Regression Wavelet Analysis for Progressive-Lossy-to-Lossless Coding of Remote-Sensing Data

Naoufal Amrani∗, Joan Serra-Sagristà∗, Miguel Hernández-Cabronero‡ and Michael Marcellin†

∗Dep. of Information and Communications Eng., Universitat Autònoma de Barcelona, Spain
‡Dep. of Computer Science, University of Warwick, UK
†Dep. of Electrical and Computer Engineering, University of Arizona, USA

Abstract

Regression Wavelet Analysis (RWA) is a novel wavelet-based scheme for coding hyperspectral images that employs multiple regression analysis to exploit the relationships among spectral wavelet-transformed components. The scheme is based on a pyramidal prediction, using different regression models, to increase the statistical independence in the wavelet domain. For lossless coding, RWA has proven to be superior to other spectral transforms like PCA and to the best and most recent coding standard in remote sensing, CCSDS-123.0. In this paper we show that RWA also allows progressive lossy-to-lossless (PLL) coding and that it attains a rate-distortion performance superior to those obtained with state-of-the-art schemes. To take into account the predictive significance of the spectral components, we propose a Prediction Weighting scheme for JPEG2000 that captures the contribution of each transformed component to the prediction process.

1 Introduction

Remote-sensing data has become enormously important for a wide range of applications in many different fields of Earth’s observation thanks to the technological evolution of optical sensors over the last years. As recent sensors can cover large areas of the Earth’s surface, producing images with unprecedented spectral and spatial resolution, the need for efficient coding techniques to improve the capabilities of storage and transmissions emerges naturally. While there are some scenarios where only lossless coding is acceptable, there are other applications where lossy coding is indispensable because only a limited amount of data can be transmitted. Techniques allowing progressive lossy-to-lossless coding are thus gaining importance in remote-sensing as they provide similar rate-distortion performance to pure lossy methods at low bitrates, and also a competitive coding performance in the lossless regime.

Most efficient coding schemes for remote-sensing data include decorrelating transforms to exploit the spatial correlation between adjacent pixels within a component and the spectral correlation between adjacent components. Some schemes apply 3D transforms [1, 2], commonly based on wavelets, to handle the joint correlation in both spatial and spectral dimensions. Other somewhat different schemes apply separately a 1D spectral transform followed by a 2D spatial transform, using the most convenient one in each dimension. Generally, the correlation removal along the spectral dimension is considered critical in coding hyperspectral images due to its significant impact on the coding performance [3].

In particular, Principal Component Analysis (PCA) is the optimal decorrelating transform for Gaussian sources [4, 5]. It is widely applied for multicomponent images as a
spectral transform since it also yields an excellent coding performance. However, PCA is a
data-dependent transform, entailing the need to compute it for each individual image before
its application. As a consequence, its computational complexity is substantial due to the
covariance matrix calculation, the extraction of eigenvectors, the matrix factorization and
the integer mapping (integer implementation is needed for lossless coding [6, 7]). Hence,
in scenarios where the data has a significant size in the spectral dimension, such as IASI
images—with more than 8000 components—, PCA and its integer implementation are not
feasible. Additionally, PCA is not a component-scalable transform, i.e., the recovery of
any single image component depends on every transformed component.

Since the complexity reduction of PCA is demanded, a number of approaches have been
proposed, including (i) approaches based on spatially and/or spectrally sub-sampling the
dataset and estimating the covariance matrix using a reduced subset of coefficients [3, 8];
(ii) approaches based on divide-and-conquer strategies to approximate the PCA with a
reduced computational complexity and some amount of component scalability [9, 10]; and
(iii) approaches that consist in learning the transform on a set of images of one particular
sensor and apply the learned transform on new images from the same sensor [11]. The
the techniques of family (i) only alleviate the complexity of computing the covariance
matrix while other complexity sources still remain, such as the integer mapping for lossless
coding. On the other hand, the techniques of families (ii) and (iii) might produce significant
deterioration in the coding performance compared to the full-complexity PCA.

Beyond computational issues, another limitation of classical PCA is its linear nature.
Even though linear PCA yields competitive performance in a wide range of applications
because of its invertibility and energy compaction, its efficiency is hampered by data ex-
hibiting non-linear relationships. Thus, despite the simplicity and efficiency that linearity
provides, non-linear methods could provide performance improvements by better exploit-
ing data dependencies. Recently, several non-linear generalization of PCA have been pro-
posed to deal with data of non-linear nature [4, 12, 13]. In the adaptation of these non-linear
methods to image coding, there are two major considerations: first, the transform must be
invertible; and second, the computational complexity and the memory consumption should
be reasonable. These two requirements eliminate a large number of candidate approaches.

