossless Compression

For the encoder described above, the decoder can reproduce $x_{i,j,k}$, without loss. In this section, we discuss the addition of a

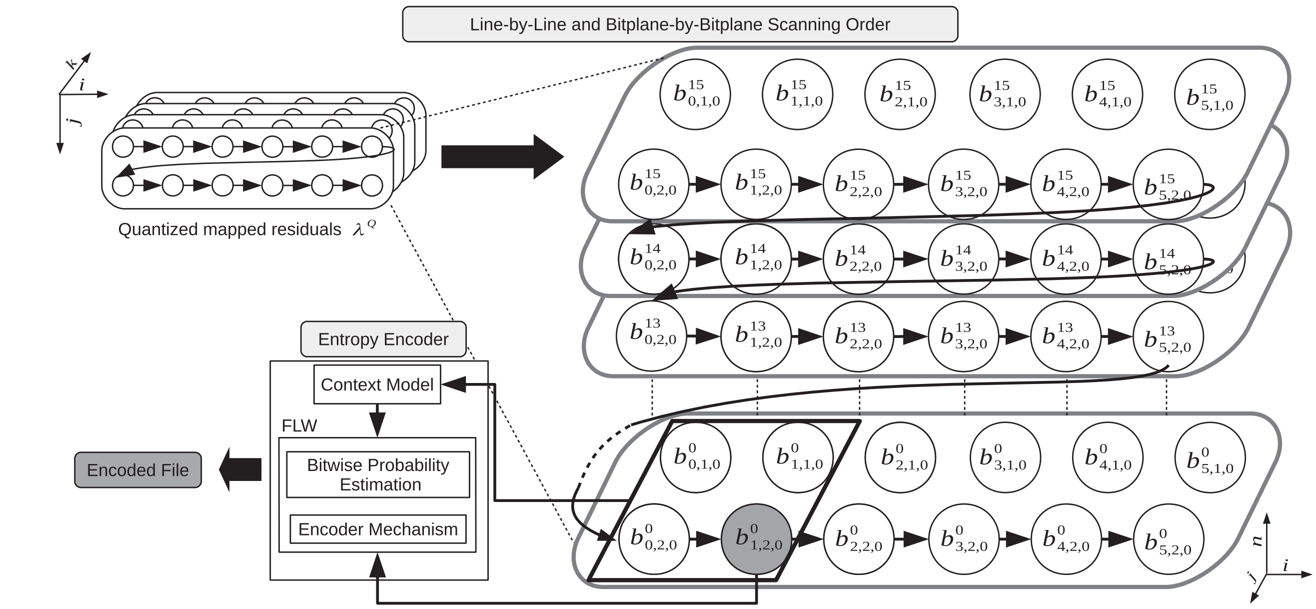


Fig. 2. Illustration of the scanning order and the entropy encoder.

quantizer, which results in higher compression ratios, but at the expense of some loss of fidelity in the decompressed image. The simplest and most effective way to design a prediction-based lossy compression algorithm is to quantize the prediction error $\Lambda_{i,j,k}$ with a quantizer Q, resulting in quantized-then-dequantized version $\widehat{\Lambda}_{i,j,k}$ (and, in consequence, $\widehat{\lambda}_{i,j,k}$). The resulting quantization index is referred to as $\Lambda^{Q}_{i,j,k}$ and its remapped version is denoted by $\lambda^{Q}_{i,j,k}$. Subsequent predictions $\widetilde{x}_{i,j,k}$ are calculated using previous reconstructed (lossy) samples $\widehat{x}_{i,j,k}$, which are obtained by implementing a decoder in the encoder [12], [33]. The decoder creates the reconstructed (lossy) image samples via

$$\widehat{x}_{i,j,k} = \widehat{\Lambda}_{i,j,k} + \widetilde{x}_{i,j,k}. \tag{3}$$

It is worth noting that the errors in the reconstructed pixels are identical to the errors introduced in the prediction errors by the quantizer. That is, $x_{i,j,k} - \widehat{x}_{i,j,k} = \Lambda_{i,j,k} - \widehat{\Lambda}_{i,j,k}$. Thus, the errors in reconstructed pixels can be precisely controlled by controlling the individual quantization errors. This is the basis of "near-lossless compression."

III. Lightweight Binary Arithmetic Coder With Context Model

The entropy encoder presented in this paper works with binary symbols. To this end, we denote the nth bit of the binary representation of $\lambda^{Q}_{i,j,k}$ by $b^{n}_{i,j,k}$, with $N-1 \geq n \geq 0$. Here, N is chosen to provide a sufficient number of bits to represent all the $\lambda^{Q}_{i,j,k}$, being $b^{N-1}_{i,j,k}$ the most significant bit.

To facilitate use with on-board sensors, our proposal processes data in a line-by-line fashion. Once a line is scanned, predicted, and mapped to positive values, it is entropy encoded on a bitplane-by-bitplane basis. The entropy encoder makes use of context model patterns obtained using a context window that contains symbols coded previously to the current symbol. The top left of Fig. 2 displays the quantized and remapped prediction residuals λ^{Q}. The binary representation of these samples is shown on the right, while the bottom left portrays the entropy encoder, which is fed by the current bit to be encoded and its context. The bit to be encoded is shaded in blue, while the context window is framed with a rectangle.

A. Context Model

Let $\mathbf{M}$ be the set of all possible patterns that can occur within the context window, with context $m \in \mathbf{M}$ being a particular realization, resulting in a context index $c \in \mathbf{C} = \{0, \ldots, C-1\}$. These context indices (loosely referred to as contexts in what follows) are determined by a modeling function $F : \mathbf{M} \rightarrow \mathbf{C}$. For each bit b to be coded, a probability model is used, corresponding to its context c. In particular, the probability model estimates the conditional probability $p(b|c) = p(b|F(m))$. After encoding, the probability model is updated with the latest coded bit b. That is, $p(b|c)$ is estimated on the fly. Specifically, our probability model estimates the probability $p(b = 0|c)$. A careful design of the context model is required to obtain high coding efficiency. This task is complicated by the goal of achieving low encoder complexity for the purpose of operating on onboard remote sensing scenarios.

