Regression Wavelet Analysis for Lossless Coding of Remote-Sensing Data

Naoufal Amrani, Joan Serra-Sagristà, *Senior Member, IEEE*, Valero Laparra, Michael W. Marcellin, *Fellow, IEEE*, and Jesus Malo

***Abstract*—A novel wavelet-based scheme to increase coefficient independence in hyperspectral images is introduced for lossless coding. The proposed regression wavelet analysis (RWA) uses multivariate regression to exploit the relationships among wavelet-transformed components. It builds on our previous nonlinear schemes that estimate each coefficient from neighbor coefficients. Specifically, RWA performs a pyramidal estimation in the wavelet domain, thus reducing the statistical relations in the residuals and the energy of the representation compared to existing wavelet-based schemes. We propose three regression models to address the issues concerning estimation accuracy, component scalability, and computational complexity. Other suitable regression models could be devised for other goals. RWA is invertible, it allows a reversible integer implementation, and it does not expand the dynamic range. Experimental results over a wide range of sensors, such as AVIRIS, Hyperion, and Infrared Atmospheric Sounding Interferometer, suggest that RWA outperforms not only principal component analysis and wavelets but also the best and most recent coding standard in remote sensing, CCSDS-123.**

***Index Terms*—Redundancy in hyperspectral images, remote sensing data compression, transform coding via regression, wavelet-based transform coding.**

I. Introduction

REMOTE-sensing data have become enormously important for a myriad of applications addressed to the Earth's observation. Recent sensors can cover large geographical areas, producing images of unprecedented spectral and spatial resolution. For instance, the Infrared Atmospheric Sounding Interferometer (IASI) instrument onboard the MetOp satellite captures 8359 spectral channels with a 60° field of view, about 1530 lines per orbit, and 14 orbits per day at an acquisition bit depth of 16 bits per pixel (bpp) close to 20 GB daily. Hence, the need for efficient coding techniques for remote-sensing data becomes more and more imperative to improve the capabilities of storage and transmission.

Manuscript received October 16, 2015; revised February 29, 2016 and April 19, 2016; accepted May 6, 2016. Date of publication June 8, 2016; date of current version August 2, 2016. This work was supported in part by the Spanish Ministry of Economy and Competitiveness and by the European Regional Development Fund under Grants TIN2015-71126-R, TIN2012-38102-C03-00 and BFU2014-58776- R, and by the Catalan Government under Grant 2014SGR-691.

N. Amrani and J. Serra-Sagristà are with Universitat Autònoma de Barcelona, 08193 Barcelona, Spain (e-mail: naoufal.amrani@uab.cat).

V. Laparra and J. Malo are with Universitat de València, 46980 València, Spain.

M. W. Marcellin is with the University of Arizona, Tucson, AZ 85721 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.2016.2569485

Most efficient coding techniques for remote-sensing data are based on a redundancy reduction transform that exploits the relations in the spectral and spatial dimensions. The problem of appropriate signal representation for transform coding is equivalent to the feature extraction problem in statistical learning [4]. Some hyperspectral coding techniques apply a discrete wavelet transform (DWT) in all three dimensions (two spatial and one spectral) [5], [6]. Other techniques apply a 2-D DWT spatially while using a different transform in the spectral dimension [7]–[10]. The transform in the spectral dimension is considered crucial in coding hyperspectral images due to its significant impact on the coding performance [11]. Generally, transforms that provide better exploitation of correlation, resulting in better energy compaction, are the ones that yield larger coding gain [12].

In particular, principal component analysis (PCA), also known as the Karhunen–Loève transform, is the optimal decorrelating transform for Gaussian sources [13]–[15]. It is widely applied to multicomponent images because of its excellent performance as a spectral decorrelator. However, PCA is a data-dependent transform, entailing the need to compute it for each individual image before its application. Additionally, its computational complexity is substantial due to the covariance matrix calculation, the extraction of eigenvectors, and the matrix factorization/integer mapping (when integer implementation is needed for lossless coding [16]). Also, the PCA is not a component-scalable transform, i.e., the recovery of any single image component depends on every transformed component. In scenarios where the input data consist of a small number of spectral components, such as multispectral images, PCA and its integer implementation can be acceptable. Nevertheless, in scenarios where the input data have a significant size in the spectral dimension, such as IASI images, PCA and its integer implementation are not feasible.

In light of this fact, a number of approaches have been proposed to reduce the computational complexity of PCA while trying to minimize the degradation in coding performance. One approach is based on subsampling the data set and estimating the covariance matrix using a reduced subset of coefficients spatially and/or spectrally [7], [11], [17]. In this way, the complexity of computing the covariance matrix can be alleviated. However, other complexity sources, such as the eigenvector extraction, the matrix multiplication, and the integer mapping for lossless coding, remain. A second approach suggests divide-and-conquer strategies to approximate the PCA while providing reduced computational complexity and some amount of component scalability. These strategies are based mainly on clustered PCA [18], [19] or on multiple pairwise PCA [20]. In

most cases, the lossless coding performance of these strategies falls somewhat below that of the full-complexity PCA, yet a third approach consists in learning the transform on a set of images of one particular sensor in order to obtain an efficient transform that can be applied to new images from the same sensor [21]–[24].

Beyond computational issues, a major conceptual limitation of PCA is related to its focus on producing decorrelated (rather than independent) components. In particular, PCA focuses on the covariance matrix (second-order relations) and neglects higher order moments, which may be relevant in non-Gaussian signals. Remote sensing hyperspectral images have been shown to be non-Gaussian both in the spatial and spectral dimensions [25]–[27]. Since the efficiency of transform coding is attached to the degree of statistical independence achieved [28], approaches focused on decorrelation may lead to inefficient representations due to higher order statistical relations still being present after decorrelation.

The (theoretical) consideration of higher order moments in independent component analysis (ICA) [29], and the (experimental) fact that ICA filters are qualitatively similar to wavelet functions in hyperspectral imagery [25] may be the fundamental reasons for the widespread use of wavelets in transform coding of hyperspectral images [5], [6], [30].

Another limitation of classical PCA (and also classical ICA) is its linear nature. Linear PCA and linear ICA (and their fixed-basis approximations, DCT and wavelets) do not completely achieve statistical independence in hyperspectral images. Similar to the case of photographic images [31]–[35], residual dependence has been reported for hyperspectral imagery in these representations [26], [27]. Thus, despite the simplicity and efficiency that linearity provides, *in principle*, nonlinear methods could provide performance improvements by better exploiting the dependence present in the data.

The aforementioned discussion suggests the pursuit of nonlinear generalizations of PCA, ICA, and their fixed-basis versions, DCT and wavelets, respectively. In [1] and [3], the following families of nonlinear generalizations of feature extraction transforms were reviewed: (i) kernel and spectral techniques such as kernel-PCA, kernel-ICA, or local linear embedding [36]–[38]; (ii) neural networks and autoencoders [39]–[42]; and (iii) techniques based on curvilinear features [1], [3], [13], [34], [43]–[45]. In the adaptation of these feature extraction ideas to image coding, there are two major considerations of importance: First, the transform must be invertible, and second, the computational complexity and the memory consumption should be reasonable. These requirements eliminate a large number of candidate approaches. For example, many nonlinear generalizations of ICA are not invertible (for instance, [34], [43], and [46]). With this in mind, the most promising approaches seem to lie within the invertible techniques of families (ii) and (iii).

In [2], we explored lossless hyperspectral image coding using curvilinear techniques (family iii) based on principal polynomial analysis (PPA) [1]. PPA exploits regression to remove nonlinear dependence that remains after linear feature extraction (e.g., after classical PCA). In [1], it was shown that PPA achieves higher energy compaction and statistical independence than PCA. However, in practical hyperspectral coding, the appropriate handling of side information and the sequential error introduced by the integer mapping dramatically penalize the coding gain. Additionally, the computational complexity and memory requirements of these nonlinear transforms are even larger than those of the original PCA, which is already demanding [13].

As a result, rather than using sophisticated nonlinear data-dependent feature extraction, we focus on theoretically suboptimal but simpler traditional transforms (such as the DWT) while adopting predictive techniques [47], [48] to exploit any remaining *posttransform* statistical dependence. This idea has been applied to encode residual errors in the spatial domain after exploiting smoothness [49] in photographic images. In works more closely related to that proposed here, predictive schemes have been employed in the wavelet domain based on the known relations of image coefficients [32], [50].

Building upon this, this paper introduces the regression wavelet analysis (RWA). This transform can be seen as a predictive scheme to reduce redundancy after wavelet analysis has been performed in the spectral dimension of hyperspectral images. RWA shares the principal properties of the DWT such as component scalability and low complexity while yielding excellent performance for lossless coding. Specifically, RWA yields superior performance to that of the best comparable spectral decorrelation techniques such as PCA. Significantly, RWA also overperforms the most recent and most competitive prediction-based hyperspectral coding technique, CCSDS-123 [51].

