Regression Wavelet Analysis for Near-lossless Remote Sensing Data Compression

Sara Álvarez-Cortés, Joan Serra-Sagristà, *Senior Member, IEEE*, Joan Bartrina-Rapesta, Michael Marcellin, *Fellow, IEEE*

***Abstract*—Regression Wavelet Analysis (RWA) is one of the current state-of-the-art lossless compression techniques for remote sensing data. This paper presents the first regression-based near-lossless compression method. It is built upon RWA, a quantizer, and a feedback loop to compensate the quantization error. Our near-lossless RWA (NLRWA) proposal can be followed by any entropy coding technique. Here, NLRWA is coupled with a bitplane-based coder that supports progressive decoding. This successfully enables gradual quality refinement and lossless and near-lossless recovery. A smart strategy for selecting the NLRWA quantization steps is also included. Experimental results show that the proposed scheme outperforms state-of-the-art lossless and near-lossless compression methods in terms of compression ratios and quality retrieval.**

***Index Terms*—lossless and near-lossless compression, pyramidal multiresolution scheme, regression wavelet analysis, remote sensing data compression.**

I. Introduction

SATELLITES carry on-board hyperspectral sensors that collect enormous volumes of data, with large spectral and spatial resolutions. Recording this information places excessive demands on bandwidth and on on-board storage capacity, meaning that part of the data could go uncaptured, or that part of the acquired data could be immediately discarded without further processing. Data compression has proven to be a convenient means to mitigate these issues, and meet the requirements of space missions. Different remote sensing data compression techniques provide lossless, lossy, and/or near-lossless recovery.

Lossless coding ensures perfect reconstruction at the price of low compression ratios. Multiband Context-based Adaptive Lossless Image Coding (M-CALIC) [1] is one of the most renowned methods. It is a context-based adaptive system that uses a nonlinear predictor. In 2012, the Consultive Committee for Space Data Systems (CCSDS) proposed the standard CCSDS-123.0-B-1 [2], which is formed by a predictor, a mapper function, and an entropy encoder. It entails minor computational cost and exploits the redundancy within 3-D spatial and spectral neighborhoods of pixels. During the last decade, techniques based upon recursive least squares methods have been presented [3]–[9]. Among them, Regression Wavelet Analysis (RWA) [3] yields similar coding performance at lower execution times. Indeed it is at least 398, 70 [10] and 17 [4] times faster when compared respectively to [5], [6] and [4], which are some of the techniques that provide the lowest lossless bitrates nowadays. Regarding the computational complexity, RWA benefits from the low complexity and simple implementation of the integer Haar Discrete Wavelet Transform (DWT). References [3], [10]–[12] provide an extensive study of the arithmetic operations required for each operation involved in the RWA scheme, concluding that its number of operations is lower than that of other efficient transforms such as the reversible Karhunen-Loève Transform. Besides, RWA is suitable to incorporate several light regression models. Consequently, it obtains one of the best coding performance trade-off concerning compression ratio and computational cost.

Lossy techniques enable high compression ratios at the expense of allowing loss in decoding. Commonly, lossy pipelines apply quantization prior to an entropy encoder, and a rate control allocation stage afterwards. Lossy wavelet-based techniques, such as JPEG 2000 [13], are well-known for attaining excellent performance in terms of mean squared error (MSE). Unfortunately, it does not provide any guarantees on the error incurred by individual pixels. Other major lossy contributions [14]–[16] extend the CCSDS-123.0-B-1 framework. The fast and lightweight rate control algorithm of Valsesia et al. [14] achieves comparable or better coding performance than [15] and [16], while decreasing the computational complexity.

Near-lossless compression aims at higher compression ratios than lossless methods by allowing some loss of fidelity in reconstruction. They bound the l_∞-norm -equivalently, the peak absolute error (PAE) or maximum absolute distortion (MAD)- via setting an error tolerance value Λ. This user-specified parameter sets the maximum admissible absolute error so that PAE $\leq \Lambda$ provides a guaranteed bound on the error incurred by individual pixels. Near-lossless compression is used in remote sensing applications such as appraisals of climate changes, natural resources and disasters, and also for farming and military purposes.

Near-lossless techniques can be classified into: prediction-based coding followed by quantization; and two-stage near-lossless coders.

Prediction-based followed by quantization techniques compute first a prediction of a pixel's value from previously

Manuscript accepted September 1, 2019. This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO) and by the European Regional Development Fund (FEDER) under Grant RTI2018-095287-B-I00, by the Catalan Government under Grant 2017SGR-463 and by Universitat Autònoma de Barcelona under Grant UAB-PIF-472/2015.

S. Álvarez-Cortés, J. Serra-Sagristà and J. Bartrina-Rapesta are with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain (e-mail: salvarez@deic.uab.cat). M. W. Marcellin is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721-0104, USA.

encoded pixels. They provide near-lossless compression by introducing a quantization feedback loop and by including the corresponding reconstruction function in the coder. M-CALIC provides near-lossless compression and is one of the most relevant techniques in this category. Other prominent and low-complexity techniques, which may be amenable for on-board computation, are the two near-lossless adaptations of the lossless compression standard CCSDS-123.0-B-1, henceforth referred to as NLCCSDS-123 [17], and CCSDS-123-AC [18]. Both NLCCSDS-123 and CCSDS-123-AC rely on the predictor and mapper of CCSDS-123.0-B-1. CCSDS-123-AC includes a lightweight contextual arithmetic encoder that defines a context model and computes the probabilities that will be used by a fixed-length arithmetic encoder. CCSDS-123-AC improves the performance of both NLCCSDS-123 and M-CALIC, being the state-of-the-art in this compression modality.

Two-stage near-lossless coders generally include a first stage that generates a lossy reconstructed scene (image), whereas the second stage quantizes the difference between the reconstructed and the original scene and finally encodes the quantized signal with an entropy encoder. The performance of this strategy strongly depends on the distortion introduced in the lossy stage. Methods such as [19] and [20] do not include any selection criterion to determine the lossy bitrate. Other methods such as [21], [22] and [23] achieve competitive coding performance after using computationally expensive iterative approaches to identify the optimal lossy bitrate. The embedded two-stage near-lossless coder [23] yields the state-of-the-art compression performance in this category.

This paper presents a low-complexity approach that constitutes the first near-lossless technique based on regression in a pyramidal multiresolution scheme and specifically on RWA. Although some recent publications [10]–[12], [24]–[29] introduce several modifications of RWA [3] –mainly in the lossless regime– and investigate their compression performance, no near-lossless implementation has ever been presented. Our proposal 1) yields the same lossless performance as the original RWA; 2) employs a novel and low-cost strategy to select the quantization steps for near-lossless reconstruction; and 3) can be followed by any entropy coder. Here, we report results when employing JPEG 2000. This allows progressive lossy-to-lossless/near-lossless transmission, while minimizing the error propagation and optimizing both signal-to-noise ratio (SNR) and PAE performance.

The paper is organized as follows: Section II introduces our novel near-lossless scheme. Section III describes the mathematical derivation that allows us to control the PAE, and puts forward a smart criterion for selecting the quantization steps. Section IV presents experimental results and provides comparison to other state-of-the-art techniques. Finally, Section V brings forward our conclusions.

II. Near-lossless Regression Wavelet Analysis

Regression Wavelet Analysis [3] exploits the correlation of a scene in the spectral dimension. It is composed of two sequential blocks: a simple integer Haar Discrete Wavelet Transform followed by a regression operation. This second block is performed through an Ordinary Least-Square (OLS) method which predicts the Haar wavelet details from the Haar approximations of the same decomposition level. The two blocks are repeated pyramidally, level by level, from the first to the highest possible decomposition level $L = \lceil log_2(z) \rceil$, where z denotes the number of spectral components of the original scene.

