We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

10

Discrete Wavelet Transform and Optimal Spectral Transform Applied to Multicomponent Image Coding

Isidore Paul Akam Bita[1], Michel Barret[2], Florio Dalla Vedova[1], Jean-Louis Gutzwiller[2] and, Dinh-Tuan Pham[3]
[1]*LUXSPACE Sarl, Chateau de Betzdorf, Betzdorf*
[2]*SUPELEC, Information Multimodality and Signal Team, 2, rue E. Belin, Metz*
[3]*Jean Kuntzmann Laboratory, 51 rue des Mathématiques, Grenoble Cedex 9*
[1]*Luxembourg*
[2,3]*France*

1. Introduction

These last years, research activities on multicomponent image compression have been expanded, due to the development of multispectral and hyperspectral image sensors which supply larger and larger amount of data. The end-users of such images become also more numerous and have various needs and various applications. The future earth observation systems, for instance, will use multi-, super- and hyper- spectral image sensors with higher resolutions leading to bigger amount of transmitted data. However the channel bandwidth for transmission is limited and therefore there is an interest of conceiving compression systems (onboard and on the ground) of multicomponent images which are not application dependent and which are compatible with the diversity of end-users' needs. The components of a multicomponent image generally represent the same scene with different views depending on the wavelength. For data from different sensors, a preliminary step of image registration is therefore required as there is a high degree of dependence (or redundancies) between the various components: the usual spatial redundancy (between different pixels in each component) and the spectral redundancy (between the components).
During the past two decades, different solutions have been proposed for multicomponent image coding. A solution currently adopted consists of using two different transformations, each with the goal of reducing only one of the two redundancies. In (Dragotti et al., 2000), a 2-D discrete wavelet transform (DWT) is used to reduce the spatial redundancies in each component while the Karhunen Loève transform (KLT) is applied to reduce the spectral ones. In that paper, the quantization and entropy coding are achieved thanks to the well known SPIHT (Set Partitioning in Hierarchical Trees) codec by Said and Pearlman (Said & Pearlman, 1996) in its original version and in a modified version including VQ (vector quantization). In the same way, with the use of the 2-D DWT of (Antonini et al., 1992) (usually called the Daubechies 9/7), the authors of (Vaisey et al., 1998) use a lattice VQ with a stack run coder as quantization and entropy coding. More recently in (Rucker et al., 2005), the KLT associated with the Daubechies 9/7 2-D DWT and with EBCOT (Taubman, 2000; Taubman & Marcellin, 2002) for quantizing and entropy coding has been tested on

hyperspectral images with different bit-allocations between components. It is shown that the Post Compression Rate-Distortion (PCRD) optimizer of EBCOT applied across multiple bands gives the best rate-distortion performance. Another solution consists of using a 3-D DWT for reducing both the spatial and spectral redundancies with only one transform. This approach is generally applied to hyperspectral images as in (Christophe et al., 2006). An overview of 3-D wavelet-based techniques and more can be found in (Fowler & Rucker, 2007). The two above mentioned solutions are compatible with the JPEG2000 Part 2 standard. The JPEG2000 standard is well known and well spread today. Moreover the KLT used in JPEG2000 Part 2 is considered as the best existing lossy compression techniques for hyperspectral images at medium and high bit rates (Du & Fowler, 2007; Penna et al., 2007). The KLT consists in a Principal Component Analysis (PCA), well known of statisticians, where all the components are kept. However, the rather great computational complexity of the KLT hinders its adoption in practice — specially on satellite platforms — and recent works propose different solutions in order to pass round this problem. One approach consists in reducing the complexity of the covariance matrix computation. This is done by randomly sampling the entire image in order to obtain a small sample of the pixels' population on which the covariance matrix is computed (Du & Fowler, 2008; Penna et al., 2007). Another approach consists in computing a kind of KLT average on a set of images (the learning basis) issued from only one sensor and using it on other images obtained with the same sensor. This sub-optimal transform is called exogenous KLT in (Thiebaut et al., 2006) and the computational complexity of the second approach is compatible with satellite platforms. Both approaches are fruitful: the rate-distortion performance sacrifice compared with the true KLT is very slight, whereas the computational burden is significantly reduced. In the second approach, the exogenous KLT matrix is known by the decoder, hence there is no need to transmit it.