This paper is organized as follows. Section II describes and formalizes the proposed RWA scheme and the selection of the regression model. Section III investigates several relevant features of RWA, including its redundancy reduction ability and computational complexity. In Section IV, we assess the coding performance achieved by RWA for images corresponding to a wide range of remote sensing scenarios. Finally, Section V puts forward our conclusions.

II. RWA

To establish the necessary notation, this section begins with a review of the DWT. The discussion then proceeds to an analytical description of the RWA and the corresponding regression model. The section concludes by suggesting a fast implementation for RWA.

A. DWT

The DWT provides a multiresolution decomposition of signals into approximation $\mathbf{V}$ and details $\mathbf{W}$. In what follows, we consider the 1-D DWT in the component (spectral) direction of a multicomponent image. The DWT can be computed with a pyramidal algorithm based on convolution [52]. The algorithm is illustrated in Fig. 1. We begin by considering a general formulation for the DWT that maps real numbers to real numbers. We later consider transforms that map integers to integers.

Suppose that an original multicomponent image $\mathbf{V}^0$ has $z = 2^d$ components with each component having m spatial samples.

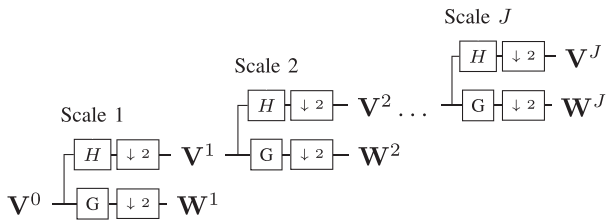

Fig. 1. DWT decomposition with J levels.

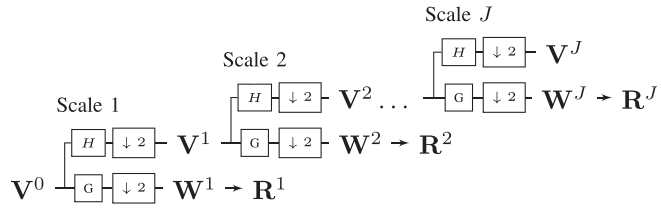

Fig. 2. RWA decomposition with J levels.

Then, we write $\mathbf{V}^0 \in \mathbb{R}^{m\times z}$ and

$$\mathbf{V}^0 = \left[\mathbf{V}^0(1), \ldots, \mathbf{V}^0(z)\right], \quad \mathbf{V}^0(i) = \mathbf{V}_i^0 \in \mathbb{R}^{m\times 1}.$$

Then, the wavelet representation of $\mathbf{V}^0$ with J levels for $1 \leq J \leq \log_2(z)$ is given by

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

where the one-level DWT decomposition of each $\mathbf{V}^{j-1}$ is given by

$$\mathbf{DWT}(\mathbf{V}^{j-1}, 1) = (\mathbf{V}^j, \mathbf{W}^j) \tag{2}$$

with

$$\mathbf{V}^j(n) = \sum_k h(2n-k)\mathbf{V}^{j-1}(k) \tag{3}$$

$$\mathbf{W}^j(n) = \sum_k g(2n-k)\mathbf{V}^{j-1}(k). \tag{4}$$

In these expressions, $h(n)$ and $g(n)$ are the impulse responses of low-pass and high-pass analysis filters H and G, respectively. At each scale or level j, the signal $\mathbf{V}^{j-1} \in \mathbb{R}^{m\times(z\cdot 2^{-j+1})}$ is decomposed into the approximation signal $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ and the detail signal $\mathbf{W}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ at half resolution each. The approximation signal $\mathbf{V}^j$ usually contains most of the information of the previous signal $\mathbf{V}^{j-1}$, whereas the detail signal $\mathbf{W}^j$ contains the information difference between $\mathbf{V}^{j-1}$ and $\mathbf{V}^j$. The decomposition is usually repeated in cascade form on the approximation signal $\mathbf{V}^j$ as shown in Fig. 1, and the maximum number of iterations that can be performed is $d = \log_2(z)$.

To reconstruct the original data, the DWT components are passed through low- and high-pass synthesis filters to obtain

$$\begin{aligned}\mathbf{V}^{j-1}(n) &= \mathbf{DWT}^{-1}\left((\mathbf{V}^j, \mathbf{W}^j), 1\right)\\ &= \sum_k \tilde{h}(n-2k)\mathbf{V}^j(k) + \sum_k \tilde{g}(n-2k)\mathbf{W}^j(k).\end{aligned} \tag{5}$$

If the original signal $\mathbf{V}^0$ has $z = 2^d$ components, then the representations $\mathbf{V}^j$ and $\mathbf{W}^j$ have $z\cdot 2^{-j}$ components each and the DWT representation $(\mathbf{V}^J, (\mathbf{W}^j)^{1\leq j\leq J})$ has the same total number of components as the original signal $\mathbf{V}^0$. With proper boundary handling procedures, the number of components z need not be a power of two. In this case, $z\cdot 2^{-j}$ is rounded up or down so that the total number of components at each level is always the same.

Rather than the convolution implementation described earlier, the DWT decomposition can be performed using a lifting scheme [53], [54]. This implementation facilitates the inclusion of rounding steps to obtain invertible transforms that map integers to integers. Such transforms are often called "reversible." In the wavelet literature, it has been demonstrated that the DWT approximately decorrelates some stochastic and nonstochastic processes [55]. This means that any two distinct within-scale or between-scale coefficients are approximately decorrelated, with the correlation decaying as the separation between scales increases.

B. RWA

The proposed RWA scheme generalizes the DWT by applying regression to tackle the redundancy that still remains in the DWT domain. At each scale j, each detail component $\mathbf{W}^j(i) = \mathbf{W}_i^j \in \mathbb{R}^{m\times 1}$ is predicted from the information contained in the approximation components $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ within the same scale j. This prediction $\widehat{\mathbf{W}}_i^j = f_i(\mathbf{V}^j)$ is then removed to obtain

$$\mathbf{R}^j = \mathbf{W}^j - \widehat{\mathbf{W}}^j. \tag{6}$$

A regression model $f_i(\mathbf{V}^j)$ is used to estimate the conditional mean of each $\mathbf{W}_i^j \in \mathbb{R}^{m\times 1}$ (dependent variable) from some or all the approximation components $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ (independent variables). Later in this paper, we will propose three different models that address the issues concerning estimation accuracy, computational complexity, and spectral scalability. Note that the prediction functions f might be linear or nonlinear, and they might use a small or a large set of neighbor coefficients. These possibilities for prediction through regression lead to a range of particular implementations of RWA.

As illustrated in Fig. 2, the resulting RWA affects only the detail components at each level of the transform. The approximation components are unchanged from those of the DWT.

Following the notation of (2) and (1), the one-level RWA and the J-level RWA become:

$$\mathbf{RWA}(\mathbf{V}^{j-1}, 1) = (\mathbf{V}^j, \mathbf{R}^j) \tag{7}$$

$$\mathbf{RWA}(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{R}^j)^{1\leq j\leq J}\right) \tag{8}$$

respectively.

The RWA transform is easily inverted. At each scale j, the estimate $\widehat{\mathbf{W}^j}$ is computed from $\mathbf{V}^j$. The approximation at level $j-1$ is then obtained by computing

$$\mathbf{W}^j = \mathbf{R}^j + \widehat{\mathbf{W}^j} \tag{9}$$

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

By repeating in cascade for $j = J, J-1, \ldots, 1$, the reconstruction $\mathbf{V}^0$ is achieved from its RWA representation $(\mathbf{V}^J, (\mathbf{R}^j)^{1\le j\le J})$.

C. Regression Model

To find the estimate $\widehat{\mathbf{W}}^j$ of the detail components $\mathbf{W}^j$ based on the approximation components $\mathbf{V}^j$, three different estimation models are discussed in this section. The first model, called the *maximum model*, is the most general. It employs all of the approximation components from $\mathbf{V}^j$ in the computation of the prediction of each detail component $\mathbf{W}_i^j$. The second model is called the *restricted model*. For a given detail component to be predicted, this model determines a subset of components from $\mathbf{V}^j$ to include in the prediction in order to preserve the component scalability of the original DWT. The third and final model is called the *exogenous model*. This model has a low computational cost and is similar to the exogenous model proposed for the PCA [56] or for the orthogonal optimal spectral transform [21], [22].