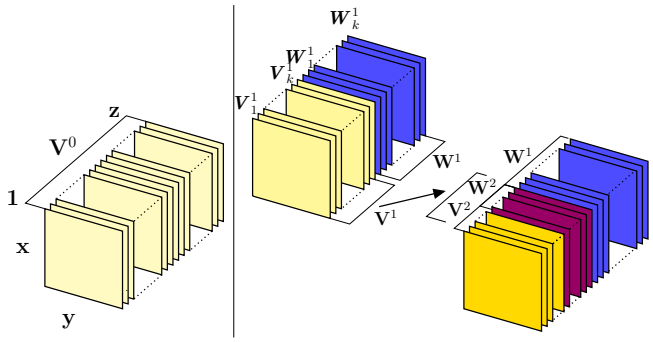

Figure 1: *Left:* original scene $\mathbf{V}^0$ with z spectral components, and $x \cdot y$ spatial samples. *Right:* first and second DWT decomposition levels, i.e., respectively $\mathbf{DWT}(\mathbf{V}^0, 1) = \left(\mathbf{V}^1, \mathbf{W}^1\right)$, and $\mathbf{DWT}(\mathbf{V}^1, 1) = \left(\mathbf{V}^2, \mathbf{W}^2\right)$ or, equivalently, $\mathbf{DWT}(\mathbf{V}^0, 2) = \left(\mathbf{V}^2, \mathbf{W}^2, \mathbf{W}^1\right)$, where $k = z \cdot 2^{-1}$.

As explained in detail below, our near-lossless RWA (NL-RWA) adaptation begins by applying Haar-DWT at the highest decomposition level (first operation). The approximation and detail components at this level constitute the input signal for the regression (second operation). After the regression/prediction computation, the difference between the original Haar wavelet details and their predictions is obtained. The result is known as the residuals. The residuals are then quantized (third operation) and then dequantized and passed through one level of inverse Haar wavelet transform (fourth operation). The resulting approximations form the input to the linear regression for the next lower level. Computation proceeds from the second to the fourth operation iteratively from level $L-1$ down to level 1. A detailed explanation is provided next.

A. First operation: Integer Haar Discrete Wavelet Transform

Let us consider a scene $\mathbf{V}^0 \in \mathbb{R}^{m \times z}$ with z spectral components and $m = x \cdot y$ spatial samples, where z is a power of two (a suitable boundary handling procedure is used otherwise), $\mathbf{V}^0 = \left[\mathbf{V}^0(1), \ldots, \mathbf{V}^0(z)\right]$ and $\mathbf{V}^0(i) = \mathbf{V}^0_i \in \mathbb{R}^{m \times 1}$. An integer Discrete Wavelet Transform (DWT) decomposition level on the original scene is denoted as

$$\mathbf{DWT}(\mathbf{V}^0, 1) = \left(\mathbf{V}^1, \mathbf{W}^1\right). \tag{1}$$

Here, $\mathbf{V}^1 \in \mathbb{R}^{m \times (z \cdot 2^{-1})}$ and $\mathbf{W}^1 \in \mathbb{R}^{m \times (z \cdot 2^{-1})}$ refer, respectively, to the half-resolution DWT approximation and detail components at the first decomposition level. A second decomposition level can be computed on the approximation components $\mathbf{V}^1$, maintaining the details $\mathbf{W}^1$ unchanged. After the second decomposition level, the transformed scene is

composed of the approximations $\mathbf{V}^2$ at the second level and the details from the first $\mathbf{W}^1$ and second $\mathbf{W}^2$ levels. This process is iteratively applied until level L. The application of L decomposition levels to the original scene $\mathbf{V}^0$ is denoted as

$$\mathbf{DWT}(\mathbf{V}^0, L) = \left(\mathbf{V}^L, (\mathbf{W}^j)^{1\leq j\leq L}\right). \tag{2}$$

See Fig. 1 for a graphical explanation.

In order to secure perfect reconstruction, a lifting scheme [30] is employed. The Haar-DWT that considers this scheme corresponds to the Haar S-Transform, and its forward equations at level j are

$$\textit{Forward:} \left\{ \begin{array}{rcl} \mathbf{W}_i^j = & \mathbf{V}_{2i}^{j-1} - \mathbf{V}_{2i-1}^{j-1} \\ \mathbf{V}_i^j = & \mathbf{V}_{2i-1}^{j-1} + \lfloor \frac{1}{2}\mathbf{W}_i^j \rfloor, \end{array} \right. \tag{3}$$

where $i \in \mathbb{I} = \{1, .., z \cdot 2^{-j}\}$ corresponds to the spectral component.

B. Second operation: Ordinary Least-Squares Method

An Ordinary Least-Squares [31] problem (OLS) is solved to obtain the prediction of the details at level j, $\widehat{\mathbf{W}}^j$. It consists of a regression operation that minimizes the sum of the squares of the distances between the original and the predicted details:

$$\underset{\boldsymbol{\beta}^j}{\operatorname{argmin}} \|\mathbf{W}^j - \widehat{\mathbf{W}}^j\|_2. \tag{4}$$

The OLS is solved by applying a regression model. Two regression models are considered here: Maximum and Parsimonious.

The *Maximum* model [3] delivers the most accurate predictions, but uses all the $k{=}z{\cdot}2^{-j}$ approximation components at each level j to predict each detail component. The predicted detail component i at level j, $\widehat{\mathbf{W}}_i^j$, is computed as

$$\widehat{\mathbf{W}}_i^j = \boldsymbol{\beta}_{i,0}^j + \boldsymbol{\beta}_{i,1}^j \mathbf{V}_1^j + \cdots + \boldsymbol{\beta}_{i,k}^j \mathbf{V}_k^j. \tag{5}$$

The regression coefficients $\boldsymbol{\beta}^j$ need to be stored and transmitted as side information (SI), which amounts to $\frac{z^2}{3}(1 - \frac{1}{2^{2j}}) + z(1 - \frac{1}{2^j})$ [3].

The *Parsimonious* model [12] does not employ all the $z{\cdot}2^{-j}$ approximation components at level j to obtain the regression coefficients, but, at most, $2r + 1$. r is a natural number and specifies the number of previous and subsequent spectral neighbors of the component i considered by the OLS operation. The SI size is thus smaller. Specifically, the number of $\boldsymbol{\beta}_i^j$ involved in each prediction equates to $\min\{(2r + 2), (z{\cdot}2^{-j} + 1)\}$.

For both models, the regression coefficients $\boldsymbol{\beta}^j$ are computed for each individual scene.

Next, the prediction residuals are computed as

$$\mathbf{R}^j = \mathbf{W}^j - \textit{round}(\widehat{\mathbf{W}}^j). \tag{6}$$

The predictions $\widehat{\mathbf{W}}^j$ are rounded to operate with integer values.

C. Third operation: USDZ Quantization

Each component i of the residuals at each level j, $\mathbf{R}_i^j$, is quantized with a Uniform Scalar Dead-Zone Quantizer (USDZQ), which delivers a symmetric behavior around 0. Let $\mathbf{c}_i^j$ be a coefficient of component i of the residuals at level j. The quantized coefficient is obtained as follows:

$$\tilde{\mathbf{c}}_i^j = \text{USDZQ}\left(\mathbf{c}_i^j\right) = \text{sign}(\mathbf{c}_i^j) \left\lfloor \frac{|\mathbf{c}_i^j|}{\Delta_i^j} \right\rfloor. \tag{7}$$

Δ_i^j refers to the quantization step for component $\mathbf{R}_i^j$, and *sign* is a function that extracts the sign value of coefficient $\mathbf{c}_i^j$.