In [14] we explored lossless hyperspectral image coding using curvilinear techniques
based on Principal Polynomial Analysis (PPA) [13]. In that work, it was shown that PPA
achieves higher energy compaction and statistical independence than PCA. However, in
practical hyperspectral image coding, appropriate handling of side information and the se-
quential error introduced by the integer mapping dramatically penalize the coding gain.
Additionally, the computational complexity and memory requirements of these non-linear
transforms were even larger than those of the original PCA, which is already computa-
tionally expensive [4].

Alternatively, rather than using sophisticated non-linear methods for remote-sensing
coding, we focus here on theoretically suboptimal but simpler traditional transforms (such
as the Discrete Wavelet Transform) while adopting predictive techniques [15, 16] to exploit
any remaining post transform statistical dependence.

Following this line of work, in [17] we introduced the Regression Wavelet Analysis
(RWA) transform for lossless coding. 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 compared to that of the best spectral decorrelation techniques like PCA, and also compared to the most recent and most competitive prediction-based hyperspectral coding technique, CCSDS-123.0 standard [18].

In this paper, we extend the benefits of RWA to progressive lossy-to-lossless coding of remote-sensing data by considering the predictive significance of the transformed components. To address this issue, we propose a weighting scheme that assigns a weight to each spectral component depending on its contribution to the prediction process. Weighting strategies are useful to improve the quality of the recovered data by allocating more bits to the significant coefficients for an algorithm design [19]. Unlike pure lossless coding, where it is not necessary to consider these weights to achieve the aforementioned performance when using RWA, in progressive lossy-to-lossless the loss of information at intermediate or low bitrates may have an error propagation effect for coding schemes based on pyramidal prediction like RWA.

This paper is organized as follows. Section 2 introduces briefly the RWA algorithm and the regression model. Section 3 puts forward the proposed Prediction Weighting scheme. Section 4 reports the experimental results. Finally, section 5 concludes.

2 Regression Wavelet Analysis

This section begins with a review of the DWT, providing the necessary notation. It then goes on to describe the Regression Wavelet Analysis algorithm and the regression model.

2.1 Discrete Wavelet Transform

Let us suppose that an original multicomponent image \( V^0 \) has \( z=2^d \) spectral components with each component having \( m \) spatial samples. Then we write \( V^0 \in \mathbb{R}^{m \times z} \) and

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

The wavelet representation of \( V^0 \) with \( J \) levels, for \( 1 \leq J \leq \log_2(z) \), is given by

\[
\text{DWT}(V^0, J) = (V^J, (W^j)_{1 \leq j \leq J}) ,
\]

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

\[
\text{DWT}(V^{j-1}, 1) = (V^j, W^j)
\]

At each scale or level \( j \) the signal \( V^{j-1} \in \mathbb{R}^{m \times (z \cdot 2^{-j+1})} \) is decomposed into the approximation signal \( V^j \in \mathbb{R}^{m \times (z \cdot 2^{-j})} \) and the details signal \( W^j \in \mathbb{R}^{m \times (z \cdot 2^{-j})} \) at half resolution each. The approximation signal \( V^j \) usually contains most of the information of the previous signal \( V^{j-1} \), whereas the details signal \( W^j \) contains the information difference between \( V^{j-1} \) and \( V^j \). The decomposition is usually repeated in cascade form on the approximation signal \( V^j \) as shown in Fig. 1.
2.2 Regression Wavelet Analysis

The Regression Wavelet Analysis (RWA) scheme generalizes the discrete wavelet transform (DWT) by applying regression to tackle the correlation that still remains in the DWT domain. At each scale $j$, each details component $W^j(i) = W^j_i \in \mathbb{R}^{m \times 1}$ is predicted from the information contained in the approximation components $V^j \in \mathbb{R}^{m \times (2^j - 1)}$ within the same scale $j$. This prediction $\hat{W}_i^j = f_i(V^j)$ is then removed to obtain a residual

$$R^j = W^j - \hat{W}^j.$$  \hspace{1cm} (3)

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

$$\text{RWA}(V^{j-1}, 1) = (V^j, R^j)$$
$$\text{RWA}(V^0, J) = (V^J, (R^j)_{1 \leq j \leq J}).$$  \hspace{1cm} (4), (5)

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

$$W^j = R^j + \hat{W}^j$$
$$V^{j-1} = \text{DWT}^{-1}((V^j, W^j), 1).$$  \hspace{1cm} (6), (7)
In general, RWA can be based on any wavelet filter, nevertheless, as we focus on progressive lossy-to-lossless (PLL) coding, and for faster implementation, hereafter, we will refer to RWA as the one based on integer Haar filter (S-transform) [20], which produces integer coefficients \((V_j, W_j)\) with a quantized estimation removal \(R_j = W_j - Q(\hat{W}_j)\), where the quantization is applied through a simple rounding operation.