1) Maximum Model: The *maximum model* involves all $k = z \cdot 2^{-j}$ components of the approximation $\mathbf{V}^j \in R^{m\times(z\cdot 2^{-j})}$ in order to estimate each detail component $\mathbf{W}_i^j \in \mathbb{R}^{m\times 1}$, $i \in I = \{1, \ldots, k = z\cdot 2^{-j}\}$. The form of each estimator is given by

$$\widehat{\mathbf{W}}_i^j = \boldsymbol{\beta}_{i,0}^j + \boldsymbol{\beta}_{i,1}^j \mathbf{V}_1^j + \cdots + \boldsymbol{\beta}_{i,k}^j \mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}. \tag{11}$$

Since all components of the approximation are included, this form provides the most general linear estimator possible in this formulation. In this model, the estimator parameters (regression coefficients) $\boldsymbol{\beta}^j$ are found for each individual image using the least squares method [57] that minimizes the sum of squares of the distances between the original components and the estimated ones

$$\min: \left\| \mathbf{W}_i^j - \widehat{\mathbf{W}}_i^j \right\|_2.$$

2) Restricted Model: The *restricted model* employs only a small number of approximation components, with the goal of preserving as much as possible the component scalability of the original DWT. In other words, the number of transformed components that need to be accessed to reconstruct an image component should be as small as possible. In the DWT context, the number of needed transformed components depends on the width of the synthesis filters. As an example, consider the Haar DWT. For this simple transform, only two transformed components from level j are required to reconstruct one approximation component at level $j-1$, one approximation component, and one detail component. More generally, for a given DWT, a number $|I_p|$ of approximation components $\mathbf{V}_{i\in I_p}^j$ and a number $|I_q|$ of detail components $\mathbf{W}_{i\in I_q}^j$ from level j are required to reconstruct an approximation component $\mathbf{V}_r^{j-1}$ at level $j-1$. We denote this by

$$\left(\mathbf{V}_{i\in I_p}^j, \mathbf{W}_{i\in I_q}^j\right) \xrightarrow{\text{Reconst.}} \mathbf{V}_r^{j-1}.$$

From the opposite point of view, a given single detail component $\mathbf{W}_i^j$ is involved in the reconstruction of t components at level $j-1$. Let us then associate $\mathbf{W}_i^j$ with these t components $(\mathbf{V}_{r1}^{j-1}, \ldots, \mathbf{V}_{rt}^{j-1})$, and for each single component $\mathbf{V}_{ri}^{j-1}$, we associate it with the set of the approximation components $(\mathbf{V}_{i\in I_{ri}}^j)$ needed for its reconstruction

$$\mathbf{W}_i^j \xrightarrow{\text{Reconst.}} \begin{cases} \mathbf{V}_{r1}^{j-1} \xleftarrow{\text{Reconst.}} & \mathbf{V}_{i\in I_{r1}}^j \\ \quad\vdots & \quad\vdots \\ \mathbf{V}_{rt}^{j-1} \xleftarrow{\text{Reconst.}} & \mathbf{V}_{i\in I_{rt}}^j. \end{cases}$$

For the Haar filter, this association is simple

$$\mathbf{W}_i^j \xrightarrow{\text{Reconst.}} \begin{cases} \mathbf{V}_{2i-1}^{j-1} \xleftarrow{\text{Reconst.}} & \mathbf{V}_i^j \\ \mathbf{V}_{2i}^{j-1} \xleftarrow{\text{Reconst.}} & \mathbf{V}_i^j. \end{cases}$$

Now, if, in the regression model, to predict $\mathbf{W}_i^j$, we involve the approximation components $\mathbf{V}_{i\in\mathfrak{I}}^j$, where $\mathfrak{I} = \cup(I_{r1}, \ldots, I_{rt})$ is the union of these t sets of indices of approximation components in the scale j, then this model will largely preserve the scalability of the original DWT. Note that the intersection $\cap(I_{r1}, \ldots, I_{rt})$ could be empty, depending on the wavelet filter used. It is beyond the scope of this paper to discuss the model that exactly preserves the component scalability for any arbitrary filter. Nevertheless, we propose a model that perfectly preserves the scalability for the Haar filter, with $\mathfrak{I} = \{i\}$. It is given by

$$\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}_i^j\right].$$

In this paper, we propose the following model for Haar, involving $\mathbf{V}_i^j$ and its second- and third-order terms:

$$\widehat{\mathbf{W}}_i^j = \boldsymbol{\beta}_{i,0}^j + \boldsymbol{\beta}_{i,1}^j \mathbf{V}_i^j + \boldsymbol{\beta}_{i,2}^j \left(\mathbf{V}_i^j\right)^2 + \boldsymbol{\beta}_{i,3}^j \left(\mathbf{V}_i^j\right)^3. \tag{12}$$

The *restricted model* preserves the component scalability requirement of the original DWT by involving a restricted number of components in the prediction model. An added benefit due to the restricted number of components is a reduction of computational complexity. Higher order terms of $(\mathbf{V}_i^j)^n$ can be added to improve the predictive power. Additional improvements may be had by adding more components $\mathbf{V}_{k\neq i}^j$, at the cost of a decrease in scalability and an increase in complexity and side information corresponding to the additional regression coefficients.

3) Exogenous Model: The *maximum model* employs a large number of components in each estimator, and accurate estimation is expected. However, this large model can result in large computational cost when computing the least squares solution for each image. This method may also result in excessive side information needed to be transmitted in order to inform the decoder of the model parameters. Since each hyperspectral image from the same sensor may have similar statistical relationships

TABLE I
SIZE OF THE SIDE INFORMATION. z IS THE NUMBER OF COMPONENTS, AND l IS THE NUMBER OF WAVELET DECOMPOSITION LEVELS

Model	Number of coefficients	$z=224$ $l=8$
RWA (*Maximum*)	$\frac{z^2}{3}(1-\frac{1}{2^{2l}})+z(1-\frac{1}{2^l})$	16947
RWA (*Restricted*)	$2z(1-\frac{1}{2^l})$	446
RWA (*Exogenous*)	–	–
PCA	z^2+z	50400

among its components, the *exogenous model* computes the regression coefficients $\overline{\boldsymbol{\beta}}^j$ over a set of training images and then uses these coefficients for all new images that come from the same sensor. As a consequence, the computational cost of RWA is reduced considerably to roughly the same as that of the DWT. Additionally, no side information needs to be stored for each individual image. The model is given by

$$\widehat{\mathbf{W}}_i^j = f_i[\mathbf{V}^j] = \overline{\boldsymbol{\beta}}_{i,0}^j + \overline{\boldsymbol{\beta}}_{i,1}^j \mathbf{V}_1^j + \cdots + \overline{\boldsymbol{\beta}}_{i,k}^j \mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}.$$

4) Side information: Table I reports the side information for the three proposed models of RWA in comparison with the side information required for PCA, consisting of the z^2 transform coefficients and the z means used to center the data prior to transformation. At each level j, the prediction of each component $\mathbf{W}_i^j$ requires $1+z\cdot 2^{-j}$ regression coefficients $(\boldsymbol{\beta}_{i,\cdot}^j)$ when using the *maximum model* [see (11)] and four coefficients when using the *restricted model* [see (12)].

These three regression models are suggested to be used for different—possibly overlapping—situations. The *maximum model* is suggested in most of the cases since it is expected to give the most accurate estimation, although the regression coefficients $\boldsymbol{\beta}$ are needed as side information. For situations where the computation cost or the size of the side information matters, the *exogenous model* is suggested. Finally, the *restricted model* is suggested for those cases where the computational cost matters or the spectral component scalability is required.

D. Fast RWA

The ordinary least squares method estimates the regression coefficients with a complexity cost of $O(mk^2)$ [58], [59], where k is the number of components used to form the prediction and m is the number of spatial samples. In the *maximum model*, $k = z\cdot 2^{-j}$ at level j, where z is the number of spectral components in the original image. The *restricted model* uses less input components to the predictors, leading to a reduced complexity $O(mk'^2) \quad (k'<k)$, depending on the wavelet filter. For example, for the Haar DWT, $k'=3$ [see (12)]. However, assuming that $m \gg z$, the number of spatial samples m will dominate the complexity. The fast RWA proposed here addresses this problem by randomly selecting a subset of $m'=\rho\cdot m$ samples $(m' \ll m)$. The least squares optimization is then carried out only on this small data set to obtain the regression coefficients. For images with a large spatial dimension $(m \gg z)$, the subsampling employed in the fast RWA has minimal impact on the results. However, for images with a small spatial dimension, the *exogenous model* may be a better choice, particularly when side information is considered.

III. RELEVANT PROPERTIES OF RWA

On the one hand, the redundancy reduction ability of RWA for remote sensing imagery is a major advantage from the fundamental point of view. On the other hand, from the applied perspective, component scalability and the possibility of efficient integer representations are also major advantages of RWA.

A. Reduction of Correlation and MI

Hyperspectral images usually exhibit significant redundancy along the spectral dimension. The spectral transform aims at exploiting this redundancy among components, so that they can be coded independently. In this section, we analyze the effect of the transform on mutual information (MI) and correlation. MI is not limited to linear relations between variables (as is the case with correlation), but it also takes into account eventual nonlinear relations [60]. As a result, MI is a more general description of statistical relations than correlation. MI includes correlation and non-Gaussianity [61]. Therefore, changes in MI may come either from better decorrelation or from removing higher order statistical relations.