Analogously, the dequantized coefficient can be described as $\overline{\mathbf{c}}_i^j$. The dequantization is computed as

$$\overline{\mathbf{c}}_i^j = \text{USDZQ}^{-1}\left(\tilde{\mathbf{c}}_i^j\right) = \Delta_i^j \ \tilde{\mathbf{c}}_i^j. \tag{8}$$

Let $\overline{\mathbf{R}}_i^j$ be the dequantized residual component i at level j and let $\overline{\mathbf{R}}^j$ be the set of dequantized residual components at level j.

D. Fourth operation: Reconstruction of Details

Let the same quantization step be applied for all the components within a decomposition level j, i.e., $\forall i,\ \Delta_i^j = \Delta^j$. If it is greater than 1, the dequantized residual components will contain errors due to the quantization process.

Since only the dequantized residuals $\overline{\mathbf{R}}_i^j$ will be available in the decoder, the reconstructed details $\overline{\mathbf{W}}_i^j$ are computed using $\overline{\mathbf{R}}_i^j$ as

$$\overline{\mathbf{W}}^j = \overline{\mathbf{R}}^j + \textit{round}(\widehat{\mathbf{W}}^j). \tag{9}$$

The reconstructed details $\overline{\mathbf{W}}^j$ are used both in the encoder and decoder as input signal to the inverse wavelet transform to derive the approximation components at the next Haar-DWT level $j - 1$,

$$\widetilde{\mathbf{V}}^{j-1} = \mathbf{DWT}^{-1}\left((\widetilde{\mathbf{V}}^j, \overline{\mathbf{W}}^j), 1\right). \tag{10}$$

$\widetilde{\mathbf{V}}^{L-1\leq j\leq 1}$ refers to the reconstructed approximation components of each decomposition level j. These components differ from the original approximations because of the quantization error.

Following the NLRWA notation, the inverse Haar S-Transform equations are:

$$\left\{ \begin{array}{rcl} \widetilde{\mathbf{V}}_{2i-1}^{j-1} & = & \widetilde{\mathbf{V}}_i^j - \lfloor \frac{1}{2}\overline{\mathbf{W}}_i^j \rfloor \\ \widetilde{\mathbf{V}}_{2i}^{j-1} & = & \overline{\mathbf{W}}_i^j + \widetilde{\mathbf{V}}_{2i-1}^{j-1}. \end{array} \right. \tag{11}$$

$\widetilde{\mathbf{V}}^{j-1}$ is fed back into the system to produce the regression coefficients at level $j-1$ (Eq. 4). Fig. 2 graphically represents the whole NLRWA procedure for level j.

III. Reconstruction Error Control

This section introduces how NLRWA controls the peak absolute error. In addition, a smart strategy to select the quantization steps is proposed. By using this criterion, competitive compression ratios and scene quality retrieval are obtained.

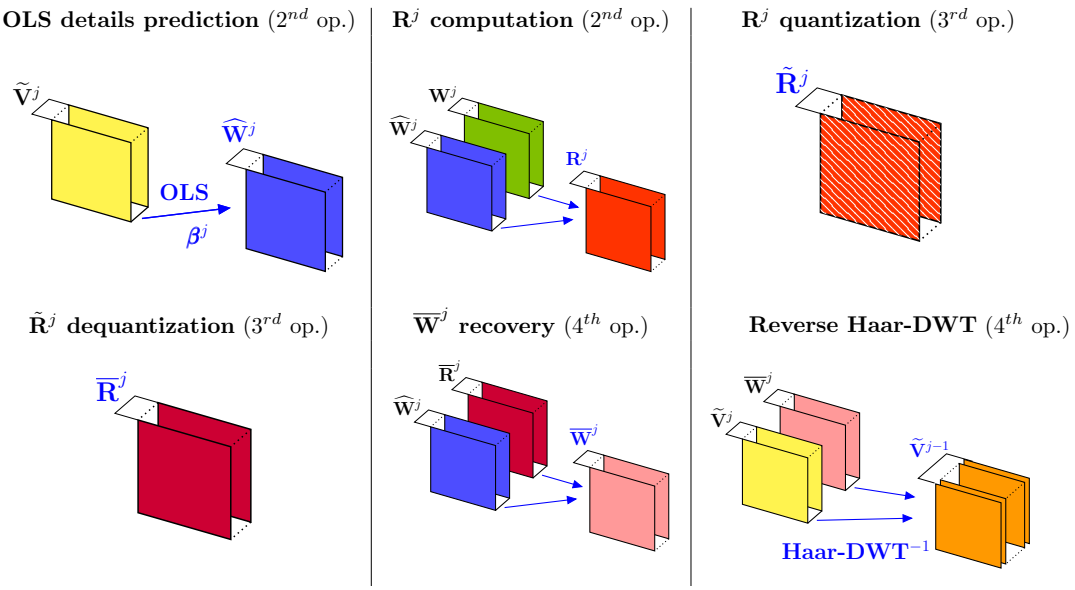

Figure 2: From *left* to *right* and from *top* to *bottom*, the NLRWA coding process is rendered for the decomposition level j and after the Haar S-Transform application, i.e., from the second operation to the fourth. All these stages are computed by the coder, which obtains the details prediction signal $\widehat{\mathbf{W}}^j$ in the second operation and employs it in the fourth operation.

A. Peak Absolute Error Restriction

In this section, we show that NLRWA can control the largest absolute error to a tolerance value Λ. Let $\mathrm{v}^0_{p,b}$ represent the pixel of the scene $\mathbf{V}^0$ located at position (p, b), where $1 \leq p \leq m$ and $1 \leq b \leq z$, p corresponds to the spatial sample and b to the spectral component (band) location. Denote the reconstructed scene after applying NLRWA by $\widetilde{\mathbf{V}}^0$, and let $\widetilde{\mathrm{v}}^0_{p,b}$ be the pixel of the recovered scene at the same location. Then, the error is limited to

$$\Lambda \geq \mathrm{PAE} = ||\mathbf{V}^0 - \widetilde{\mathbf{V}}^0||_\infty = \max_{p,b} |\mathrm{v}^0_{p,b} - \widetilde{\mathrm{v}}^0_{p,b}|. \tag{12}$$

Lemma 1. *For NLRWA, Λ is equal to*

$$\Lambda = \sum_{j=1}^{L} \left\lfloor \frac{1}{2}\, \Delta^j \right\rfloor. \tag{13}$$

This equation depicts the largest possible cumulative error introduced by the quantization stage in the final reconstruction. Note that the superscript j refers to the decomposition level of the quantization steps and not to a power.

Proof: let the error in the quantized residual be denoted by $\epsilon_{\mathrm{R}}{}^j_i = \overline{\mathbf{R}}^j_i - \mathbf{R}^j_i$. Examination of the USDZQ reveals that $-(\Delta^j - 1) \leq \epsilon_{\mathrm{R}}{}^j_i \leq (\Delta^j - 1)$. Thus,

$$|\epsilon_{\mathrm{R}}{}^j_{\max}| = \max_i |\epsilon_{\mathrm{R}}{}^j_i| = \max_i |\overline{\mathbf{R}}^j_i - \mathbf{R}^j_i| \leq \Delta^j - 1. \tag{14}$$

From Eq. 6, the recovered details at level j (Eq. 9) are then $\overline{\mathbf{W}}^j_i = \mathbf{R}^j_i + \epsilon_{\mathrm{R}}{}^j_i + \mathit{round}(\widehat{\mathbf{W}}^j_i) = \mathbf{W}^j_i + \epsilon_{\mathrm{R}}{}^j_i$.