A simple strategy for context modeling employs a context window that contains only the three nearest causal neighbors as depicted in Fig. 2. We consider several choices for the context modeling function F. The first ignores all samples within the context window except the one directly above the sample of interest. This is indicated in Fig. 3(a). Three other choices are shown in Fig. 3(b)–(d). The notations V, H, HV, and HVD are used in Fig. 3, where V (vertical) denotes the sample above the bit to be encoded, H (horizontal) denotes the sample to the left, and D (diagonal) denotes the sample to

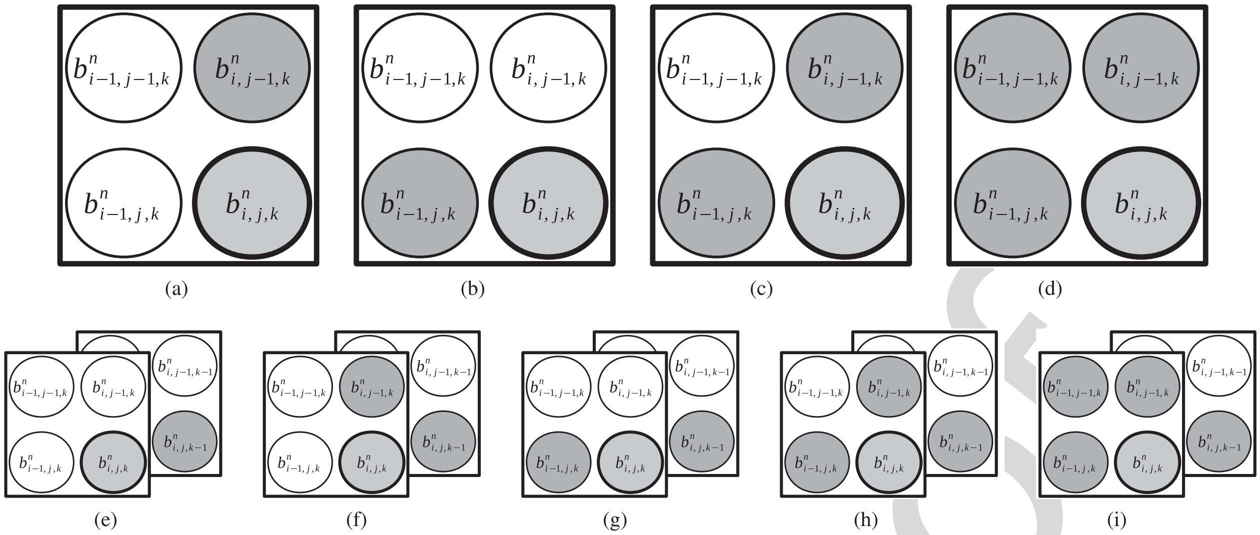


Fig. 3. Illustration of different context models to encode $b^n_{i,j,k}$. (a) V. (b) H. (c) HV. (d) HVD. (e) S. (f) VS. (g) HS. (h) HVS. (i) HVDS.

TABLE I
CONTEXT ASSIGNMENTS FOR THE V, H, HV, AND HVD MODELING FUNCTIONS

c	V	H	HV		HVD		
	$s^n_{i,j-1,k}$	$s^n_{i-1,j,k}$	$s^n_{i,j-1,k}$	$s^n_{i-1,j,k}$	$s^n_{i,j-1,k}$	$s^n_{i-1,j,k}$	$s^n_{i-1,j-1,k}$
0	0	0	0	0	0	0	0
1	1	1	0	1	0	0	1
2			1	0	0	1	0
3			1	1	0	1	1
4					1	0	0
5					1	0	1
6					1	1	0
7					1	1	1

the left and above. To take advantage of dependencies between spectral components, the preceding spectral component $k-1$ can be included in the context window. In this case, S (spectral) denotes the coregistered sample in the previous spectral component. The inclusion of this sample gives rise to five additional modeling functions as shown in Fig. 3(e)–(i). Note that if only samples H and S are employed by the modeling function, only the current scanned line must be stored in memory. For all other modeling functions, the previous and the current lines are necessary.

Rather than the actual bit (from bitplane n) of each neighboring sample, the so-called "significance state" is employed to compute the context c. To this end, let $s^n_{i,j,k}$ denote the significance state of the sample at location i, j, k at bitplane n. A value of 1 indicates that the sample contains a 1 at bitplane n or higher. Table I shows how c is derived from the significance states of the neighbors for the V, H, HV, and HVD modeling functions. The S modeling function results in two states, i.e., $c \in \{0, 1\}$. The VS, HS, HVS, and HVDS modeling functions result in twice the number of states than their counterparts that do not employ S. They are not shown in Table I for the sake of space. The experimental results for all context modeling functions are presented in a subsequent section.

Before finishing this section, we note that the entropy coder and its associated probability models are initialized at the beginning of each bit plane of each component. In particular, the initial probability model for each context is set to a value of $p(b = 0|c) = 0.66$. The probability is biased toward 0 since, as found empirically, bits of higher bitplanes have higher probability of being 0, thus allowing FLW to adapt faster. This, together with the fact that all bitplane data from the current line (and its predecessor, when relevant) are available in the encoder, leads to the conclusion that the bitplanes of the current line can be encoded in parallel. This parallel strategy is not possible in the decoder. The use of significant states in the context formation process requires that bitplanes be decoded sequentially. We note that the probabilities are reset ($p(b = 0|c) = 0.66$) at the beginning of each component without penalizing the coding performance. This is because only 2^{12} symbols are encoded with the default probability value, which on average for the image corpora used, corresponds to the 0.06% of the total symbols per band to be encoded.

B. Bitwise Probability Estimation

As mentioned before, FLW was devised to reduce computational costs through the use of FLWs, which avoids a renormalization operation, but is not aimed to reduce the computational load derived from probability estimation [28].

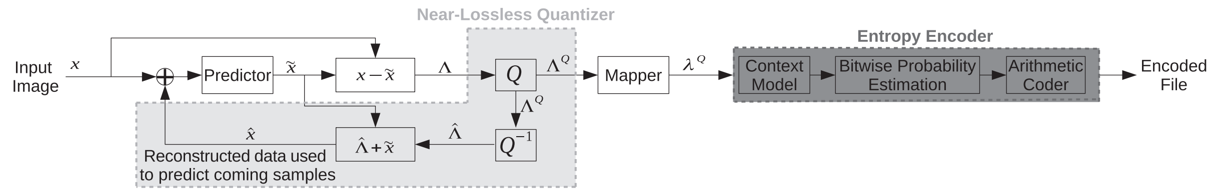


Fig. 4. Adopted coding approach.