2.3 Regression Model

Different prediction models have been proposed for RWA. The most general model is called the Maximum model. It employs all of the approximation components from \(V^j\) for the computation of the prediction of each details component \(W^j_i\). It is given by

\[
\hat{W}_i^j = f_i[V^j] = \beta_{i,0}^j + \beta_{i,1}^j V_1^j + \cdots + \beta_{i,k}^j V_k^j, \quad V_i^j \in \mathbb{R}^{m \times 1}.
\]  

(8)

For the Maximum model, the regression coefficients \(\beta^j\) are found for each individual image using the least-squares method [21] that minimizes the sum of squares of the distances between the original components and the estimated ones, i.e.,

\[
\min: \|W_i^j - \bar{W}_i^j\|_2.
\]

Since Maximum model has a non-negligible computational complexity, to alleviate its cost we resort to the application of an approach in line with those of family (iii) (see Sect. 1). The proposed Exogenous model computes the regression coefficients \(\overline{\beta}^j\) 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 for Exogenous model 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

\[
\hat{W}_i^j = f_i[V^j] = \overline{\beta}_{i,0}^j + \overline{\beta}_{i,1}^j V_1^j + \cdots + \overline{\beta}_{i,k}^j V_k^j, \quad V_i^j \in \mathbb{R}^{m \times 1}.
\]  

(9)

3 Prediction Weighting Scheme

According to the RWA design, the residuals \(R^j\) not only contribute to reconstruct the details components \(W^j\) (Eq. 6), but are also used to generate the estimates \(\hat{W}^{j-1}\) in the next inverse steps. As a result, the distortion introduced to a given residual component \(R_i^j\) at scale \(j\) will affect the reconstruction in all finer scales \(j - 1, \ldots, 1\).

The Prediction Weighting scheme that we propose aims at emphasizing the contribution of each spectral residual to the prediction process by assigning weights that capture the predictive significance. These weights will be used as input values to JPEG2000 to control the relative significance of the spectral components [19].

Consider the RWA representation \((V^j, (R^j)^{1 \leq j \leq J})\) at scale \(J\). To invert the transform, we need to generate the estimates \((\hat{W}^j)^{1 \leq j \leq J}\) used in (Eq. 6) and (Eq. 7). Let us analyze how the residuals \(R^j\) of each level \(j\) contribute to the generation of these estimates. Given
Then the relative proportion with respect to the overall number of components \( z \)
with all the residuals of the coarser scales of predictions in which it is involved, i.e.,
the number of components of \( \hat{W}^j \) is involved in the prediction of all the estimates \( \hat{W}^j \)
\( j > j > 1 \) is involved in the prediction of all the estimates \( \hat{W}^j \) in the finer scales
\( j = 1, \ldots, 1 \). Additionally, the estimates \( \hat{W}^j \in \mathbb{R}^{m \times (z-2^{-k})} \) in each scale \( k \) are formed by
\( z \cdot 2^{-k} \) components (Eq. 2).

Given that, let us consider a function \( N(\cdot) \) that relates the residuals \( R_j \) with the number
of predictions in which it is involved, i.e., the number of components of \( \hat{W}^j \in \mathbb{R}^{m \times (z-2^{-k})} \)
with \( 1 \leq k \leq j - 1 \)
\[
N(R_j) = \begin{cases} 
  z \cdot \sum_{i=1}^{j-1} (2^{-i}) & \text{if } j > 1 \\
  0 & \text{if } j = 1 
\end{cases} 
\]  \hspace{1cm} (14)

Then the relative proportion with respect to the overall number of components \( z \) will be
\[
P N(R_j) = \frac{N(R_j)}{z} = \begin{cases} 
  \sum_{i=1}^{j-1} 2^{-i} & \text{if } j > 1 \\
  0 & \text{if } j = 1 
\end{cases} 
\]  \hspace{1cm} (15)