In the reverse Haar S-Transform, for odd indexed components (Eq. 11), the approximations are reconstructed as $\widetilde{\mathbf{V}}^{j-1}_{2i-1} = \widetilde{\mathbf{V}}^j_i - \left\lfloor \frac{1}{2}\left(\mathbf{W}^j_i + \epsilon_{\mathrm{R}}{}^j_i\right) \right\rfloor =$

$$\begin{cases} \widetilde{\mathbf{V}}^j_i - \left(\left\lfloor \frac{1}{2}\mathbf{W}^j_i \right\rfloor + \left\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \right\rfloor + 1\right), & \text{if } \mathbf{W}^j_i \text{ and } \epsilon_{\mathrm{R}}{}^j_i \text{ are odd,} \\ \widetilde{\mathbf{V}}^j_i - \left(\left\lfloor \frac{1}{2}\mathbf{W}^j_i \right\rfloor + \left\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \right\rfloor\right), & \text{otherwise.} \end{cases} \tag{15}$$

Comparing the original approximation components with the reconstructed approximation components, for odd components, the produced error is

$$\delta^{j-1}_{2i-1} = \begin{cases} -\left\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \right\rfloor - 1, & \text{if } \mathbf{W}^j_i \text{ and } \epsilon_{\mathrm{R}}{}^j_i \text{ are odd,} \\ -\left\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \right\rfloor, & \text{otherwise.} \end{cases} \tag{16}$$

A similar analysis for even indexed components yields

$$\delta^{j-1}_{2i} = \begin{cases} \left\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \right\rfloor + 1, & \text{if } \mathbf{W}^j_i \text{ is even and } \epsilon_{\mathrm{R}}{}^j_i \text{ is odd,} \\ \left\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \right\rfloor, & \text{otherwise.} \end{cases} \tag{17}$$

Now, the largest distortion in the reconstructed approximation components happens when the error in the reconstructed residuals is the largest at all levels j. In such a case, for even Δ^j values, $\epsilon_{\mathrm{R}}{}^j_{\max}$ will be odd, and vice versa. In summary, the highest possible absolute error can be expressed as

$$|\delta^{j-1}_{\max}| = \max_i |\delta^{j-1}_i| \leq \begin{cases} \left\lfloor \frac{1}{2}(\Delta^j - 1) \right\rfloor + 1, & \text{if } \Delta^j \text{ even,} \\ \left\lfloor \frac{1}{2}(\Delta^j - 1) \right\rfloor, & \text{if } \Delta^j \text{ odd.} \end{cases} \tag{18}$$

The maximum possible error in the residuals at level j contributes to the error as

$$\lambda^j = \left\lfloor \frac{1}{2}\, \Delta^j \right\rfloor. \tag{19}$$

By induction, and after applying NLRWA iteratively from level L down to level 1, the PAE in the reconstructed scene is limited by the error tolerance value, derived as

$$\Lambda = \sum_{j=1}^{L} \lambda^j = \sum_{j=1}^{L} \left\lfloor \frac{1}{2}\ \Delta^j \right\rfloor. \qquad \blacksquare$$

We note that other integer DWT could be used instead of Haar wavelet transform in the first operation of NLRWA [10]. For such a case, Eq. 13 should be adapted accordingly.

B. Quantization Steps Selection Criterion

The number of combinations of quantization step sizes that fulfill Eq. 13 is given by

$$\mathrm{N} = \sum_{m=1}^{\min\{\Lambda, L\}} 2^m \binom{L}{m} \binom{\Lambda - 1}{m - 1} \tag{20}$$

(see [32]), which depends on the number of the highest decomposition level L and on the error tolerance value Λ. This combinatorial number grows rapidly as L and Λ increase, such that assessing the coding performance for every combination becomes unattainable.

Now, considering Eq. 13, odd quantization steps yield better performance than even quantization steps. If only odd quantization steps are considered, the possible combinations are reduced to (see [32])

$$\mathrm{N}_{\mathrm{odd}\Delta} = \binom{\Lambda + L - 1}{L - 1}. \tag{21}$$

Despite this reduction, exhaustive search over every possible combination is still prohibitive for reasonable values of L and Λ.

Rather, a heuristic selection criterion is proposed. In order to priorize the introduction of distortion into the residuals of the lowest decomposition levels (the less significant components in reconstruction), we define a quantization step-allocation formulation that, once replaced in Eq. 13, describes an approximation of a finite geometric series. This approximation enforces a positive and integer value for each quantization step by means of introducing a rounding operation. Thanks to this, higher compression ratios are achieved, while preserving better the signal's quality retrieval. Recall that the quantization steps of the residuals at level j are the same ($\Delta_i^j = \Delta^j,\ \forall i$). Given Λ fixed by the user, an odd value is assigned to the quantization step at level j according to:

$$\Delta^j = 2\left\lfloor \frac{\Lambda}{2^j} + \frac{1}{2} \right\rfloor + 1 \tag{22}$$

As an example, for a scene where L=8 wavelet decomposition levels have been applied, the quantization steps for Λ=10, $[\mathbf{R}^8, \mathbf{R}^7, \mathbf{R}^6, \mathbf{R}^5, \mathbf{R}^4, \mathbf{R}^3, \mathbf{R}^2, \mathbf{R}^1]$ are [1,1,1,1,3,3,7,11]; for Λ=25, [1,1,1,3,5,7,13,27]; for Λ=50, [1,1,3,5,7,13,27,51].

The proposed strategy may not necessarily secure the best rate-distortion results. However, it attains a very competitive coding performance, is independent of the processed scene, and is computationally efficient.

IV. Experimental Results

Experimental results of our embedded coding framework, NLRWA coupled with JPEG 2000, are reported in comparison to state-of-the-art prediction-based followed by quantization techniques, in comparison to the state-of-the-art two-stage near-lossless coding technique, and in comparison to the best performing rate control algorithm that enables lossy and near-lossless recovery. Scenes used in the experiments are available at [33].

A. Coding Pipeline

Our proposed coding method applies NLRWA through a Matlab implementation, and then JPEG 2000 compresses the NLRWA transformed data through Kakadu software. Although our framework uses JPEG 2000, other coders could be used [24]. Here, JPEG 2000 is employed mostly because of its scalability capability, competitive lossy and lossless performance, and capacity of providing progressive refinement of the scene's quality retrieval.

To obtain smooth and steady increasing rate-distortion curves with JPEG 2000, the predictive weighting scheme (PWS) [25] is applied. It attributes pyramidal weights according to the significance of the NLRWA spectral components in the reconstruction [11].

For our NLRWA approach, Maximum or Parsimonious regression model is selected depending on the number of spectral components of the original scene: Maximum model for scenes where $L \leq 8$, and Parsimonious model otherwise. Notwithstanding, other models, such as Restricted, or variants, e.g., Fast-Maximum or Fast-Restricted, can be selected. For a more detailed description, see [10].

The side information for NLRWA can be encoded by any entropy coder. Here, results are reported when applying LZMA [34]. The side information's bitrate is always added to the bitrate required to losslessly encode our NLRWA transformed scene.