TABLE II
MAIN DIFFERENCES BETWEEN CCSDS-123 AND THE ADOPTED APPROACH

	Predictor	Quantizer	Mapper	Data Scanned Supported	Context Model	Entropy Encoder
CCSDS-123	✔	✘	✔	Line or Block	✘	Sample-Adaptive or Block-Adaptive
Proposal	✔	✔	✔	Line	✔	Contextual Binary Arithmetic Coder

For each context c, FLW uses a sliding window of symbols coded with that context. The length of this window varies between $\mathcal{T}$ and $2\mathcal{T}-1$ symbols. The probability estimate is updated once every $\mathcal{V}$ symbols coded, according to

$$p(b=0|c) = \frac{Z \ll \mathcal{B}}{W} \tag{4}$$

with W representing the number of symbols within the window, Z the number of zeroes within the window, and $\mathcal{B}$ the number of bits used to express symbol probabilities. The numerator of the expression is computed by left shifting the binary representation of Z by $\mathcal{B}$ bits. The size of the window is incremented each time a symbol is encoded using context c until $W = 2\mathcal{T}$, at which time the window size is immediately reduced to $\mathcal{T}$ and the number of zeroes within the window is updated according to $Z \leftarrow Z - Z'$ and $Z' \leftarrow Z$, with Z' being the number of zeroes coded during the most recent $\mathcal{T}$ symbols.

In the original approach of FLW as formulated above, $p(b|c)$ is computed via a division operation to achieve maximum accuracy. Such a division may tax the on-board computational resources in a remote sensing scenario. To reduce computational complexity, we propose to estimate the probability through bitwise operations. The substitution of the division by bitwise operations requires that $\mathcal{V} = \mathcal{T}$ and that both are a power of two. This forces the sliding window to contain a power of two symbols, so the probability can be updated using only bit shift operations according to

$$p(b|c) = (Z \ll \mathcal{B}) \gg \log_2(W) \tag{5}$$

where W and Z are then updated through $W \leftarrow W \gg 1$ and $Z \leftarrow Z \gg 1$. Note that this update rule for Z approximates only the number of zeroes in the most recent $\mathcal{T}$ coded samples. Nevertheless, the update can be carried out in the decoder using the same approximation. At the beginning of encoding, the probability is first updated when $\mathcal{V}$ symbols are coded. Subsequently, it is updated every $\mathcal{V}/2$ symbols. The strategy proposed here can be seen as a special case of (4), which was not explored in [28].

Using (5) instead of (4) reduces the flexibility of the arithmetic coder since the updating of the probability estimates and the window size are tied together. The maximum performance achieved with the original formulation of the arithmetic coder proposed in [28] is achieved when the probability estimate is updated every symbol, i.e., $\mathcal{V} = 1$, regardless of the window size. The strategy proposed here provides a significant reduction in complexity with a minor reduction in compression performance. The experimental results provided in Section V indicate that our approach yields highly competitive compression performance.

IV. ADOPTED CODING APPROACH

Although the novel entropy encoder presented here may be incorporated in any coding system, we employ it in the CCSDS-123 coding pipeline. Fig. 4 illustrates the adopted coding approach, which employs the predictor and mapper of CCSDS-123, but adds a near-lossless quantizer (see the yellow block), and substitutes the usual CCSDS-123 encoder by our entropy encoder (see the green block). The circle containing a cross at the left side of Fig. 4 indicates that the input to the predictor is either the original pixel x (when the optional quantization is not present) or the reconstructed pixel $\hat{x}$ (when quantization is present).

The adopted coding scheme is evaluated with a uniform quantizer (UQ) and a uniform scalar deadzone quantizater (USDQ) [29]. The UQ operates over $\Lambda_{i,j,k}$ to obtain a quantization index according to

$$\Lambda^{Q}_{i,j,k} = \text{sign}(\Lambda_{i,j,k}) \left\lfloor \frac{|\Lambda_{i,j,k}| + \Delta}{2\Delta + 1} \right\rfloor \tag{6}$$

where $2\Delta + 1$ is the quantization step size. The operation to reconstruct $\hat{\Lambda}_{i,j,k}$ from its quantization index is given by

$$\hat{\Lambda}_{i,j,k} = \text{sign}\big(\Lambda^{Q}_{i,j,k}\big)(2\Delta + 1)\Lambda^{Q}_{i,j,k}. \tag{7}$$

The UQ is employed in lossless compression techniques such as JPEG-LS, M-CALIC, and 3-D-CALIC [34]. On the other hand, the USDQ quantizes $\Lambda_{i,j,k}$ to obtain a quantization index according to

$$\Lambda^{Q}_{i,j,k} = \text{sign}(\Lambda_{i,j,k}) \left\lfloor \frac{|\Lambda_{i,j,k}|}{\Delta + 1} \right\rfloor \tag{8}$$

TABLE III

SUMMARY OF DATA USED IN THE EXPERIMENTAL RESULTS. SENSOR NAME, ITS ABBREVIATION, THE NUMBER OF IMAGES FROM EACH SENSOR, AND FIRST-ORDER ENTROPIES (IN BITS PER SAMPLE) ON AVERAGE PER SENSOR ARE PROVIDED. THE LAST TWO COLUMNS INDICATE THE PREDICTOR MODE AND THE LOCAL SUM USED FOR EACH SENSOR

Sensor	Abbreviation	Number of images	Entropy	Predictor Mode	Local Sum
Aviris Calibrated	AC	5	9.77	Neighbor Oriented	Full Mode
Aviris Uncalibrated	AU	7	11.21	Neighbor Oriented	Full Mode
Airs	A	9	11.34	Neighbor Oriented	Reduced Mode
Casi	C	2	10.52	Neighbor Oriented	Reduced Mode
Crism	Cr	20	10.69	Column Oriented	Reduced Mode
Hyperion	H	4	9.53	Column Oriented	Reduced Mode
M3	M3	5	9.19	Column Oriented	Reduced Mode
Total / average	——	52	10.41	——	——

where the quantization step is $\Delta + 1$. The operation to reconstruct $\hat{\Lambda}_{i,j,k}$ from its quantization index is expressed as

$$\hat{\Lambda}_{i,j,k} = \text{sign}\left(\Lambda^{Q}_{i,j,k}\right)(\Delta + 1)\Lambda^{Q}_{i,j,k}. \tag{9}$$

Due to its straightforward implementation and excellent performance, the USDQ has been selected for the JPEG 2000 standard [24]. The USDQ partitions the range of input values into intervals all of size Δ, except for the interval that contains zero, which is of size 2Δ. This results in all absolute pixel errors $|x_{i,j,k} - \widehat{x}_{i,j,k}|$ being bounded above Δ for both quantizers.