We now define the weights for \( R_j \) as
\[
W(R_j) = \frac{1}{P N(R_j) - P N(R^{j-1})} = 2^{j-1}. 
\]  \hspace{1cm} (16)

Finally, the approximation \( \mathcal{V}^j \) of RWA (in the case of applying the maximum possible
number of levels \( J = \log_2(z) \)) consists of only one component \( \mathcal{V}^j \in \mathbb{R}^{m \times 1} \), which is
weighted by
\[
W(\mathcal{V}^j) = 2^j. 
\]  \hspace{1cm} (17)

In this way, \( W(\cdot) \) weights the transformed components \( (\mathcal{V}^j, (R_j)^{1 \leq j \leq J}) \) by different
quantities, taking into account the predictive significance of each component and the components
of the previous scale. These prediction weights play a role similar to that of the subband weighting
factors in spatial or temporal wavelet transforms [19].
4 Experimental results

In this section we evaluate the RWA transform with the proposed Prediction Weighting scheme (Weighted RWA) on a set of hyperspectral images from two different sensors: AVIRIS [22] and IASI [23]. For comparison purposes, we provide results for other reversible spectral transforms including RWA without weighting, PCA and DWT 5/3.

The coding system pipeline is shown in Fig. 3 and illustrates the compression process. The compared spectral transforms are first applied in the spectral dimension followed by a 2D JPEG2000 compression, which may include a reversible 2D DWT 5/3 transform with 5 levels in the spatial dimension and a bitplane-based entropy coding. The spatial transform is not applied for PCA and RWA. In these cases, the spatial transform provides no improvement, as most of the transformed components already have low energy. The Kakadu software implementation of JPEG2000 has been used. For Weighted RWA, the weights were introduced using the option Cweight.

![Image](image.png)

Figure 3: 1D spectral transform followed by JPEG2000 standard.

![Figure 4: PLL Rate-Distortion for AVIRIS (radiance) with 224 spectral components, 512x677 spatial resolution and 16 bpppc bitdepth.](figure4.png)

Figures 4, 5 and 6 provide the Rate-Distortion performance for the compared spectral transforms. These figures provide the relation between the target bitrate, measured in bits per pixel per component (bpppc) (including all required side information), and the reconstruction quality, reported as Peak Signal-to-Noise Ratio (PSNR), measured in dB.

As a general trend, Weighted RWA performs significantly better than RWA without weighting for progressive lossy-to-lossless coding. Additionally, RWA without weighting
is heavily affected by the error propagation at all bitrates (low, intermediate and high bitrates), which produces multiple plateaus in the Rate-Distortion performance. Weighted RWA is also superior to the DWT 5/3 spectral transform, which has a performance similar to that of the RWA without weighting. In comparison to PCA, the performance of Weighted RWA for 16 bpppc AVIRIS calibrated (resp. uncalibrated) images is somewhat worse up to a bitrate of 0.7 (resp. 2.8) bpppc; from then onwards, PCA falls clearly behind, with a difference of over 8 dB at 3 (resp. 5) bpppc. For 12 bpppc AVIRIS uncalibrated images, Weighted RWA is the most competitive spectral transform.

With regard to IASI images, as they have a large size in the spectral dimension (more than 8000 components), reversible PCA results in a prohibitive computational cost due to the recursiveness of the integer mapping [24]. For this reason, coding results for PCA on IASI images are not reported. The results of Weighted RWA and classical RWA (without weighting) are provided using the Exogenous model due to the non-negligible role of the side information (coefficients $\beta$). It is worth noting that the Exogenous coefficients $\bar{\beta}_{i,k}^j$
were computed using only one training image from IASI Level 0 and another one from IASI Level 1. The analysis of the coding performance of the Exogenous model for PLL coding is similar to that reported before: Weighted RWA is clearly superior to RWA without weighting and to DWT 5/3.

5 Conclusions

We proposed a Prediction Weighting scheme for progressive lossy-to-lossless coding of hyperspectral images with Regression Wavelet Analysis algorithm. The suggested weights are able to capture the relative significance of the transformed spectral components in the prediction process, minimizing the effect of error propagation and providing a steady quality evolution. At a much lower computational complexity than reversible PCA, our proposal yields the most competitive rate-distortion coding performance published to date.

Acknowledgements

This work has been partially supported by the Spanish Government (MINECO), by FEDER, by the Catalan Government and by Universitat Autònoma de Barcelona, under Grants TIN2015- 71126-R, TIN2012-38102-C03-03, 2014SGR-691, and UAB-PIF-472-03-1/2012.

References