B. Prediction-based followed by Quantization Coders Comparison

Table I presents an extended study of NLRWA + JPEG 2000 coding performance in comparison with CCSDS-123-AC, NLCCSDS-123 and M-CALIC, for different Λ values. All of these techniques employ the previous recovered pixel values instead of the original ones and the same quantizer, increasing their computational complexity in a similar and comparable manner.

Experimental results are presented for 26 scenes from different hyper- and ultraspectral sensors with a bit-depth of 16 bpppc (bits-per-pixel-per-component): calibrated and uncalibrated AVIRIS (referred to as CA and UA, respectively), calibrated Hyperion (CH), filtered uncalibrated Hyperion (FUH), uncalibrated IASI (UI), and AIRS Gran (AG). The scenes from the uncalibrated Hyperion (UH) corpus are filtered [35], [36] to remove the streaking artifacts along one of the spatial dimensions [37]. These artifacts appear because of the pushbroom sensor nature, and they should be dealt with for a better scene information assessment and visualization.

Table I: Average lossless ($\Lambda = 0$) and near-lossless ($\Lambda > 0$) compression results for several prediction-based techniques and our proposal, NLRWA + JPEG 2000. For all of them, USDZQ has been used for near-lossless. The best results are enhanced in bold. The coding gains of our method with respect to the other three techniques are included within parentheses. A positive difference means that our approach is better.

					Average bitrates (bpppc)				**Average SNR (dB)**		
Sensor abbreviation	Number of scenes	Number of spectral components	Average order-0 entropy (bpppc)	Λ values	CCSDS-123-AC	NLCCSDS-123	M-CALIC	NLRWA + JPEG 2000	CCSDS-123-AC & NLCCSDS-123	M-CALIC	NLRWA + JPEG 2000
CA	5	224	9.77	$\Lambda = 0$	3.66 (0.13)	3.73 (0.20)	4.03 (0.50)	**3.53**	—	—	—
				$\Lambda = 1$	**2.45** (-0.03)	2.54 (0.06)	2.87 (0.39)	2.48	59.29 (0.37)	59.29 (0.37)	**59.66**
				$\Lambda = 10$	**0.58** (-0.13)	0.94 (0.23)	0.88 (0.17)	0.71	42.94 (7.20)	41.38 (8.76)	**50.14**
				$\Lambda = 20$	**0.35** (-0.13)	0.72 (0.24)	0.53 (0.05)	0.48	37.46 (10.35)	35.38 (12.43)	**47.81**
				$\Lambda = 30$	**0.26** (-0.13)	0.63 (0.23)	0.40 (0.00)	0.40	34.15 (12.32)	31.84 (14.63)	**46.47**
UA	3	224	12.13	$\Lambda = 0$	5.87 (0.05)	5.95 (0.13)	6.13 (0.31)	**5.82**	—	—	—
				$\Lambda = 1$	**4.80** (-0.12)	4.89 (-0.03)	5.05 (0.13)	4.92	74.73 (0.06)	74.74 (0.05)	**74.79**
				$\Lambda = 10$	**1.96** (-0.56)	2.24 (-0.28)	2.22 (-0.30)	2.52	57.09 (4.19)	57.07 (4.21)	**61.28**
				$\Lambda = 20$	**1.14** (-0.41)	1.46 (-0.09)	1.40 (-0.15)	1.55	51.95 (4.93)	51.65 (5.23)	**56.88**
				$\Lambda = 30$	**0.83** (-0.27)	1.18 (0.08)	1.05 (-0.05)	1.10	48.90 (5.90)	48.20 (6.60)	**54.80**
CH	3	242	9.50	$\Lambda = 0$	**5.36** (-0.02)	5.61 (0.23)	—	5.38	—	—	—
				$\Lambda = 1$	**4.50** (-0.14)	4.75 (0.11)	—	4.64	65.63 (0.78)	—	**66.41**
				$\Lambda = 10$	**2.25** (-0.52)	2.54 (-0.23)	—	2.77	47.87 (4.82)	—	**52.69**
				$\Lambda = 20$	**1.44** (-0.57)	1.77 (-0.24)	—	2.01	42.42 (5.58)	—	**48.00**
				$\Lambda = 30$	**1.05** (-0.53)	1.40 (-0.18)	—	1.58	39.27 (5.97)	—	**45.24**
FUH	3	242	9.42	$\Lambda = 0$	4.26 (0.21)	4.37 (0.32)	4.28 (0.23)	**4.05**	—	—	—
				$\Lambda = 1$	3.08 (0.05)	3.19 (0.16)	3.10 (0.07)	**3.03**	59.67 (0.20)	59.67 (0.20)	**59.87**
				$\Lambda = 10$	0.73 (-0.20)	1.14 (0.21)	**0.61** (-0.32)	0.93	42.73 (5.84)	42.53 (6.07)	**48.57**
				$\Lambda = 20$	0.38 (-0.12)	0.79 (0.29)	**0.32** (-0.18)	0.50	37.38 (8.46)	36.24 (9.60)	**45.84**
				$\Lambda = 30$	0.26 (-0.07)	0.66 (0.33)	**0.24** (-0.09)	0.33	34.15 (10.35)	32.23 (12.27)	**44.50**
UI	4	8359	8.12	$\Lambda = 0$	2.82 (0.27)	2.89 (0.34)	2.94 (0.39)	**2.55**	—	—	—
				$\Lambda = 1$	1.53 (0.05)	1.68 (0.20)	1.74 (0.26)	**1.48**	46.95 (0.65)	46.95 (0.65)	**47.60**
				$\Lambda = 10$	**0.13** (-0.02)	0.49 (0.34)	0.41 (0.26)	0.15	31.52 (9.06)	28.34 (12.24)	**40.58**
				$\Lambda = 20$	**0.06** (-0.01)	0.42 (0.35)	0.29 (0.22)	0.07	26.26 (13.38)	21.89 (17.75)	**39.64**
				$\Lambda = 30$	**0.03** (-0.03)	0.40 (0.34)	0.26 (0.20)	0.06	23.17 (16.26)	18.36 (21.07)	**39.43**
AG	8	1501	11.39	$\Lambda = 0$	4.25 (0.28)	4.31 (0.34)	4.38 (0.41)	**3.97**	—	—	—
				$\Lambda = 1$	3.08 (0.12)	3.13 (0.17)	3.21 (0.25)	**2.96**	70.73 (0.17)	70.73 (0.17)	**70.90**
				$\Lambda = 10$	**0.61** (-0.09)	0.98 (0.28)	0.70 (0.00)	0.70	54.16 (5.49)	53.75 (5.90)	**59.65**
				$\Lambda = 20$	**0.29** (-0.03)	0.66 (0.34)	0.36 (0.04)	0.32	49.04 (8.59)	47.97 (9.66)	**57.63**
				$\Lambda = 30$	**0.20** (-0.04)	0.57 (0.33)	0.26 (0.02)	0.24	45.77 (11.32)	44.27 (12.82)	**57.09**

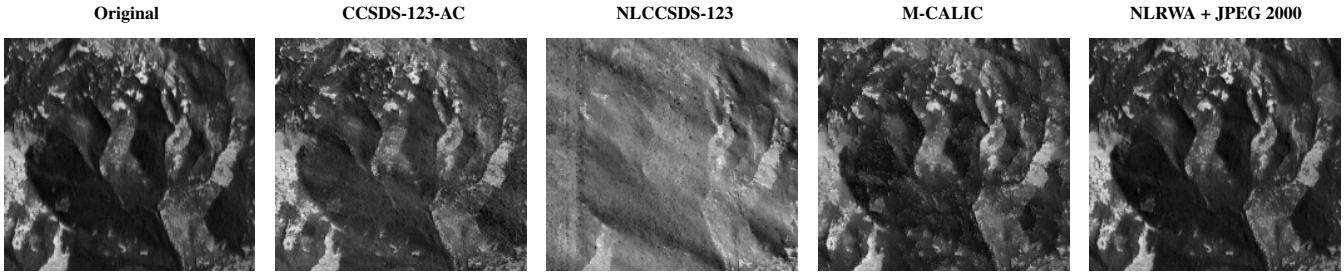

Figure 3: Crops of the spectral component 107 of the calibrated AVIRIS Yellowstone scene 18 (CA-Yellowstone sc18), and its reconstruction after applying CCSDS-123-AC, NLCCSDS-123, M-CALIC, and our proposal NLRWA + JPEG 2000. In all cases, the bitrate is about 0.5 bpppc.