Table II summarizes the main differences between CCSDS-123 and the adopted coding scheme.

V. EXPERIMENTAL RESULTS

This section presents a set of experiments aimed at the analysis and evaluation of the adopted coding scheme. First, the proposed context modeling functions are evaluated in terms of the conditional entropy of the prediction residual. The bitwise probability estimator is then evaluated via the same performance metric to determine its proper configuration. A variety of binary encoder mechanisms such as IEC, MQ, and FLW are evaluated in terms of their lossless compression performance in conjunction with the proposed context modeling and probability estimation. Finally, the resulting proposed overall approach is compared in terms of lossless and near-lossless compression performances with CCSDS-123, JPEG-LS, and M-CALIC.

For the experiments conducted in this paper, we have selected a set of images[1] collected with different sensors that are included in CCSDS MHDC-WG corpus. The sensor names and their main features are listed in Table III. The average entropy is reported for each image type. The reported values are first-order entropy; they represent the entropy of individual pixels, without accounting for any dependencies among pixels within or between components.

In [35], the impact of different CCSDS-123 parameters that control the operation of the prediction and the entropy encoder was evaluated, suggesting that a correct parameter selection had more impact on the predictor stage than in the entropy encoder stage. Concerning the prediction, the parameters local sum type, prediction mode, the number of prediction bands, and predictor adaption rate were the most critical. Extensive experimental evaluations were conducted to find suitable configurations.

In this paper, leaning on the results in [35] and after conducting an extensive evaluation also, experimental results are produced for the following parameter configuration: the local sum type and predictor mode depend on the acquisition sensor (as indicated in the last two columns of Table III); the number of prediction bands P is set to 3, since it is a good tradeoff between the computational load and the coding performance; and the predictor adaptation rate $\nu_{\max}$ is set to 3, since, in general, it yields the best performance.

For evaluating the performance of context modeling and probability estimation, we employ the conditional entropy of the prediction residuals, as mentioned above. For the work proposed here, binary entropy coding is employed. To yield results with units in bits per pixel, the binary entropies of all bitplanes are added. Since our context model estimates the probability of $p(b = 0|c)$, the conditional entropy of an image (in bits) is computed as

$$H(\lambda^{Q}) = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1}\sum_{k=0}^{K-1}\sum_{n=0}^{15} \times \begin{cases} \log_2\left(p\left(b^{n}_{i,j,k} = 0|c\right)\right) & \text{if } b^{n}_{i,j,k} = 0 \\ \log_2\left(1 - p\left(b^{n}_{i,j,k} = 0|c\right)\right) & \text{if } b^{n}_{i,j,k} = 1 \end{cases} \tag{10}$$

where λ^{Q} denotes the symbols to be entropy coded.

A. Context Modeling Function

The context model is used to select the probability model that is employed to encode the current symbol. In this first experiment, each of the probability models themselves is estimated using the high-performance method given by (4) employing $\mathcal{V} = 1$ and $\mathcal{T} = 2^{12}$, without regard to complexity.

Table IV provides the conditional entropy obtained (in bits per sample) for the different context formations defined in Section III-A, i.e., V, H, HV, HVD, S, VS, HS, HVS, and HVDS. The results from Table III suggest the following.

[1]The images used are available at http://cwe.ccsds.org/sls/docs/sls-dc/123.0-B-Info/TestData

TABLE IV

CONDITIONAL ENTROPY OF THE PREDICTION RESIDUALS (IN BITS PER SAMPLE) FOR THE CONTEXT MODELING FUNCTIONS DENOTED BY V, H, HV, HVD, S, VS, HS, HVS, AND HVDS. RESULTS ARE REPORTED ON AVERAGE FOR DIFFERENT SENSORS AND $\Delta = 0$

	Context Formation								
Sensor	V	H	HV	HVD	S	VS	HS	HVS	HVDS
AC	3.69	3.69	3.68	3.68	3.69	3.68	3.68	3.68	3.67
AU	5.01	5.00	5.00	4.99	5.00	4.99	4.99	4.99	4.99
A	4.24	4.23	4.24	4.24	4.24	4.24	4.24	4.24	4.23
C	4.85	4.85	4.84	4.84	4.85	4.84	4.84	4.83	4.83
Cr	4.15	4.21	4.14	4.14	4.20	4.12	4.19	4.12	4.12
H	4.26	4.26	4.25	4.25	4.29	4.26	4.26	4.25	4.25
M3	2.66	2.70	2.65	2.65	2.70	2.64	2.68	2.63	2.63
Average	4.12	4.14	4.11	4.11	4.14	4.11	4.13	4.10	4.10

TABLE V

CONDITIONAL ENTROPY OF THE PREDICTION RESIDUALS (IN BITS PER SAMPLE) FOR $\Delta = 0$ RESULTING FROM THE MAXIMUM PRECISION AND THE BITWISE PROBABILITY ESTIMATORS. THE V CONTEXT MODEL IS EMPLOYED IN EACH CASE. THE BEST RESULTS FOR EACH STRATEGY ARE REPRESENTED IN BOLD

	division $\mathcal{V} = 1$					bitwise operations $\mathcal{V} = \mathcal{T}$				
$\mathcal{T} =$	2^{16}	2^{14}	2^{12}	2^{10}	2^{8}	2^{16}	2^{14}	2^{12}	2^{10}	2^{8}
AC	3.70	3.69	**3.69**	3.70	3.75	3.72	3.70	**3.69**	3.70	3.77
AU	5.02	5.01	**5.01**	5.01	5.06	5.04	5.02	**5.01**	5.02	5.08
A	4.27	4.26	**4.23**	4.24	4.26	4.32	4.29	**4.24**	4.26	4.28
C	4.94	4.86	**4.85**	4.86	4.89	4.90	4.86	**4.85**	4.86	4.91
Cr	4.16	4.15	**4.15**	4.15	4.18	4.21	4.17	**4.16**	4.16	4.20
H	4.27	4.26	**4.26**	4.27	4.31	4.29	4.27	**4.26**	4.27	4.32
M3	2.69	2.66	**2.66**	2.67	2.70	2.71	2.68	**2.67**	2.68	2.72
Average	4.15	4.13	**4.12**	4.13	4.16	4.17	4.14	**4.13**	4.13	4.18

1) All of the modeling functions provide significant improvements over the pixel entropy reported in Table III.
2) The differences in performance between the modeling functions are generally small.
3) Although the context models H and S yield the worst performance on average, they are the best option when memory resources are severely limited since they need only to store samples from the current line to be encoded.
4) Adding the S sample to a context results in an improvement of only about 0.01 b/sample.
5) The V context obtains a coding benefit of 0.02 b/sample on average with respect to the H context and only adds the previous processed line to its storage requirements.