It is well known that the KLT can be suboptimal in transform coding when the data are not Gaussian. Now, under only the high resolution quantization hypothesis, nearly everything is known about the performance of a transform coding. Nevertheless, the optimal transform computation is generally considered as a difficult task and the Gaussian assumption is then used in order to simplify the calculation. Recently, the problem of computing the optimal coding transform associated with scalar variable-rate quantizers for still images was resolved under high-resolution quantization hypothesis, with mean square error as distortion and without the Gaussian assumption (Narozny et al., 2005; 2008). However, for the JPEG2000 Part2 compression scheme, the previous optimal transform computation cannot be directly applied to obtain the optimal spectral transform, because of the 2D DWT presence—see the criterion (15) in Section 4, which depends on subband statistics—. In (Akam Bita et al., 2010a), the authors solved both the problems of computing an optimal spectral transform (OST), with the constraint of orthogonality and without any constraint but invertibility, for that compression scheme, when the 2D DWT has fixed coefficients and under the only high resolution quantization hypothesis. They showed that on hyperspectral images, the orthogonal OST, called OrthOST, performs slightly but significantly better than a KLT at low, medium and high bit-rates and that the gain obtained by removing the orthogonality constraint in the computation of the OST is not significant. Further, it is not widely known that even when the input data are Gaussian, the KLT is not optimal in the above mentioned compression scheme. Indeed, after the 2D DWT, the variance of the wavelet coefficients depends on the subband they belong to (even for Gaussian data) and the KLT does not capture these various variances, while the EBCOT coder with its PCRD optimizer performing simultaneously across all the codeblocks from the entire image take them into

account. In (Akam Bita et al., 2010b), the authors introduced an orthogonal spectral transform (called JADO for Joint Approximate Diagonalization under Orthogonality constraint) using only second order statistics that has not this shortcoming, and that is optimal at high bit-rates for the JPEG2000 Part 2 compression scheme, when the data are Gaussian. They showed on natural hyperspectral images that JADO (resp. OrthOST) performs slightly but significantly better than the KLT (resp. JADO). The main drawback of the OSTs is their heavy computational cost, which is much higher than the one of a KLT or JADO (which both have roughly the same complexity).

In order to reduce the complexity of a codec based on OrthOSTs, the authors of (Akam Bita et al., 2008; 2010c; Barret et al., 2009) used the same strategy as in (Thiebaut et al., 2006): they replaced the OrthOST, which must be computed for each new encoded image, with an *exogenous* quasi optimal spectral transform. This last transform is an OrthOST computed once and for all on a learning basis constituted of images from only one spectrometer and which is then applied to any image to be coded stemming from the same spectrometer. Using either the JPEG2000 codec called Verification Model version 9 (JPEG2000, 2001) or the Bit Plane Encoder (BPE (CCSDS-1, 2007)) recommended for satellite image compression (Yeh et al., 2005) by the CCSDS (Consultative Committee for Space Data Systems), they showed that this strategy yielded good performances, sometimes better than the (non exogenous) KLT ones, in terms of bit-rate versus distortions. Four different distortions were considered: Signal to Noise Ratio (SNR), Maximum Absolute Difference (MAD), Mean Absolute Error (MAE) and Maximum Spectral Angle (MSA). Indeed, it is well-known that providing the mean square error as one distortion only is not sufficient to assess the quality of a codec for hyperspectral images (Christophe et al., 2005). However in the simulations presented in (Akam Bita et al., 2008; 2010c; Barret et al., 2009) when the VM9 is used, the computational complexity of the EBCOT coder associated with its PCRD optimizer is very high, and when the BPE is applied to encode each component of the transformed image, the complexity of the algorithm for optimal allocation between components is also very high. In both cases, the computational complexity is too high for a compression system on-board a satellite. In (Barret et al., 2011), the authors present a low complexity hyperspectral image coder based on exogenous OrthOST and zerotrees well adapted to OrthOST.

It is important to note that the point of view presented in this chapter — i.e., a compression scheme for hyperspectral images that is independent of the end-user application — is no longer justified at very low bit-rates (lower than 0.5 bits per pixel and per band). For more details on low-bit rates hyperspectral compression see (Chang et al., 2010c).