NLRWA applies the Maximum regression model for AVIRIS and Hyperion scenes and Parsimonious for the rest. Here, NLCCSDS-123 employs its sample-adaptive encoder for lossless and near-lossless coding, when $\Lambda = 1$, and block-adaptive for the rest of near-lossless results. The predictor of both NLCCSDS-123 and CCSDS-123-AC considers 3 previous spectral components, and their mode and local sum are selected depending on the corpus. The neighbor oriented mode has been used for AVIRIS, AIRS and IASI sensors, and the column oriented predictor configuration for Hyperion sensor. The local sum applied in prediction is full mode only when processing the AVIRIS and IASI corpus scenes. The reduced mode is set for the rest.

The C++ M-CALIC software implementation does not handle transformed signals with more than 15 bpppc. Due to the dynamic range extension, no results can be obtained for the CH corpus. In contrast, NLRWA induces a dynamic range expansion of only, at most, 1 bit in the detail components. This avoids severe inconveniences in systems that support only a limited bit-depth [38].

Table I reports coding performance measured in bitrate and quality. For lossless coding ($\Lambda = 0$), our approach is superior. That is, it yields the lowest rate of any of the compared methods. For near-lossless coding, the best coding performance in terms of bitrate is provided by CCSDS-123-AC, while our approach is competitive for most error tolerance values Λ and for most sensors. Concerning quality, as measured by SNR, our approach is always the best performing, with increasingly

Table II: Bitrate and SNR results of Fully Embedded Two-Stage Coder [23] in comparison with our proposal, NLRWA + JPEG 2000, at several Λ values. Again, bold font indicates highest coding performance.

	Bitrate (bpppc)								**SNR (dB)**					
	[23]				NLRWA + JPEG 2000				[23]			NLRWA + JPEG 2000		
Scene \ Λ	*0*	*1*	*5*	*32*	*0*	*1*	*5*	*32*	*1*	*5*	*32*	*1*	*5*	*32*
UA12-Hawaii	**2.45**	**1.00**	**0.17**	**0.07**	2.54	1.47	0.46	0.15	51.50	47.44	41.95	**53.07**	**48.19**	**44.22**
UA12-Maine	**2.61**	**1.16**	**0.23**	**0.07**	2.69	1.61	0.58	0.22	54.87	49.80	40.00	**56.46**	**51.22**	**46.41**
CA-Yellowstone sc00	3.76	**2.24**	**0.70**	**0.17**	**3.74**	2.70	1.30	0.42	59.04	50.57	42.77	**60.56**	**53.22**	**46.65**
UA-Yellowstone sc00	**5.95**	**4.37**	**2.54**	**0.53**	6.08	5.19	3.74	1.25	75.00	63.24	52.62	**76.30**	**67.18**	**55.57**
CH-Agricultural	6.11	**4.53**	**2.77**	**0.84**	**5.49**	4.76	3.60	1.61	65.43	53.83	42.23	**67.60**	**58.42**	**45.84**
CH-Coral Reef	5.80	**4.21**	**2.40**	**0.60**	**5.12**	4.37	3.19	1.31	61.06	49.87	38.52	**63.25**	**54.16**	**42.29**
CH-Urban	6.14	**4.55**	**2.80**	**0.88**	**5.50**	4.77	3.60	1.66	66.20	54.60	43.12	**68.37**	**59.21**	**46.76**
UH-Erta Ale	**4.54**	**3.00**	**1.26**	**0.11**	4.79	3.86	2.44	0.60	57.92	46.75	39.01	**59.25**	**50.42**	**39.65**
UH-Lake Monona	**4.64**	**3.08**	**1.37**	**0.16**	4.99	4.07	2.66	0.73	59.55	48.33	40.50	**60.88**	**51.97**	**40.74**

larger differences as Λ grows.

Fig. 3 depicts a crop of a component of a CA scene. For a fair visual comparison, first we encode the scene for different error tolerance values and then we choose the encoded scene that requires a bitrate as close as possible to 0.5 bpppc. It can be noticed that the crop corresponding to NLRWA + JPEG 2000 conserves better the details and is less noisy.

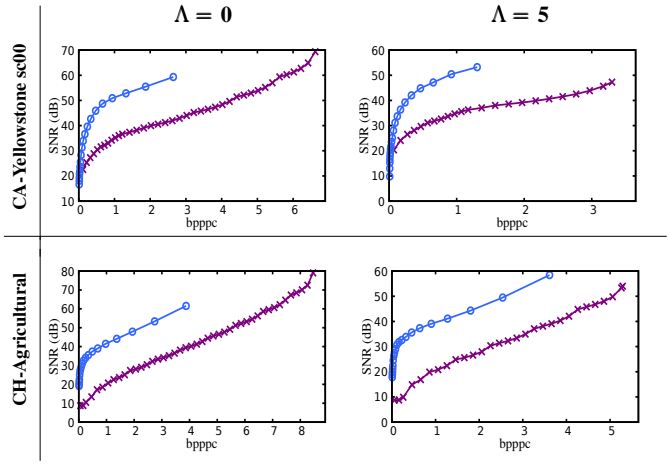

Figure 4: *Blue* with *circles* and *purple* with *crosses* curves represent the rate-distortion performance of, respectively, NLRWA and [23] with KLT replaced by Haar-DWT.

C. Two-Stage Near-lossless Coders Comparison

Table II reports bitrates and quality (SNR) results for the 9 scenes presented in the original work of Beerten et al. [23] in comparison with our approach. UA12 refers to the UA sensor that has stored the scenes with a bit-depth of 12 bpppc.

Both [23] and our approach use a bitplane-based entropy coder that allows for some degree of embeddedness and enables progressive lossy-to-lossless/near-lossless transmission. Concerning the quality of the reconstructed scenes, given any error tolerance value Λ, our proposal is always superior, as happened for the case of comparing against prediction-based coding techniques.

Concerning the bitrate performance, a cursory glance to Table II shows the good behaviour of [23] (with the exception of results for the CH corpus in the lossless regime). However, [23] applies a Karhunen-Loève-Transform (KLT) before JPEG 2000 lossy layer. KLT efficiently decorrelates the spectral dimension of a scene, but entails a high computational complexity and a non-negligible side information. When processing scenes with a very large number of spectral components, e.g., the scenes recorded by AIRS or IASI sensors, KLT results to be computationally untenable. Contrary to this, our proposal not only provides competitive bitrates at significantly lower computational complexity, but also benefits from its minor side information requirements, so that NLRWA is applicable also for scenes with more than 8,000 spectral components.

A fair comparison between NLRWA and the state-of-the-art method [23] in terms of computational complexity is provided in Fig. 4, where the rate-distortion curves for both techniques employ the Haar-DWT. In this case, our approach not only provides again better quality, but also better bitrate.