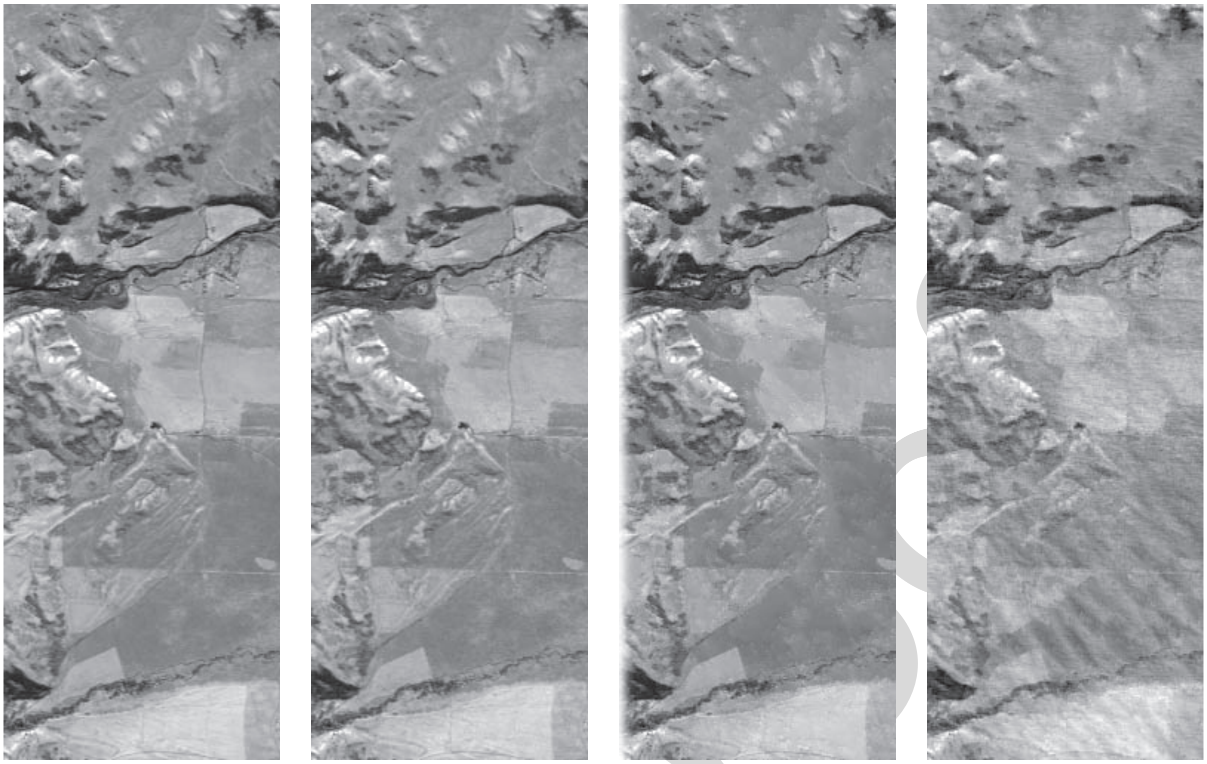
In what follows, we select context model V for further evaluation due to its favorable tradeoff among the performance, memory resources, and computational load.

B. Probability Estimation

This section reports the results obtained by the two different probability estimation strategies discussed in Section III. In particular, Table V reports the conditional entropy of the prediction residuals resulting from the two different probability estimation strategies. In both cases, the V context model is employed. The left of Table V presents results for the maximum precision technique (using division), as defined by (4). These results are shown for different values of $\mathcal{T}$, but $\mathcal{V} = 1$. The right side of Table V presents results for the bitwise strategy, as defined by (5). The same values of $\mathcal{T}$ are explored, but always with $\mathcal{V} = \mathcal{T}$, as required to avoid division. The results suggest that $\mathcal{T} = 2^{12}$ attains the highest performance for both strategies. A larger $\mathcal{T}$ degrades the coding performance because the window may contain symbols that are not correlated with the current one. A smaller $\mathcal{T}$ degrades the coding performance because there are insufficient symbols to reliably estimate the probabilities $p(b|C)$. The results of Table V also indicate that the low-complexity strategy that employs bitwise operations is as competitive as that employing division. Although not tabulated here for the sake of space, these results hold for the other context modeling functions considered in the previous sections.

C. Entropy Coding

We note that the context model and probability estimator proposed here can be used with any entropy encoder that codes binary symbols according to a given probability model, such as MQ, IEC or the adopted FLW. Table VI provides the actual



(a) (b) (c) (d)

Fig. 5. Visual comparison for the "Aviris Calibrated Yellowstone sc00" image. (a) Original. (b) Proposed approach at 0.43 b/sample ($\Delta = 20$). (c) M-CALIC at 0.42 b/sample ($\Delta = 30$). (d) CCSDS-123 at 0.50 b/sample ($\Delta = 80$).

TABLE VI
CODING PERFORMANCE (IN Bits per Sample) OF THE PROPOSED APPROACH USING MQ, IEC, AND FLW ENTROPY ENCODING. ALL RESULTS EMPLOY CONTEXT MODEL V AND BITWISE PROBABILITY ESTIMATION WITH $\mathcal{T} = \mathcal{V} = 2^{12}$

Sensor	MQ	IEC	FLW
AC	3.74	3.72	3.71
AU	5.06	5.04	5.03
A	4.30	4.29	4.28
C	4.90	4.88	4.87
Cr	4.20	4.18	4.18
H	4.31	4.29	4.28
M3	2.71	2.70	2.69
Average	4.17	4.16	4.15

compression results (in bits per sample) obtained using the MQ, IEC, and FLW entropy coders. In each case, the results are obtained with context model V and the bitwise estimator with $\mathcal{T} = \mathcal{V} = 2^{12}$. From these results, we can see that, on average, FLW yields slightly better results than IEC and MQ.