Here, we measure the correlation and MI between the coefficients for the three more representative transformations in this paper: wavelets (Haar with eight decomposition levels), PCA, and the proposed RWA (*maximum model* using the Haar filter). We have taken the Haar and PCA transformations because they provide extreme cases of previous work from the literature. In particular, the Haar transform has low complexity and provides fine grain scalability but has modest compression performance. On the other hand, the PCA provides superior compression performance at the cost of high complexity and no scalability.

The following analysis employed the hyperspectral image "Hawaii" (314 368 spatial samples and 224 spectral components). The matrices depicted as images in Fig. 3 represent the statistical relations between the coefficients of transformed versions of the image. The left column of matrices represents correlation, while the right column of matrices represents MI. Each row of matrices corresponds to one (transformed) image representation. For a given matrix, the element M_{kl} represents the quantity of interest (correlation or MI) between the k-*th* and l-*th* components.

The degree to which these matrices appear to be diagonal has been used to describe the suitability of a domain for scalar (coefficientwise) coding [33], [62], [63]. Specifically, better representations for transform coding are those with small off-diagonal values. To quantify this property, the average value of the off-diagonal coefficients is depicted in the figure. The nondiagonal nature of the matrices in the top row shows that the original spectral domain is highly unsuitable for scalar coding. Subsequent rows indicate that significant improvements can be obtained via the application of transforms.

To compare the residual correlation that exists after the application of the three transforms, we turn our attention to the left-hand side of the figure. As can be seen there, the DWT

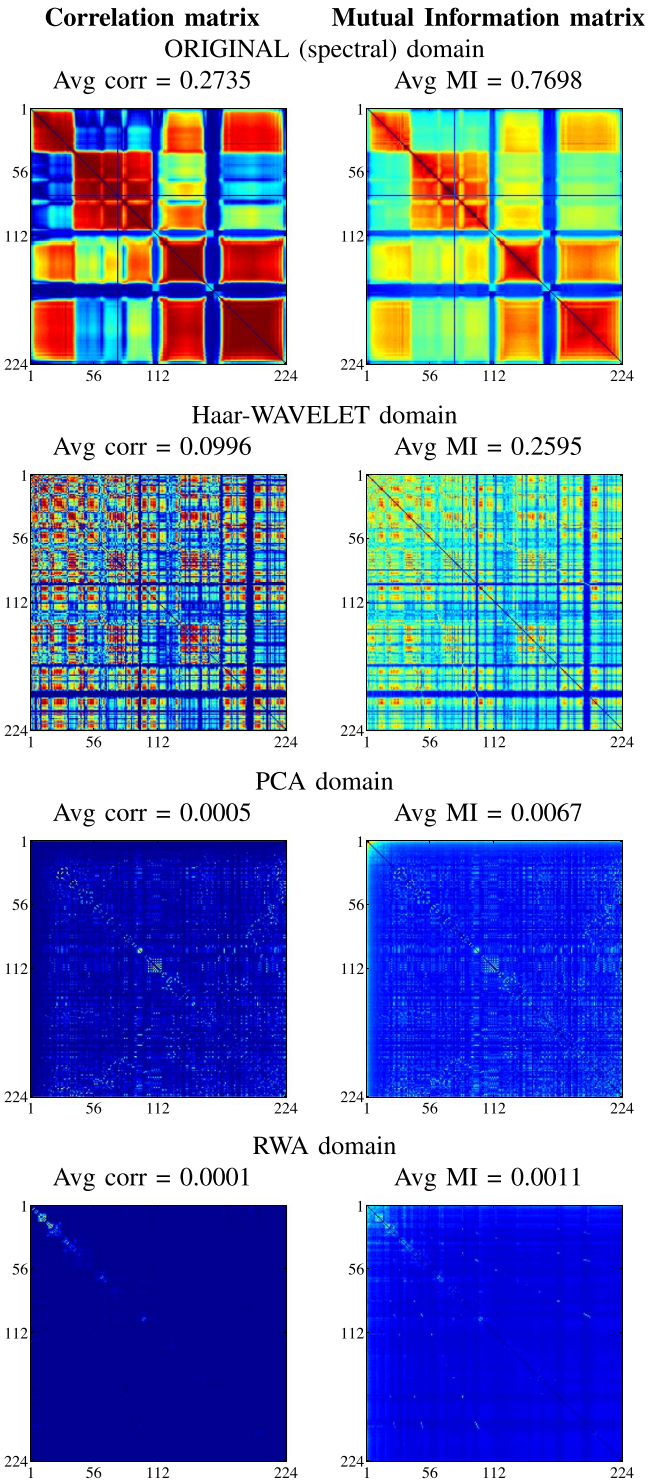

Fig. 3. (Left column) Correlation and (right column) MI matrices for the coefficients of an illustrative hyperspectral image, "Hawaii" with 224 spectral bands. Each row of the figure corresponds to a different representation domain. The blue-to-red color map indicates low-to-high dependence values. Average values are computed using the magnitude of the off-diagonal elements. MI is given in bits. Matrices in each column are normalized in the same way for fair comparison. For the purpose of visualizing the small off-diagonal values, their fourth roots are displayed.

provides significant decorrelation. However, strong correlations still exist among and between the components of certain subbands in the transform domain. PCA achieves significantly better decorrelation than the DWT. By definition, it is designed to diagonalize the covariance matrix. However, in practical applications, total decorrelation is hard to achieve due to implementation constraints (e.g., reversible integer-to-integer implementation) and the difference between the sample and the actual covariance matrices. Thus, some residual correlation is still present. It can be seen that, after applying RWA (*maximum model*), almost full decorrelation is reached in the transform domain, as evidenced by the improved diagonal appearance and the smaller off-diagonal average. This illustrates the clear advantage of RWA over the usual Haar DWT and even a small advantage over PCA. Similar comments can be made for MI, as reported in bits, in the right side of the figure. These observations serve to explain why RWA obtains superior compression results compared to the usual DWT as well as comparable performance to that of PCA. These statements are verified in the compression results presented hereinafter.

TABLE II
COMPONENT SCALABILITY FOR DIFFERENT TRANSFORMS. THE NUMBER OF REQUIRED TRANSFORM COMPONENTS TO RECONSTRUCT ONE SINGLE ORIGINAL COMPONENT. z IS THE NUMBER OF COMPONENTS, AND l IS THE NUMBER OF TRANSFORM LEVELS

Method	Requirement components	z= 224 l=3
RWA (*Maximum model*)	$z(1-\frac{1}{2^l})$	196
RWA (*Restricted model*)	$l+1$	4
RWA (*Exogenous model*)	$z(1-\frac{1}{2^l})$	196
PCA	z	224
DWT 5/3 (even component)	$2l+1$	7
DWT 5/3 (odd component)	$3l+2$	11
Haar	$l+1$	4

Note that MI can be expressed as a sum of a correlation-dependent index, a global negentropy, and a marginal negentropy [61]. Reducing the correlation implies reducing the MI in most cases. However, given the invariance of global negentropy to linear transforms, once the correlation has been removed (through PCA) and the marginal negentropy (or sparsity) has been maximized (through ICA), the global negentropy can only be reduced through nonlinear means. The results in Fig. 3, obtained using linear regression, suggest that the prediction in RWA increases the sparsity of the representation. More sophisticated (nonlinear) regression techniques could be included in the RWA framework [functions f_i in (6)] to reduce the global negentropy at the cost of training, complexity, and side information.

B. Reversible Integer RWA and Dynamic Range

For lossless coding, it is required that the redundancy reduction transform maps integer coefficients to integer coefficients. For wavelets, the integer versions can be achieved by applying the lifting scheme [64], [65]. For instance, integer versions of the Haar filter are computed and reversed as follows:

$$\text{Forward}: \begin{cases} \mathbf{W}_i^j = \mathbf{V}_{2i}^{j-1} - \mathbf{V}_{2i-1}^{j-1} \\ \mathbf{V}_i^j = \mathbf{V}_{2i-1}^{j-1} + \left\lfloor \frac{1}{2}\mathbf{W}_i^j \right\rfloor \end{cases} \tag{13}$$

$$\text{Reverse}: \begin{cases} \mathbf{V}_{2i-1}^{j-1} = \mathbf{V}_i^j - \left\lfloor \frac{1}{2}\mathbf{W}_i^j \right\rfloor \\ \mathbf{V}_{2i}^{j-1} = \mathbf{W}_i^j + \mathbf{V}_{2i-1}^{j-1}. \end{cases} \tag{14}$$

TABLE III
COMPUTATIONAL COMPLEXITY IN FLOPS FOR RWA. z IS THE NUMBER OF SPECTRAL COMPONENTS, m IS THE NUMBER OF SAMPLES PER COMPONENT, AND l IS THE NUMBER OF WAVELET DECOMPOSITION LEVELS. k_i IS THE NUMBER OF DETAIL COMPONENTS EMPLOYED IN THE PREDICTION AT LEVEL i