Table III: Bitrate, SNR and PAE results of the rate control presented by Valsesia et al. [14] and our near-lossless proposal. The mean of the squared Pearson correlation coefficients of each scene, $\overline{r}^2$, is included. z refers to the number of spectral components.

Scene	z	$\overline{r}^2$	[14]			NLRWA + JPEG 2000		
			Bitrate	SNR	PAE	Bitrate	SNR	PAE
CA-Yellowstone sc00	224	0.65	1.00	44.60	255	0.91	51.28	8
			2.00	57.79	25	1.73	55.37	3
			3.00	64.10	25	2.69	60.56	1
			4.00	71.13	3	3.74	∞	0
AG sc09	1501	0.65	1.00	53.86	18	0.96	60.40	7
			2.00	63.04	4	1.86	64.53	3
			2.99	67.28	3	2.87	70.15	1
			4.02	79.25	1	3.90	∞	0
CASI-T0477F06-NUC	72	0.44	1.00	41.69	49	0.96	45.80	24
			2.00	50.92	7	2.06	51.64	7
			3.00	57.64	4	3.05	56.56	3
			4.00	62.03	3	4.02	62.45	1
CRISM-sc167-NUC	545	0.40	1.00	44.14	48	1.02	41.07	39
			2.00	52.59	7	2.00	48.25	14
			3.00	59.37	3	2.97	54.43	6
			4.00	64.25	2	4.06	61.61	2
LANDSAT MOUNTAIN	6	0.53	1.00	27.13	10	1.00	21.20	28
			2.00	34.17	3	2.03	32.54	4
			3.00	39.37	3	3.03	38.74	1
			3.73	∞	0	4.06	∞	0
MODIS-MOD01DAY	14	0.65	1.00	38.70	230	1.11	28.27	491
			2.00	49.59	134	2.00	38.09	139
			3.00	58.66	17	2.99	46.11	52
			4.00	63.90	7	4.02	53.00	20

D. Lossy Rate Control-based Coders Comparison

Table III depicts bitrate and quality (SNR and PAE) values of our proposal and of the rate control presented by Valsesia et al. [14] for all the scenes presented in that paper.

Notice that [14] only provides end-to-end rate-controlled outcomes given a prefixed target bitrate, while our approach,

thanks to JPEG 2000's rate control algorithm, permits progressive decoding of a codestream. In this experiment, several Λ values have been tested in our approach, choosing those that provide the closest bitrate to that reported in [14]. This last technique yields near-lossless outcomes in the sense of being able to determine the PAE that results when compressing a scene at a specified target bitrate. This determination can only be made post-compression, but does not require decompression of the codestream. On the other hand, the scheme is not designed to achieve a desired target error in reconstruction.

NLRWA considers all (Maximum) or several neighbor (Parsimonious) spectral components per spatial sample as regressors in the prediction stage. Therefrom, it is very suited to apply when processing scenes with a large number of spectral components, i.e., for the CA Yellowstone scene 00 (CA-Yellowstone sc00), the AIRS Gran scene 09 (AG sc09) and CASI-T0477F06-NUC. For them, NLRWA + JPEG 2000 is superior in PAE-bpppc and PAE-SNR performance. Conversely, LANDSAT MOUNTAIN and MODIS-MOD01DAY scenes have a low number of spectral components. This may lead to lower Λ values than prescribed by the user, penalizing the coding performance of our approach. For these two scenes, the quantization steps criterion of Subsection III-B has been modified, allowing the assignment of larger quantization step values to each decomposition level.

The CRISM sensor is affected by common artifacts present in pushbroom sensors. Although a non uniformity calibration (NUC) has been applied for the CRISM scene, it contains groups of spectral components still strongly compromised, as shown by the low mean of the squared Pearson correlation coefficient (0.40). For this sensor, [14], whose predictor takes into consideration the correlation between adjacent spatial samples, outperforms our proposal.

V. Conclusion

This manuscript introduces NLRWA, the first near-lossless compression technique based on regression, in particular, on the pyramidal multiresolution regression wavelet analysis. It expands the state-of-the-art lossless compression technique RWA by introducing quantization and a feedback loop to compensate the quantization error. We also provide a smart criterion, which is independent of the scene, to select a unique quantization steps combination. This criterion helps avoiding iterative computations while producing very competitive coding performances in terms of compression ratio and quality of the reconstruced scene. NLRWA can be followed by any entropy coder. Here, NLRWA is coupled with JPEG 2000, enabling progressive lossy-to-lossless/near-lossless decoding.

Experimental results indicate that NLRWA + JPEG 2000 considerably outperforms rate control-based algorithms in both bitrate and scene quality reconstruction for scenes with a large number of spectral components. When comparing with two-stage near-lossless coders, our approach always yields superior quality retrieval, and achieves competitive compression ratios at significantly lower computational cost. With respect to the most competitive prediction-based followed by quantization techniques such as CCSDS-123-AC, NLCCSDS-123 and M-CALIC, our approach NLRWA + JPEG 2000 always yields reconstructed scenes with the highest quality, and obtains outstanding compression ratios while offering progressive decoding and some degree of embeddedness.

References

[1] E. Magli, G. Olmo, and E. Quacchio, "Optimized Onboard Lossless and Near-lossless Compression of Hyperspectral Data Using CALIC," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 1, pp. 21–25, Feb. 2004, DOI: 10.1109/LGRS.2003.822312.

[2] Consultative Committee for Space Data Systems (CCSDS), "Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1." *CCSDS. Blue Book*, 2012. https://public.ccsds.org/Pubs/123x0b1ec1.pdf

[3] N. Amrani, J. Serra-Sagristà, V. Laparra, M. W. Marcellin, and J. Malo, "Regression Wavelet Analysis for Lossless Coding of Remote-Sensing Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5616–5627, Jun. 2016, DOI: 10.1109/TGRS.2016.2569485.

[4] A. C. Karaca, and M. K. Güllü, "Lossless hyperspectral image compression using bimodal conventional recursive least-square," *Remote sensing letters*, vol. 9, no. 1, pp. 31–40, Aug. 2017, DOI: 10.1080/2150704X.2017.1375612.

[5] J. Mielikainen, and B. Huang, "Lossless Compression of Hyperspectral Images Using Clustered Linear Prediction with Adaptive Prediction Length," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1118–1121, Apr. 2012, DOI: 10.1109/LGRS.2012.2191531.

[6] F. Gao, and S. Guo, "Lossless Compression of Hyperspectral Images Using Conventional Recursive Least-Squares Predictor with Adaptive Prediction Bands," *SPIE Journal of Applied Remote Sensing*, vol. 10, no. 1, pp. 015010, Feb. 2016, DOI: 10.1117/1.JRS.10.015010.

[7] J. Song, Z. Zhang, and X. Chen, "Lossless Compression of Hyperspectral Imagery Via RLS Filter," *IEEE Electronics Letters*, vol. 49, no. 16, pp. 992–994, Aug. 2013, DOI: 10.1049/el.2013.1315.

[8] 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 Processing Letters*, vol. 22, no. 12, pp. 2194–2198, Jun. 2015, DOI: 10.1109/LSP.2015.2443913.

[9] J. Song, L. Zhou, C. Deng, and J. An, "Lossless compression of hyperspectral imagery using a fast adaptive-length-prediction RLS filter," *Remote sensing letters*, vol. 10, no. 4, pp. 401–410, Jan. 2019, DOI: 10.1080/2150704X.2018.1562257.