D. Lossless and Near-Lossless Compression

The results reported in this section compare the lossless performance of the proposed approach with those of JPEG-LS, M-CALIC, and CCSDS-123. Additionally, we compare its *near-lossless* performance with those of JPEG-LS and M-CALIC and the implementation of CCSDS-123. Different quantizers have been combined with our proposal and CCSDS-123, to obtain an as fair as possible comparison. In particular, the UQ and the USDQ discussed in Section IV are compared.

M-CALIC and the near-lossless version of CCSDS-123 are considered to be state of the art in terms of compression performance and computational complexity, and JPEG-LS is a standard technique with near-lossless features. All results for the proposed scheme are produced using the FLW arithmetic coder, context model V, and the bitwise probability estimator having $\mathcal{V} = \mathcal{T} = 2^{12}$. The results reported in Table VII indicate that our method outperforms both M-CALIC and CCSDS-123 in terms of lossless coding ($\Delta = 0$) for all sensors. In the near-lossless regime ($\Delta > 0$), the proposed approach outperforms M-CALIC when the USDQ is used and in most cases for the UQ. In particular, M-CALIC obtains slightly better results than our proposal only for images acquired with sensors AIRS and Hyperion when the UQ is used. On the other hand, the proposed system always outperforms the near-lossless extension of CCSDS-123 for both quantizers. In addition, in general, for the same Δ value, the coding performance is better for the USDQ than for UQ. Although achieved bit rates vary widely from image to image, low bit rates can be obtained for all images with a modest value of PAEs (maximum absolute pixel error).

E. Visual Comparison

To evaluate visual performance, we show a region cropped from an image encoded at the "same" bit rate by the proposed approach with the UQ, M-CALIC, and CCSDS-123. For CCSDS-123, we employ the block-adaptive coder since we want to compare the images at a bit rate lower than 1 b/sample.

TABLE VII

LOSSLESS ($\Delta = 0$) AND NEAR-LOSSLESS ($\Delta > 0$) COMPRESSION RESULTS FOR THE PROPOSED APPROACH. FOR COMPARISON, THE RESULTS FOR JPEG-LS, M-CALIC, AND CCSDS-123 ARE INCLUDED. BOTH A UQ AND A USDQ HAVE BEEN USED IN OUR PROPOSED APPROACH AND IN OUR NEAR-LOSSLESS EXTENSION TO CCSDS-123 TO PRODUCE RESULTS FOR $\Delta > 0$. THE RESULTS ARE REPORTED IN BITS PER SAMPLE (LOWER IS BETTER)

		CCSDS-123				JPEG-LS	M-CALIC	Our Proposal	
		with UQ		with USDQ				with UQ	with USDQ
Sensor	Δ values	Sample adaptive	Block adaptive	Sample adaptive	Block adaptive				
AC	$\Delta = 0$	3.73	3.91	3.73	3.91	6.41	3.87	**3.71**	**3.71**
	$\Delta = 10$	1.20	0.97	1.20	0.94	2.45	0.76	0.62	**0.60**
	$\Delta = 20$	1.10	0.74	1.10	0.72	1.81	0.50	0.38	**0.36**
	$\Delta = 30$	1.07	0.64	1.07	0.63	1.48	0.40	0.28	**0.27**
AU	$\Delta = 0$	5.06	5.23	5.06	5.23	7.47	5.13	**5.03**	**5.03**
	$\Delta = 10$	1.69	1.68	1.85	1.78	3.41	1.46	**1.39**	1.52
	$\Delta = 20$	1.35	1.19	1.40	1.21	2.68	0.95	**0.87**	0.90
	$\Delta = 30$	1.24	0.98	1.26	0.99	2.30	0.73	**0.65**	0.67
A	$\Delta = 0$	4.29	4.48	4.29	4.48	6.85	4.28	**4.27**	**4.27**
	$\Delta = 10$	1.23	1.12	1.18	0.96	2.62	0.73	0.76	**0.62**
	$\Delta = 20$	1.10	0.75	1.07	0.65	1.86	0.41	0.39	**0.30**
	$\Delta = 30$	1.06	0.63	1.05	0.56	1.50	0.31	0.28	**0.22**
C	$\Delta = 0$	4.97	5.15	4.97	5.15	6.79	4.91	**4.87**	**4.87**
	$\Delta = 10$	1.47	1.48	1.51	1.47	2.64	1.12	**1.10**	**1.10**
	$\Delta = 20$	1.25	1.06	1.25	1.03	1.94	0.68	0.67	**0.64**
	$\Delta = 30$	1.17	0.88	1.18	0.87	1.58	0.52	0.49	**0.48**
Cr	$\Delta = 0$	4.40	4.50	4.40	4.50	5.10	6.91	**4.18**	**4.18**
	$\Delta = 10$	1.64	1.66	1.63	1.42	1.83	2.75	1.26	**0.99**
	$\Delta = 20$	1.43	1.34	1.39	1.05	1.47	2.02	0.92	**0.64**
	$\Delta = 30$	1.34	1.17	1.30	0.89	1.29	1.64	0.76	**0.50**
H	$\Delta = 0$	4.37	4.57	4.37	4.57	6.24	4.80	**4.28**	**4.28**
	$\Delta = 10$	1.38	1.45	1.21	1.08	2.74	1.02	1.06	**0.68**
	$\Delta = 20$	1.24	1.19	1.09	0.74	2.06	0.52	0.77	**0.35**
	$\Delta = 30$	1.18	1.04	1.06	0.63	1.68	0.33	0.62	**0.25**
M	$\Delta = 0$	2.81	2.97	2.81	2.97	4.24	5.18	**2.69**	**2.69**
	$\Delta = 10$	1.26	1.21	1.11	0.74	1.38	1.32	0.76	**0.33**
	$\Delta = 20$	1.17	1.01	1.07	0.60	1.20	0.78	0.56	**0.20**
	$\Delta = 30$	1.14	0.89	1.06	0.55	1.00	0.53	0.46	**0.15**
Average	$\Delta = 0$	4.23	4.40	4.23	4.40	6.16	5.01	**4.15**	**4.15**
	$\Delta = 10$	1.41	1.37	1.39	1.20	2.44	1.31	1.00	**0.83**
	$\Delta = 20$	1.23	1.04	1.20	0.86	1.86	0.84	0.65	**0.48**
	$\Delta = 30$	1.17	0.89	1.14	0.73	1.55	0.64	0.51	**0.36**

We note that none of the schemes compared here includes precise rate control. For this reason, we have employed the following methodology: 1) encode an image using a variety of different quantization step sizes for each compression scheme and 2) choose those encoded images that yield bit rates as close as possible for the three algorithms. We note that a close match was not obtained in the case of CCSDS-123, so a step size was chosen to afford a higher bit rate than that of the proposed approach, thus giving an advantage to CCSDS-123 in terms of visual performance.