Operation	Floating-point operations (FLOP) for lossless RWA
Integer Haar	$8(1-\frac{1}{2^l})mz$
$\boldsymbol{\beta}=(\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$	$\sum_{i=1}^{l}(2m-1)(k_i+1)^2+(k_i+1)^3+(\frac{z}{2^i})(k_i+1)\,[(2m-1)+(2k_i+1)]$
$\widehat{\mathbf{W}}=\boldsymbol{\beta}\mathbf{V}$	$2\sum_{i=1}^{l}(2k_i-1)m\frac{z}{2^i}$
Apply and Remove	$2m(z-1)$

Given that, we propose a reversible integer version of RWA based on an integer DWT decomposition $(\overline{\mathbf{V}}^j, \overline{\mathbf{W}}^j)$ and a quantized estimation removal $\overline{\mathbf{R}}^j = \overline{\mathbf{W}}^j - Q(\widehat{\mathbf{W}}^j)$. The scalar quantization is here performed through a simple rounding operation. The reversible integer RWA representation is then given by

$$\overline{\mathbf{RWA}}(\mathbf{V}^0, J) = \left(\overline{\mathbf{V}}^J, \left(\overline{\mathbf{R}}^j\right)^{1\le j\le J}\right).$$

In general, RWA can be based on any wavelet transform. Nevertheless, as this paper focuses on lossless coding, and for faster implementation, we focus exclusively on RWA based on the integer Haar transform [see (13)]. This transform, aside from its simplicity, provides another advantage, which consists of largely preserving the dynamic range of the original domain. Note that, with the integer Haar filter, the dynamic range can be expanded by only 1 b in the detail signal. In some applications, large dynamic range expansion can lead to serious problems, particularly for existing devices or systems that support only a limited bit depth [66].

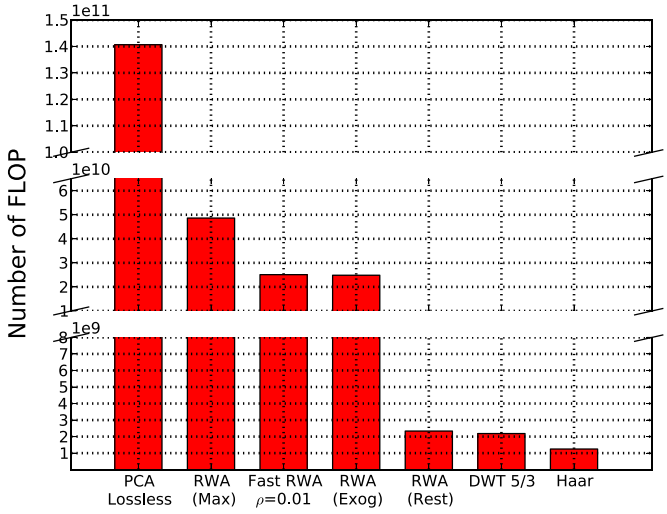

Fig. 4. Cost comparison in FLOPs for different transforms applied to the uncalibrated Yellowstone image with 224 spectral components and a spatial resolution of 512×680. The values reported for PCA and DWT 5/3 are from [18].

C. Component Scalability and Computational Cost

Table II reports the scalability of several spectral transforms in terms of the number of transformed components required to reconstruct a single original component. For the *restricted model*, the dependence for RWA is the same as that for the wavelet transform (Haar in this case).

The computational cost of RWA is dominated by the estimation of regression coefficients $\boldsymbol{\beta}=(\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$ and the generation of the predictions $\widehat{\mathbf{W}}=\boldsymbol{\beta}\mathbf{V}$. In the case of the *exogenous model*, the estimation of $\boldsymbol{\beta}$ is performed offline and does not contribute to the complexity of encoding. The computational cost of RWA is at its highest when using the *maximum model*. However, the cost of other efficient transforms such as PCA is still significantly higher. Furthermore, the integer implementation of the lossless PCA is usually performed by a factorization of the transform matrix [67], [68]. The resulting implementation is typically not computationally efficient and does not offer a high degree of parallelization due to the recursiveness of the associated reconstruction process.

Table III details the computational cost of RWA in floating-point operations (FLOPs), and Fig. 4 compares the computational cost of different transforms applied to a typical uncalibrated image from the AVIRIS sensor.

IV. EXPERIMENTAL RESULTS

This section presents experimental results for the proposed RWA applied to the lossless coding of hyperspectral images. Comparisons are provided with the most prominent methods in the state of the art.

A. Data Set and Coding Pipeline

Experimental evaluations were conducted using a set of images from three different sensors: AVIRIS [69], Hyperion [70], and IASI [71]. Table IV provides detailed information about these images. As described in the table, the AVIRIS images include five uncalibrated and five calibrated images corresponding to five scenes in Yellowstone. These images have a bit depth of 16 bits per pixel per component (bpppc). Also included are two uncalibrated AVIRIS images (Maine and Hawaii), each having a bit depth of 12 bpppc. All of the AVIRIS images have 224 spectral components and 512 lines. The widths of these images vary between 614 and 680. The Hyperion sensor produces images with 242 spectral components, each having a bit depth of 12 bpppc and a width of 256. The number of lines varies from image to image. The IASI is composed of a Fourier transform spectrometer and an integrated imaging subsystem. IASI Level 0 images have 8359 spectral components, each having 1528 lines of width 60. Level 1 IASI images are of size $8461 \times 1530 \times 60$.

The proposed coding system pipeline is shown in Fig. 5. A 1-D transform is applied in the spectral dimension followed by the 2-D JPEG2000 compression of each resulting transformed component. To this end, the Kakadu software implementation of JPEG2000 is employed with five levels of reversible 2-D

TABLE IV
DATA SET INFORMATION FOR AVIRIS, HYPERION, AND IASI SENSORS. z IS THE NUMBER OF SPECTRAL COMPONENTS, y IS THE HEIGHT, AND x IS THE WIDTH

Corpus	Image
AVIRIS (Uncal) z=224, y=512 x=680	Yellowstone, sc: 0, 3, 10, 11, 18 Hawaii (x=614) Maine
AVIRIS (Cal) z=224, y=512, x=677	Yellowstone, sc: 0, 3, 10, 11, 18
Hyperion (Uncal) z=242 x=256	ErtaAle (y=3187) Lake Monona (y=3176) Mt. St. Helens (y=3242)
Hyperion (Cal) z=242 x=256	Agriculture (y=3129) Coral Reef (y=3127) Urban (y=2905)
IASI Level 0 z=8359 y=1528 x=60	L0 1: 20091007093900Z L0 2: 20091007143900Z L0 3: 20100319050300S6 L0 4: 20120718075700Z L0 5: 20130116133300Z L0 6: 20130916080300Z
IASI Level 1 z=8461 y=1530 x=60	L1 1: 20130816230553Z L1 2: 20130817004753Z L1 3: 20130817041457Z L1 4: 20130817055657Z L1 5: 20130817073857Z

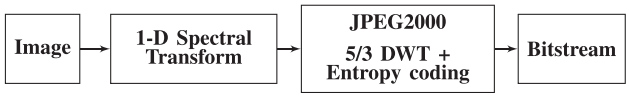

Fig. 5. Proposed coding pipeline: One-dimensional spectral transform followed by JPEG2000 standard, which includes a 2-D spatial DWT followed by bit-plane coding.

DWT 5/3. When a wavelet-based transform (including the proposed RWA) is employed in the spectral dimension, the maximum possible number of levels ($\lceil \log_2(z) \rceil$) is employed. For comparison with RWA, experimental results are provided for four other reversible spectral transforms, including the Haar and 5/3 wavelet transforms as well as PCA and pairwise orthogonal transform (POT) (a low-complexity approximation of PCA [20]). Also, for the purpose of comparison, experimental results are reported for the state-of-the-art predictive methods M-CALIC [72] and CCSDS-123 [51]. Rather than performing predictions in the transform domain as in the case of RWA, these methods perform prediction in the original (pixel) domain and are of particular interest since both are deemed appropriate for onboard hyperspectral image coding [73].

As noted in the aforementioned paragraph, five levels of spatial DWT are applied as part of the 2-D JPEG2000 coding process applied to the transformed components. This was found to yield good performance for all images and all transformed techniques with a small number of notable exceptions. For the for AVIRIS and IASI images, the PCA and RWA (*maximum model*) spectral transforms were found to be so efficient that no spatial DWT was necessary. In fact, in these cases, the application of the spatial DWT resulted in slightly inferior performance as compared to omitting the spatial DWT. For this reason, 0 levels of spatial DWT were performed as part of the 2-D JPEG2000 compression process in these cases.