[10] S. Álvarez-Cortés, N. Amrani, and J. Serra-Sagristà, "Low Complexity Regression Wavelet Analysis Variants for Hyperspectral Data Lossless Compression," *International Journal of Remote Sensing*, Taylor & Francis, p. 1–30, Sep. 19, 2017, DOI: 10.1080/01431161.2017.1375617.

[11] S. Álvarez-Cortés, N. Amrani, M. Hernández-Cabronero, and J. Serra-Sagristà, "Progressive Lossy-to-Lossless Coding of Hyperspectral Images through Regression Wavelet Analysis," *International Journal of Remote Sensing*, vol. 39, no. 7, pp. 2001–2021, Jul. 2017, DOI: 10.1080/01431161.2017.1343515.

[12] N. Amrani, J. Serra-Sagristà, and M. W. Marcellin, "Low Complexity Prediction Model for Coding Remote-Sensing Data with Regression Wavelet Analysis," *In Data Compression Conference (DCC), IEEE*, p. 112–121, May 2017, DOI: 10.1109/DCC.2017.61.

[13] D. S. Taubman, and M. W. Marcellin, "JPEG 2000 Image Compression Fundamentals, Standards and Practice," *Norwell, Massachusetts 02061 USA: Kluwer Academic Publishers*, 2002.

[14] D. Valsesia, and E. Magli, "Fast and Lightweight Rate Control for Onboard Predictive Coding of Hyperspectral Images,", *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, pp. 394–398, Jan. 2017, DOI: 10.1109/LGRS.2016.2644726.

[15] D. Valsesia, and E. Magli, "A Novel Rate Control Algorithm for Onboard Predictive Coding of Multispectral and Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6341–6355, Jan. 2014, DOI: 10.1109/TGRS.2013.2296329.

[16] M. Conoscenti, R. Coppola, and E. Magli, "Constant SNR, Rate Control, and Entropy Coding for Predictive Lossy Hyperspectral Image Compression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7431–7441, Sep. 2016, DOI: 10.1109/TGRS.2016.2603998.

[17] I. Blanes, E. Magli, and J. Serra-Sagristà, "A Tutorial on Image Compression for Optical Space Imaging Systems," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 3, pp. 8–26, Oct. 2014, DOI: 10.1109/MGRS.2014.2352465.

[18] J. Bartrina-Rapesta, I. Blanes, F. Aulí-Llinàs, J. Serra-Sagristà, V. Sanchez, and M. W. Marcellin, "A Lightweight Contextual Arithmetic Coder for On-Board Remote Sensing Data Compression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4825–4835, May 2017, DOI: 10.1109/TGRS.2017.2701837.

[19] G. Carvajal, B. Penna, and E. Magli, "Unified Lossy and Near-lossless Hyperspectral Image Compression based on JPEG 2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 593–597, Nov. 2008, DOI: 10.1109/LGRS.2008.2000651.

[20] C. W. Chen, T. C. Lin, S. H. Chen, and T. K. Truong, "A Near-lossless Wavelet-based Compression Scheme for Satellite Images," in *Computer Science and Information Engineering, WRI World Congress on*, vol. 6, pp. 528–532, Jul. 2009, DOI: 10.1109/CSIE.2009.933.

[21] R. Ansari, N. Memon, and E. Ceran, "Near-lossless Image Compression Techniques", *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 486–494, Jul. 1998, DOI: 10.1117/1.482591.

[22] X. Wu, and P. Bao, 'Near-lossless image compression by combining wavelets and CALIC", *IEEE Conference Record of the 31 Asilomar Conference on Signals, Systems and Computers (Cat. No. 97CB36136)*, vol. 2, pp. 1427–1431, Nov. 1997, DOI: 10.1109/ACSSC.1997.679139.

[23] J. Beerten, I. Blanes, and J. Serra-Sagristà, "A Fully Embedded Two-Stage Coder for Hyperspectral Near-lossless Compression", *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1775–1779, May 2015, DOI: 10.1109/LGRS.2015.2425548.

[24] S. Álvarez-Cortés, J. Bartrina-Rapesta, and J. Serra-Sagristà, "Multilevel Split Regression Wavelet Analysis for Lossless Compression of Remote Sensing Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 15400–1544, Jul. 2018, DOI 10.1109/LGRS.2018.2850938.

[25] N. Amrani, J. Serra-Sagristà, M. Hernández-Cabronero, and M. Marcellin. 2016. "Regression Wavelet Analysis for Progressive-Lossy-to-Lossless Coding of Remote-Sensing Data." *Data Compression Conference DCC. IEEE.* pp. 121–130, Dec. 2016, DOI: 10.1109/DCC.2016.43.

[26] N. Amrani, J. Serra-Sagristà, and M. W. Marcellin, "Unbiasedness of regression wavelet analysis for progressive lossy-to-lossless coding," *In Picture Coding Symposium (PCS)*, pp. 1–5, Dec. 2016, DOI: 10.1109/PCS.2016.7906316.

[27] J. García-Sobrino, J. Serra-Sagristà, and J. Bartrina-Rapesta, "Hyperspectral IASI L1C data compression", *Sensors*, vol. 17, no. 6, Jun. 2017, DOI: 10.3390/s17061404.

[28] E. Ahanonu, M. W. Marcellin, and A. Bilgin, "Clustering Regression Wavelet Analysis for Lossless Compression of Hyperspectral Imagery", *Data Compression Conference, IEEE*, Mar. 2019, DOI: 10.1109/DCC.2019.00063.

[29] E. Ahanonu, M. W. Marcellin, and A. Bilgin, "Lossless image compression using reversible integer wavelet transforms and convolutional neural networks", *Data Compression Conference, IEEE*, Jul. 2018, DOI: 10.1109/DCC.2018.00048.

[30] A. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo, "Wavelet Transforms that Map Integers to Integers", *Applied and computational harmonic analysis,* vol. 5, no. 3, pp. 332–369, Aug. 1996, DOI: 10.1006/acha.1997.0238.

[31] J. Nocedal, and S. J. Wright, "Least-Squares Problems," *Numerical optimization, Springer,* pp. 245–269, 2006, DOI: 10.1007/978-0-387-74759-0_329.

[32] J. Serra-Sagristà, "Enumeration of Lattice Points in l1 Norm," *Information processing letters*, vol. 76, no. 1-2, pp. 39–44, Nov. 2000, DOI: 10.1016/S0020-0190(00)00119-8.

[33] Consultative Committee for Space Data Systems (CCSDS), "123.0-B-Info TestData," 2015. http://cwe.ccsds.org/sls/docs/SLS-DC/123.0-B-Info/TestData

[34] I. Pavlov, "Lzma sdk (software development kit)", https://www.7-zip.org/sdk.html, 2007.

[35] I. Blanes, and J. Serra-Sagristà, "Pairwise Orthogonal Transform for Spectral Image Coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, p. 961–972, Oct. 2010, DOI: 10.1109/TGRS.2010.2071880.

[36] Consultative Committee for Space Data Systems (CCSDS). 2015. "Lossless Multispectral & Hyperspectral Image Compression CCSDS 120.2-G-1." *CCSDS. Green Book. https://public.ccsds.org/Pubs/120x2g1.pdf*

[37] U.S. Geological Survey and NASA, "Earth Observing 1, Hyperion Website."http://eo1.usgs.gov/hyperion.php.

[38] I. Blanes, M. Hernández-Cabronero, F. Aulí-Llinas, J. Serra-Sagristà, and M. W. Marcellin, "Isorange Pairwise Orthogonal Transform", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, p. 3361–3372, Dec. 2014, DOI: 10.1109/TGRS.2014.2374473.