The results of this process are shown in Fig. 5 for a crop from component 122 of the image "Aviris Calibrated Yellowstone sc00." The bit rates obtained are 0.43, 0.42, and 0.50 for the proposed approach, M-CALIC, and CCSDS-123, respectively. The reader is invited to zoom in to see the specific visual artifacts arising from the different compression schemes. Fig. 5 indicates that the image obtained by the proposed approach has higher visual quality than those by (near lossless) CCSDS-123 and M-CALIC. In particular, the proposed approach preserves edges and textures very well, while M-CALIC results in smoothness and loss of texture. CCSDS-123 also removes texture, but also introduces an annoying "banding" effect, due to the high step size required to reach 0.50 b/sample.

VI. CONCLUSION

This paper proposes an entropy encoder based on an efficient definition for a context model and the associated strategy to estimate probabilities for use in a fixed-length arithmetic encoder using low-cost bitwise operations. These contributions are incorporated in a coding approach that employs the predictor included in CCSDS-123. A near-lossless quantizer has also been deployed. The entropy encoder works on a line-by-line and bitplane-by-bitplane scanning order. The experimental results indicate that the use of a single neighbor for the context formation is enough to properly exploit the contextual information in the arithmetic encoder and that it is possible to estimate the probability employing bitwise operations without penalizing the coding efficiency. Further results indicate that, on average, our proposal improves the current standard version of CCSDS-123 for lossless coding by more than 0.1 b/sample. Compared with M-CALIC, our proposal provides an average improvement of 0.86 b/sample for lossless, whereas for near-lossless, the benefit ranges from 0.13 to 0.31 b/sample, depending on the allowed PAE.

REFERENCES

[1] *Consultative Committee for Space Data Systems (CCSDS)*. [Online]. Available: http://www.ccsds.org AQ:1

[2] *Lossless Data Compression*, document CCSDS-121.0-B-1, Blue Book, May 1997. AQ:2

[3] *Image Data Compression*, document CCSDS-122.0-B-1, Blue Book, Nov. 2005.

[4] *Lossless Multispectral & Hyperspectral Image Compression*, document CCSDS-123.0-B-1, Blue Book, May 2012.

[5] *JPEG-LS Lossless and Near-Lossless Compression for Continuous-Tone Still Images*, document, ISO/IEC, 1999. AQ:3

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

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

[8] G. Carvajal, B. Penna, and E. Magli, "Unified lossy and near-lossless hyperspectral image compression based on JPEG 2000," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 593–597, Oct. 2008.

[9] C.-W. Chen, T.-C. Lin, S.-H. Chen, and T.-K. Truong, "A near lossless wavelet-based compression scheme for satellite images," in *Proc. WRI World Congr. Comput. Sci. Inf. Eng.*, vol. 6. Mar. 2009, pp. 528–532.

[10] S.-C. Tai, T.-M. Kuo, C.-H. Ho, and T.-W. Liao, "A near-lossless compression method based on CCSDS for satellite images," in *Proc. Int. Symp. Comput., Consum. Control*, Jun. 2012, pp. 706–709.

[11] J. Song, Z. Zhang, and X. Chen, "Lossless compression of hyperspectral imagery via RLS filter," *Electron. Lett.*, vol. 49, no. 16, pp. 992–994, Aug. 2013.

[12] I. Blanes, E. Magli, and J. Serra-Sagrista, "A tutorial on image compression for optical space imaging systems," *IEEE Geosci. Remote Sens. Mag.*, vol. 2, no. 3, pp. 8–26, Sep. 2014.

[13] D. Valsesia and E. Magli, "A novel rate control algorithm for onboard predictive coding of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6341–6355, Oct. 2014.

[14] J. Beerten, I. Blanes, and J. Serra-Sagristà, "A fully embedded two-stage coder for hyperspectral near-lossless compression," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1775–1779, Aug. 2015.

[15] W. Jiaji, K. Wanqiu, M. Jarno, and H. Bormin, "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2194–2198, Dec. 2015.

[16] G. Fang and G. Shuxu, "Lossless compression of hyperspectral images using conventional recursive least-squares predictor with adaptive prediction bands," *J. Appl. Remote Sens.*, vol. 10, no. 1, p. 015010, Feb. 2016.

[17] N. Amrani, J. Serra-Sagristà, V. Laparra, M. W. Marcellin, and J. Malo, "Regression wavelet analysis for lossless coding of remote-sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5616–5627, Sep. 2016.

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

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

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

[21] M. J. Slattery and J. L. Mitchell, "The Qx-coder," *IBM J. Res. Develop.*, vol. 42, no. 6, pp. 767–784, Nov. 1998.

[22] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 620–636, Jul. 2003.

[23] *Information Technology—Lossy/Lossless Coding of Bi-Level Images*, ISO/IEC Standard 14 492:2001, Dec. 2001.

[24] *Information Technology—JPEG 2000 Image Coding System—Part 1: Core Coding System*, ISO/IEC, Standard 15 444-1, Dec. 2000. AQ:4

[25] *H.264—Advanced Video Coding for Generic Audiovisual Services*, ITU-T, Standard, May 2003.

[26] *H.265—High Efficiency Video Coding*, ITU-T, Standard, Apr. 2015. AQ:5

[27] M. D. Reavy and C. G. Boncelet, "An algorithm for compression of bilevel images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 669–676, May 2001.

[28] F. Aulí-Llinàs, "Context-adaptive binary arithmetic coding with fixed-length codewords," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1385–1390, Aug. 2015.