Hyperion and IASI images are tall and narrow. That is, they have a large number of lines compared to columns. For this reason, experiments were performed on versions of these images that were spatially rotated 90° prior to compression. In the following section, results are reported for the rotation (0° or 90°) that provided the best performance, algorithm by algorithm. Specifically, for the uncalibrated Hyperion images, the 90° rotation was used for all algorithms except M-CALIC. For the calibrated Hyperion images, the 90° rotation was used for all algorithms except M-CALIC and CCSDS-123. For the IASI images, the 90° rotation was only used for the POT-based algorithm. In most cases, the differences between rotating and not rotating are minor. On the other hand, for CCSDS-123 and the POT-based algorithm, the differences can be significant. The POT is a line-based transform that produces side information for each line of the input image. Hence, the rotation results in a small number of lines giving rise to a reduced amount of side information.

For reproducibility of results, the basic Matlab source code for our implementations of the PCA, Haar, DWT 5/3, and RWA spectral transforms is available online at [74].

B. Results

Table V provides the lossless coding results for all systems under test in terms of the bit rate, in bpppc. All necessary side information is included in the results reported in the table. The compression efficiency of each algorithm can be appreciated by observing the degree to which its resulting bit rate falls below the bit depth of the original image (e.g., 12 or 16 bpppc).

It can be observed that, overall, the proposed RWA outperforms the state-of-the-art methods included in the comparison. RWA, in one of its modalities, is the best in three out of the six data subsets (three sensors, calibrated and uncalibrated). When it is not the best, it yields results within 0.1 bpppc of the best technique. The worst such cases occur for Hyperion uncalibrated images that tend to have streaking artifacts due to the nature of the pushbroom sensors [70]. More detailed observations are now provided for each specific sensor.

For both uncalibrated and calibrated (radiance) AVIRIS images, on average, RWA (*maximum model*) outperforms PCA and POT by more than 0.1 and 0.5 bpppc, respectively. RWA also outperforms the predictive methods M-CALIC and CCSDS-123 by more than 0.2 bpppc. When compared with the HAAR and DWT 5/3 wavelet spectral transforms, the coding gain of RWA is larger than 1.2 and 0.9 bpppc, respectively.

For uncalibrated Hyperion images, the coding performances of Haar, DWT 5/3, PCA, and RWA (*maximum model*) are very similar. CCSDS-123 provides the best results, followed by POT and RWA (*restricted model*), which benefits from the 2-D spatial transform more than PCA or RWA (*maximum model*). In this case, RWA (*restricted model*) provides improvement over the (usual) wavelet spectral transforms by about 0.2 bpppc with a low computational cost. For calibrated Hyperion images, RWA (*maximum model*) achieves a coding performance similar to that of M-CALIC and outperforms POT and PCA by around 0.4 bpppc. It outperforms CCSDS-123 by 0.3 bpppc and improves over the wavelet transforms by more than 0.7 bpppc.

TABLE V
LOSSLESS BIT RATE FOR HAAR, DWT 5/3, POT, PCA, AND RWA SPECTRAL TRANSFORMS WITH JPEG2000, AND M-CALIC AND CCSDS-123 CODING TECHNIQUES IN TERMS OF BPPPC. THE LOWER THE BIT RATE, THE BETTER THE PERFORMANCE

Image / Transform or coding technique	Haar	DWT 5/3	POT	PCA	RWA (Restricted)	RWA (Maximum)	Fast RWA (ρ)	RWA (Exogenous)	M-CALIC	CCSDS-123
AVIRIS Uncalibrated ($\rho = 0.01$)										
Yellowstone 00	7.52	7.16	6.64	**6.05**	6.67	6.07	6.11	Train	6.34	6.40
Yellowstone 03	7.37	7.03	6.46	**5.88**	6.50	5.90	5.94	5.95	6.16	6.26
Yellowstone 10	6.28	6.11	5.77	**5.44**	5.81	5.46	5.52	5.55	5.54	5.66
Yellowstone 11	6.91	6.63	6.24	**5.74**	6.44	5.76	5.81	5.81	5.93	5.93
Yellowstone 18	7.52	7.15	6.70	**5.95**	6.76	5.97	6.01	6.05	6.35	6.50
Average 16 bpppc	7.12	6.82	6.36	**5.81**	6.44	5.83	5.88	5.89	6.06	6.15
Maine	3.52	3.43	3.09	3.22	3.08	**2.69**	2.71	Train	2.89	2.77
Hawaii	3.37	3.27	2.98	3.20	2.95	**2.55**	2.56	2.82	2.84	2.70
Average 12 bpppc	3.44	3.35	3.03	3.21	3.01	**2.62**	2.63	2.68	2.86	2.73
Global Average	6.07	5.83	5.41	5.06	5.46	**4.91**	4.95	4.97	5.14	5.17
AVIRIS Calibrated (Radiance, $\rho = 0.01$)										
Yellowstone 00	5.13	4.75	4.45	3.85	4.45	3.74	**3.73**	Train	4.13	3.90
Yellowstone 03	5.00	4.64	4.30	3.74	4.31	**3.58**	3.60	3.63	3.98	3.76
Yellowstone 10	4.01	3.84	3.66	3.40	3.65	**3.18**	3.20	3.25	3.39	3.35
Yellowstone 11	4.55	4.25	4.08	3.67	4.06	**3.47**	3.49	3.52	3.73	3.58
Yellowstone 18	5.12	4.71	4.50	3.80	4.50	**3.65**	3.67	3.72	4.10	3.92
Average 16 bpppc	4.76	4.44	4.20	3.69	4.20	**3.52**	3.54	3.57	3.87	3.70
Hyperion Uncalibrated (Rotated 90°, $\rho = 0.01$)										
Erta Ale	4.40	4.44	**4.21**	4.49	4.26	4.46	4.47	Train	4.76	4.24
Lake Monona	4.52	4.55	4.41	4.62	4.46	4.63	4.64	4.63	4.92	**4.36**
Mt. St. Helens	4.57	4.57	4.31	4.54	4.40	4.52	4.53	4.52	4.83	**4.28**
Average	4.54	4.56	4.31	4.55	4.38	4.54	4.55	4.54	4.83	**4.29**
Hyperion Calibrated (Rotated 90°, $\rho = 0.01$)										
Agricultural	6.20	6.35	5.92	5.88	5.55	**5.44**	**5.44**	Train	**5.44**	5.73
Coral Reef	5.54	5.80	5.44	5.53	5.19	5.11	5.11	5.18	**5.05**	5.47
Urban	6.29	6.39	5.99	5.89	5.58	5.46	5.46	5.50	**5.44**	5.75
Average	6.01	6.18	5.78	5.76	5.44	5.34	5.34	5.37	**5.31**	5.65
IASI Uncalibrated (Level 0, $\rho = 0.1$)										
IASI L0 1	3.00	3.02	2.87	–	2.85	2.35+0.9	2.38+0.9	Train	2.89	2.90
IASI L0 2	2.99	3.02	2.86	–	2.84	2.34+0.9	2.37+0.9	**2.39**	2.87	2.88
IASI L0 3	2.98	3.01	2.86	–	2.84	2.35+0.9	2.38+0.9	**2.40**	2.87	2.90
IASI L0 4	3.03	3.05	2.89	–	2.87	2.35+0.9	2.38+0.9	**2.41**	2.91	2.91
IASI L0 5	2.96	2.99	2.86	–	2.89	2.35+0.9	2.38+0.9	**2.40**	2.86	2.90
IASI L0 6	3.06	3.08	2.90	–	2.89	2.37+0.9	2.39+0.9	**2.41**	2.92	2.91
Average	3.00	3.02	2.87	–	2.85	2.35+0.9	2.38+0.9	**2.40**	2.88	2.90
IASI Calibrated (Level 1, $\rho = 0.1$)										
IASI L1 1	7.37	7.46	7.12	7.22+2.4	7.10	6.42+0.9	6.45+0.9	Train	6.86	6.60
IASI L1 2	7.37	7.09	7.12	7.05+2.4	7.10	6.41+0.9	6.44+0.9	**6.48**	6.88	6.61
IASI L1 3	7.38	7.11	7.14	–	7.13	6.42+0.9	6.45+0.9	**6.49**	6.88	6.59
IASI L1 4	7.33	7.06	7.12	–	7.10	6.43+0.9	6.46+0.9	**6,50**	6.87	6.56
IASI L1 5	7.40	7.12	7.15	–	7.13	6.43+0.9	6.46+0.9	**6.50**	6.87	6.60
Average	7.37	7.16	7.13	7.13+2.4	7.11	6.42+0.9	6.45+0.9	**6.49**	6.87	6.59

For IASI images, which have more than 8000 components, lossless PCA encounters two serious problems. First is a huge increment in computational cost, which results in an implausible execution time. Second is a huge increase in the required side information. For instance, the PCA matrix for IASI Level 1 images has a size of 8461×8461, while the spatial dimension is only 1530×60, implying a side information of 2.4 bpppc. For this reason, results for PCA applied to IASI images are only reported for a couple of examples. On the other hand, RWA (*maximum model*) does not suffer from extreme complexity, but side information plays a nonnegligible role. This problem can be solved by using RWA (*exogenous model*). As noted in the table, the results for this method were obtained using only one training image per sensor to learn the regression coefficients. For IASI Level 0 images, the RWA *exogenous model* outperforms POT, M-CALIC, and CCSDS-123 by about 0.4 bpppc. It is superior to the Haar and 5/3 wavelet transforms by about 0.5 bpppc. For IASI Level 1 images, the RWA *exogenous model* achieves competitive coding results compared to CCSDS-123 while outperforming PCA, POT, and wavelets by more than 0.5 bpppc and M-CALIC by about 0.2 bpppc.