[29] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*. Norwell, MA, USA: Kluwer, 2002.

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

[31] A. Kiely and M. Klimesh, "A new entropy coding technique for data compression," Jet Propuls. Lab., California Inst. Technol., Pasadena, CA, USA, Tech. Rep., Jun. 2001. AQ:6

[32] *Lossless Multispectral & Hyperspectral Image Compression*, document CCSDS-120.2-G-1, Green Book, Dec. 2015.

[33] N. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.

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

[35] E. Augé, J. E. Sánchez, A. Kiely, I. Blanes, and J. Serra-Sagristà, "Performance impact of parameter tuning on the CCSDS-123 lossless multi- and hyperspectral image compression standard," *J. Appl. Remote Sens.*, vol. 7, no. 1, p. 074594, 2013.

Joan Bartrina-Rapesta received the B.Sc., B.E., M.S. (Hons.), and Ph.D. (Hons.) degrees in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2002, 2004, 2006, and 2009, respectively.

Since 2012, he has been an Associate Professor with the Department of Information and Communications Engineering, UAB. He has co-authored numerous papers in journals and conferences and has guided Ph.D. students. His research interests include image coding topics, computing, and transmission.

Dr. Bartrina-Rapesta was a recipient of a Doctoral Fellowship from UAB. He is a Reviewer for magazines and symposiums.

Ian Blanes (S'05–M'11–SM'17) received the B.S., M.S., and Ph.D. degrees in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2007, 2008, and 2010, respectively.

In 2010, he was a Visiting Post-Doctoral Researcher with the Centre National d'Etudes Spatiales, Paris, France. Since 2011, he has been actively involved in the creation of new on-board data compression standards within the framework of the CCSDS Multispectral Hyperspectral Data Compression Working Group. Since 2003, he has been with the Group on Interactive Coding of Images, UAB, where he is currently an Associate Professor. His research interests include data compression in space-borne instruments.

Dr. Blanes was the second-place finisher in the 2007 Best Computer-Science Student Awards by the Spanish Ministry of Education.

Francesc Aulí-Llinàs (S'06–M'08–SM'14) received the B.E. and Ph.D. (*cum laude*) degrees in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2002 and 2006, respectively.

From 2002 to 2015, he was consecutively funded in competitive fellowships, including a Ramón y Cajal Grant awarded with the i3 certificate, when he carried out two post-doctoral research stages with Prof. D. Taubman and M. Marcellin. Since 2016, he has been an Associate Professor (with the Full Professor certificate) with the Department of Information and Communications Engineering, UAB. He has participated and supervised various projects funded by the Spanish government and the European Union. His research interests include image and video coding, computing, and transmission.

Dr. Aulí-Llinàs is a Reviewer for magazines and symposiums.

Joan Serra-Sagristà (S'97–M'05–SM'11) received the Ph.D. degree in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 1999.

From 1997 to 1998, he was with the University of Bonn, Bonn, Germany, funded by DAAD. He has co-authored over 100 publications. He is currently an Associate Professor (hab. Full professor) with the Department of Information and Communications Engineering, UAB. His research interests include data compression, with special attention to image coding for remote sensing applications.

Dr. Serra-Sagristà was a recipient of the Spanish Intensification Young Investigator Award in 2006. He serves or has served as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON IMAGE PROCESSING, and as the Committee Chair for Data Compression Conference.

Victor Sanchez received the M.Sc. degree from the University of Alberta, Edmonton, AB, Canada, in 2003, and the Ph.D. degree from the University of British Columbia, Vancouver, BC, Canada, in 2010.

From 2011 to 2012, he was a Post-Doctoral Researcher with the Video and Image Processing Laboratory, University of California, Berkeley, CA, USA. In 2012, he was a Visiting Lecturer with the Group on Interactive Coding of Images, Universitat Autònoma de Barcelona, Barcelona, Spain. He is currently an Associate Professor with the Department of Computer Science, University of Warwick, Coventry, U.K. His research has been funded by the Consejo Nacional de Ciencia y Tecnologia, Mexico, the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, the FP7 and H2020 programs of the European Union, and the Engineering and Physical Sciences Research Council, U.K. He has authored several technical papers in these areas and co-authored a book (Springer, 2012). His research interests include signal and information processing with applications to multimedia analysis, security, image and video coding, and communications.

Michael W. Marcellin (S'81–M'87–SM'93–F'02) received the B.S. degree in electrical engineering from San Diego State University, San Diego, CA, USA, in 1983, and the M.S. and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, TX, USA, in 1985 and 1987, respectively.

Since 1988, he has been with the University of Arizona, Tucson, AZ, USA, where he currently holds the title of Regents' Professor, and is the International Foundation for Telemetering Chaired Professor. He has authored or co-authored more than 200 publications in these areas. His research interests include digital communication and data storage systems, data compression, and signal processing.

Dr. Marcellin is a recipient of numerous honors, including six teaching awards.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

PLEASE NOTE: We cannot accept new source files as corrections for your paper. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.

Please be aware that authors are required to pay overlength page charges ($200 per page) if the paper is longer than 6 pages. If you cannot pay any or all of these charges please let us know.

This pdf contains 2 proofs. The first half is the version that will appear on Xplore. The second half is the version that will appear in print. If you have any figures to print in color, they will be in color in both proofs.

The "Open Access" option for your paper expires when the paper is published on Xplore in an issue with page numbers. Papers in "Early Access" may be changed to Open Access.

If you have not completed your electronic copyright form (ECF) and payment option please return to Scholar One "Transfer Center." In the Transfer Center you will click on "Manuscripts with Decisions" link. You will see your article details and under the "Actions" column click "Transfer Copyright." From the ECF it will direct you to the payment portal to select your payment options and then return to ECF for copyright submission.

AQ:1 = Please confirm the title and also provide the accessed date for ref. [1].
AQ:2 = Please confirm the title, document no., and organization name for refs. [2]–[4], and [32].
AQ:3 = Please confirm the title, organization name, and year for ref. [5]. Also provide the document no.
AQ:4 = Please confirm the title, organization name, standard no., and year for ref. [24].
AQ:5 = Please confirm the title, organization name, and year for refs. [25] and [26]. Also provide the standard no.
AQ:6 = Please confirm the author name, title, department name, organization name, organization location, and year for ref. [31]. Also provide the report no.