With regard to the results for calibrated versus uncalibrated images, we note that, for the AVIRIS sensor, the calibrated images yield better performance, while for Hyperion and IASI, we

find the opposite situation, where uncalibrated images achieve better performance. Similar results were observed in [20] for AVIRIS and Hyperion.

It is worth pointing out that fast RWA is a straightforward approach to reduce the cost of computing the regression coefficients. For the AVIRIS and Hyperion sensors, only 1% of the pixels were used for this computation while achieving a performance almost identical to that obtained when using 100% of the pixels. For the IASI sensor, 10% of the pixels had to be used in order to approximately maintain the same performance due to the low degree of freedom $(m - z - 1)$. It is also worth highlighting the fact that the RWA *exogenous model* eliminates the need for side information as well as the cost of computing the regression coefficients online.

V. Conclusion

Remote sensing is becoming increasingly widespread and finds application in a growing number of fields. At the same time, the size of transmitted remote-sensing data has increased and is foreseen to continue increasing, which begs for efficient ways to improve its storage and dissemination. Data compression plays a key role in this regard. In particular, lossless data compression has seen numerous recent advances, spanning prediction-based to transform-based techniques. In this regard, much work has focused on the spectral dimension of hyperspectral images. PCA usually provides the best performance among transform-based techniques but entails some drawbacks. Several approaches have been proposed to partially tackle these drawbacks while yielding competitive coding performance.

In this paper, we introduce a novel spectral redundancy reduction transform that builds upon the low-complexity Haar wavelet transform and exploits the remaining redundancy among wavelet-transformed components through regression analysis. The suggested RWA is capable of estimating, via regression, the detail components from the approximation components resulting from the wavelet transform. RWA allows for an integer-to-integer implementation and perfect reconstruction and, thus, for lossless compression.

The regression model can be devised to account for finer estimation accuracy, for finer spectral component scalability, or for lower computational complexity, which give rise to the *maximum model*, the *restricted model*, and the *exogenous model*, respectively. The first model involves all the approximation components in the regression, the second model involves only the approximation components as required to maintain the same component scalability as the baseline Haar wavelet transform, and the third model computes the regression coefficients based on training data from the appropriate hyperspectral sensor and employs the learned coefficients to all the other images in the same sensor. However, another variant is a fast RWA implementation, which performs a fast prediction using the least squares methods based only on spatially subsampled data, reducing the time to compute the regression coefficients.

Extensive experimental results for the lossless compression of images from three different popular and widely used hyperspectral sensors have been carried out. Specifically, both calibrated and uncalibrated images have been employed from the airborne AVIRIS sensor as well as the satellite-based Hyperion and IASI sensors. The resulting coding performance suggests that RWA yields, overall, the best achievement as compared to other spectral transforms such as Haar, DWT 5/3, PCA, or its low-complexity approximation POT. RWA also provides superior performance compared to other prominent prediction-based coding techniques such as M-CALIC and the current CCSDS-123 standard.

To summarize, RWA provides a tradeoff between computational complexity and coding performance that makes it an appealing approach for remote sensing lossless data compression. It offers additional desirable features such as limited dynamic range increase and superior component scalability.

References

[1] V. Laparra, S. Jiménez, D. Tuia, G. Camps-Valls, and J. Malo, "Principal polynomial analysis," *Int. J. Neural Syst.*, vol. 24, no. 7, 2014, Art. no. 1440007.

[2] N. Amrani, V. Laparra, G. Camps-Valls, J. Serra-Sagrista, and J. Malo, "Lossless coding of hyperspectral images with principal polynomial analysis," in *Proc. IEEE ICIP*, Oct. 2014, pp. 4023–4026.

[3] V. Laparra, J. Malo, and G. Camps-Valls, "Dimensionality reduction via regression in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1026–1036, Aug. 2015.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer-Verlag, 2009.

[5] X. Tang and W. A. Pearlman, "Three-dimensional wavelet-based compression of hyperspectral images," in *Hyperspectral Data Compression*. New York, NY, USA: Springer-Verlag, 2006, pp. 273–308.

[6] J. E. Fowler and J. T. Rucker, "3D wavelet-based compression of hyperspectral imager," in *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ, USA: Wiley, 2007, pp. 379–407.

[7] B. Penna, T. Tillo, E. Magli, and G. Olmo, "A new low complexity KLT for lossy hyperspectral data compression," in *Proc. IEEE IGARSS*, 2006, vol. 7, pp. 3525–3528.

[8] J. Zhang and G. Liu, "A novel lossless compression for hyperspectral images by context-based adaptive classified arithmetic coding in wavelet domain," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 461–465, Jul. 2007.

[9] J. Zhang, J. E. Fowler, and G. Liu, "Lossy-to-lossless compression of hyperspectral imagery using three-dimensional TCE and an integer KLT," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 814–818, Oct. 2008.

[10] F. García-Vilchez and J. Serra-Sagristà, "Extending the CCSDS recommendation for image data compression for remote sensing scenarios," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, pp. 3431–3445, Oct. 2009.

[11] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Transform coding techniques for lossy hyperspectral data compression," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1408–1421, May 2007.

[12] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Progressive 3-D coding of hyperspectral images based on JPEG2000," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 125–129, Jan. 2006.

[13] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer-Verlag, 2002.

[14] I. P. Akam Bita, M. Barret, and D. Pham, "On optimal transforms in lossy compression of multicomponent images with JPEG2000," *Signal Process.*, vol. 90, no. 3, pp. 759–773, 2010.

[15] M. Effros, H. Feng, and K. Zeger, "Suboptimality of the Karhunen-Loeve transform for transform coding," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1605–1619, Aug. 2004.

[16] P. Hao and Q. Shi, "Reversible integer KLT for progressive-to-lossless compression of multiple component images," in *Proc. ICIP*, Sep. 2003, vol. 1, pp. I-633–I-636.

[17] Q. Du and J. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *Int. J. High Perform. Comput. Appl.*, vol. 22, no. 4, pp. 438–448, 2008.

[18] I. Blanes and J. Serra-Sagristà, "Clustered reversible-KLT for progressive lossy-to-lossless 3d image coding," in *Proc. IEEE DCC*, Mar. 2009, pp. 233–242.

[19] I. Blanes, J. Serra-Sagrista, M. Marcellin, and J. Bartrina-Rapesta, "Divide-and-conquer strategies for hyperspectral image processing: A review of their benefits and advantages," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 71–81, May 2012.

[20] I. Blanes and J. Serra-Sagristà, "Pairwise orthogonal transform for spectral image coding," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 961–972, Mar. 2011.

[21] M. Barret, J.-L. Gutzwiller, and M. Hariti, "Low-complexity hyperspectral image coding using exogenous orthogonal optimal spectral transform (OrthOST) and degree-2 zerotrees," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1557–1566, May 2011.

[22] I. P. Akam Bita, M. Barret, F. D. Vedova, and J.-L. Gutzwiller, "Lossy and lossless compression of MERIS hyperspectral images with exogenous quasi-optimal spectral transforms," *J. Appl. Remote Sens.*, vol. 4, Jul. 2010, Art. no. 041790.

[23] I. P. Akam Bita, M. Barret, and D. Pham, "On optimal orthogonal transforms at high bit-rates using only second order statistics in multicomponent image coding with JPEG2000," *Signal Process.*, vol. 90, no. 3, pp. 753–758, Jan. 2010.

[24] M. Barret, J.-L. Gutzwiller, I. P. Akam Bita, and F. D. Vedova, "Lossy hyperspectral images coding with exogenous quasi optimal transforms," in *Proc. IEEE DCC*, Mar. 2009, pp. 411–419.

[25] P. Birjandi and M. Datcu, "Multiscale and dimensionality behavior of ICA components for satellite image indexing," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 103–107, Jan. 2010.

[26] G. Camps-Valls, D. Tuia, L. Gomez-Chova, S. Jimenez, and J. Malo, *Remote sensing image processing*, ser. Synthesis Lectures on Image, Video and Multimedia Processing, A. Bovik, Ed. Austin, TX, USA: Morgan & Claypool, 2011.

[27] V. Laparra and R. Santos-Rodriguez, "Spatial/spectral information tradeoff in hyperspectral images," in *Proc. IEEE IGARSS*, Jul. 2015, pp. 1124–1127.

[28] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, ser. The Kluwer International Series in Engineering and Computer Science. Boston, MA, USA: Kluwer, 1992, no. 159.

[29] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, USA: Wiley, 2001.

[30] X. Tang, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding," in *Proc. Electron. Imaging, Int. Soc. Opt. Photon.*, 2003, pp. 1037–1047.

[31] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc 31st Asilomar Conf. Signals, Syst. Comput.*, Nov. 2–5, 1997, vol. 1, pp. 673–678.

[32] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.

[33] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, "Non-linear image representation for efficient perceptual coding," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 68–80, Jan. 2006.

[34] J. Malo and J. Gutiérrez, "V1 non-linear properties emerge from local-to-global non-linear ICA," *Network: Comput. Neurol. Syst.*, vol. 17, no. 1, pp. 85–102, 2006.

[35] J. Malo and V. Laparra, "Psychophysically tuned divisive normalization approximately factorizes the PDF of natural images," *Neurol. Comput.*, vol. 22, no. 12, pp. 3179–3206, 2010.

[36] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[37] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[38] J. Arenas-García, K. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 16–29, Jul. 2013.

[39] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE J.*, vol. 37, no. 2, pp. 233–243, 1991.

[40] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[41] M. Scholz, M. Fraunholz, and J. Selbig, *Nonlinear Principal Component Analysis: Neural Networks Models and Applications*. Berlin, Germany: Springer-Verlag, 2007, ch. 2, pp. 44–67.

[42] V. Laparra, G. Camps, and J. Malo, "Iterative Gaussianization: From ICA to random rotations," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 537–549, Apr. 2011.

[43] U. Ozertem and D. Erdogmus, "Locally defined principal curves and surfaces," *J. Mach. Learn. Res.*, vol. 12, no. 4, pp. 1249–1286, Apr. 2011.

[44] V. Laparra, S. Jimenez, G. Camps, and J. Malo, "Nonlinearities and adaptation of color vision from sequential principal curves analysis," *Neural Comput.*, vol. 24, no. 10, pp. 2751–2788, 2012.

[45] V. Laparra and J. Malo, "Visual aftereffects and sensory nonlinearities from a single statistical framework," *Front. Human Neurosci.*, vol. 9, no. 557, 2015. [Online]. Available: http://www.frontiersin.org/human_neuroscience/10.3389/fnhum.2015.00557/abstract

[46] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Netw.*, vol. 12, no. 3, pp. 429–439, 1999.

[47] S. Goyal and J. O Neal, "Entropy coded differential pulse-code modulation systems for television systems," *IEEE Trans. Commun.*, vol. 23, no. 6, pp. 660–666, Jul. 1975.

[48] N. Farvadin and J. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 402–418, May 1985.

[49] M. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: A low complexity, context-based, lossless image compression algorithm," in *Proc. Data Compress. Conf.*, 1996, pp. 140–149.

[50] R. Rinaldo and G. Calvagno, "Image coding by block prediction of multiresolution subimages," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 909–920, Jul. 1995.

[51] Consultative Committee for Space Data Systems (CCSDS), *Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1*, ser. Blue Book, May 2012. [Online]. Available: http://public.ccsds.org/publications/archive/123x0b1ec1.pdf

[52] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[53] A. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Proc. Int. Conf. Image Process.*, Oct. 1997, vol. 1, pp. 596–599.

[54] J. Jing Zheng, J. Yun Fang, and C. de Han, "Reversible integer wavelet evaluation for DEM progressive compression," in *Proc. IEEE IGARSS*, Jul. 2009, vol. 5, pp. V-52–V-55.

[55] P. F. Craigmile and D. B. Percival, "Asymptotic decorrelation of between-scale wavelet coefficients," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 1039–1048, Mar. 2005.

[56] C. Thiebaut, E. Christophe, D. Lebedeff, and C. Latry, "CNES studies of on-board compression for multispectral and hyperspectral images," in *Proc. Opt. Eng. Appl., Int. Soc. Opt. Photon.*, 2007, pp. 1–15.

[57] J. Nocedal and S. J. Wright, *Least-Squares Problems*. New York, NY, USA: Springer-Verlag, 2006.

[58] P. Bloomfield and W. L. Steiger, *Least Absolute Deviations: Theory, Applications, and Algorithms*. Boston, MA, USA: Birkhäuser, 1983.

[59] D. M. Bates and D. G. Watts, *Nonlinear Regression: Iterative Estimation and Linear Approximations*. New York, NY, USA: Wiley, 1988.

[60] T. M. Cover and J. A. Tomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[61] J. Cardoso, "Dependence, correlation and Gaussianity in independent component analysis," *J. Mach. Learn. Res.*, vol. 4, no. 7, pp. 1177–1203, 2003.

[62] R. Clarke, *Transform Coding of Images*. New York, NY, USA: Academic, 1985.

[63] J. Malo, R. Navarro, I. Epifanio, F. Ferri, and J. Artigas, "Non-linear invertible representation for joint statistical and perceptual feature decorrelation," *Lecture Notes Comput. Sci.*, vol. 1876, pp. 658–667, 2000.

[64] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmonic Anal.*, vol. 3, no. 2, pp. 186–200, 1996.

[65] A. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Appl. Comput. Harmonic Anal.*, vol. 5, no. 3, pp. 332–369, 1998.

[66] I. Blanes, M. Hernandez-Cabronero, F. Auli-Llinas, J. Serra-Sagrista, and M. Marcellin, "Isorange pairwise orthogonal transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3361–3372, Jun. 2015.

[67] P. Hao and Q. Shi, "Matrix factorizations for reversible integer mapping," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2314–2324, Oct. 2001.

[68] L. Galli and S. Salzo, "Lossless hyperspectral compression using KLT," *Proc. IEEE IGARSS*, 2004, vol. 1–7, pp. 313–316.

[69] Jet Propulsion Laboratory, NASA, *Airborne Visible InfraRed Imaging Spectrometer Website*. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.overview.html

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

[71] Infrared Atmospheric Sounding Interferometer (IASI). [Online]. Available: https://wdc.dlr.de/sensors/iasi/

[72] 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.

[73] L. Santos, L. Berrojo, J. Moreno, J. Lopez, and R. Sarmiento, "Multispectral and hyperspectral lossless compressor for space applications (HYLOC): A low-complexity FPGA implementation of the CCSDS 123 standard," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 757–770, Feb. 2016.
[74] N. Amrani and J. Serra-Sagristà, *RWA Coding Toolbox*, 2015. [Online]. Available: http://gici.uab.es/GiciWebPage/downloads.php

Naoufal Amrani received the B.S. degree in mathematics from the Universitat de Barcelona, Barcelona, Spain in 2011 and the M.S. degree in computer vision and the M.S. degree in high performance computer from the Universitat Autonoma de Barcelona, Barcelona, Spain, in 2012 and 2013, respectively.

Since 2013, he has been with the Group on Interactive Coding of Images, Universitat Autonoma de Barcelona, where he is currently a Ph.D. student.

Mr. Amrani was the recipient of the Capocelli Prize for the 2016 Data Compression Conference.

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

He is currently an Associate Professor with the Department of Information and Communications Engineering, UAB. From September 1997 to December 1998, he was at the University of Bonn, Bonn, Germany, funded by DAAD. His current research interests focus on data compression, with special attention to image coding for remote sensing and telemedicine applications. He has coauthored over 100 publications.

Dr. Serra-Sagristà serves as Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was the recipient of the Spanish Intensification Young Investigator Award in 2006.

Valero Laparra was born in València, Spain, in 1983. He received the B.Sc. degree in telecommunications engineering, Universitat de València in 2005, the B.Sc. degree in electronics engineering, Universitat de València in 2007, the B.Sc. degree in mathematics, Universidad Nacional de Educación a Distancia in 2010, and the Ph.D. degree in computer science and mathematics, Universitat de València in 2011.

He is a Postdoctoral Researcher in the Image Processing Laboratory at Universitat de València, València.

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, California, 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. His research interests include digital communication and data storage systems, data compression, and signal processing. He has authored or coauthored more than 200 publications in these areas. He has received numerous honors, including six teaching awards.

Jesus Malo received the M.Sc. and Ph.D. degrees in physics from the Universitat de València, València, Spain, in 1995 and 1999, respectively.

In 2000 and 2001, he worked as Fulbright Postdoctoral Researcher at the Vision Grofup of the NASA Ames Research Center (A.B. Watson) and at the Laboratory of Computational Vision of the Center for Neural Science, New York University (E.P. Simoncelli). He came back to NYU in 2013 for a semester. He is with the Image and Signal Processing Group and the Visual Statistics Group (VI(S)TA) at the Universitat de València.

Dr. Malo is a member of the Asociación de Mujeres Investigadoras y Tecnólogas. He currently serves as Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was the recipient of the Vistakon European Research Award in 1994.