In this chapter, we study the question of an optimal linear transform for reducing spectral redundancies under high resolution and variable rate constrained quantization hypothesis, when a 2-D DWT — with fixed coefficients — is applied to each component to reduce spatial redundancies and one scalar quantizer per subband and per component is used. This compression scheme, described in Section 2, is compatible with the JPEG2000 Part 2 standard. The asymptotic expression of the mean square error distortion associated with that compression scheme is given in Section 3. In Section 4, we clarify the criterion minimized by such an optimal spectral transform with mean square error distortion and we show the link between the criterion and the mutual information contrast used in Independent Component Analysis (ICA). In Section 5, we derive a criterion minimized by an OrthOST under Gaussian data assumption. Moreover, we describe in Section 6 the quasi-Newton algorithms used for the minimization of the criterion, either with the constraint of an orthogonal transform or with no constraint but invertibility or with the constraint of an orthogonal transform and the

assumption of Gaussian data. The two first algorithms are derived from an algorithm by Pham `ICAinf` described in (Pham, 2004) that performs ICA. Then in Section 7, performances of these transforms and comparisons with the KLT are given for multi- and hyper-spectral satellite images, with the four above mentioned different measures of distortion. Finally, in Section 8 we introduce quasi-optimal OrthOSTs, called *exogenous*, that have not the main drawback of heavy computational cost and we compare their performances in lossy coding with OrthOSTs.

2. Description of the separable compression scheme

2.1 Conventions and notations

We consider a multicomponent image $\mathbf{X}$ with N components $\mathbf{X}_1, \ldots, \mathbf{X}_N$. Each component $\mathbf{X}_i$ is a 2-D image with N_r rows and N_c columns. To simplify the notations and the mathematical expressions, we assume that each component is written as a row vector by scanning all its pixels row by row (for example). Then $\mathbf{X}$ is a $N \times L$ matrix, with $L = N_r N_c$. In the following, depending on the context, we shall interpret $\mathbf{X}_i$ as a 2-D image or as a row vector of dimension L. For a square matrix $\mathbf{M}$, the expressions $\det \mathbf{M}$, $\operatorname{tr} \mathbf{M}$ and $\operatorname{diag}(\mathbf{M})$ denote respectively its determinant, its trace and the diagonal matrix obtained with its diagonal elements.
In the following compression scheme, the 2-D DWT has fixed coefficients (in our tests, the Daubechies 9/7 DWT is always used), but the spectral linear transform is adapted to the data. We denote $\mathbf{W}$ the invertible $L \times L$ matrix associated with the 2-D DWT.

2.2 The separable scheme

The separable scheme is compatible with the JPEG2000 Part 2 standard. It can be described as follow:

- *Coding*. The same 2-D DWT is applied to each component $\mathbf{X}_i$ in order to reduce the spatial redundancies and a linear transform $\mathbf{A}$ is applied between the components in order to reduce the spectral redundancies. The result of the 2-D DWT applied to the entire image $\mathbf{X}$ is $\mathbf{X}\mathbf{W}^T$ and the transformed coefficients are the elements of the matrix $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{W}^T$. For each component, the wavelet coefficients of each subband are regrouped according to a fixed scan that does not depend on the component. This re-ordering corresponds to the right multiplication of $\mathbf{X}\mathbf{W}^T$ by a permutation matrix $\mathbf{P}^T$. We can suppose without loss of generality that $\mathbf{P}$ is the identity, otherwise we could replace $\mathbf{W}$ with $\mathbf{P}\mathbf{W}$. This partitioning can be written $\mathbf{X}\mathbf{W}^T = [(\mathbf{X}\mathbf{W}^T)^{(1)} \ldots (\mathbf{X}\mathbf{W}^T)^{(M)}]$, where M is the number of subbands. Then, the transformed coefficients $\mathbf{Y} = [\mathbf{Y}^{(1)} \ldots \mathbf{Y}^{(M)}]$ (where $\mathbf{Y}^{(i)} = \mathbf{A}(\mathbf{X}\mathbf{W}^T)^{(i)}$) are quantized and entropy coded with one quantizer per subband and per component (see § 7.1).
- *Decoding*. Let $\mathbf{Y}^q$ denote the matrix with the same dimension as $\mathbf{Y}$ containing the dequantized transformed coefficients. The mathematical inverse transforms are applied to $\mathbf{Y}^q$ in order to reconstruct an approximation $\widehat{\mathbf{X}} = \mathbf{A}^{-1}\mathbf{Y}^q\mathbf{W}^{-T}$ of the original image $\mathbf{X}$.

We can remark that the order of the transformations (i.e., applying first the DWT then $\mathbf{A}$, or first $\mathbf{A}$ then the DWT) has no effect on the result, since $\mathbf{Y} = \mathbf{A}(\mathbf{X}\mathbf{W}^T) = (\mathbf{A}\mathbf{X})\mathbf{W}^T$. This is why that scheme is called *separable*.

3. Expression of the distortion

In lossy or quasi-lossless coding, the quantization leads to irreversible loss of information, therefore the decoded image $\widehat{\mathbf{X}}$ is an approximation of the original image and in order to

quantify the quality of the reconstructed image, it is necessary to introduce a measure of distortion. In this section we give, under various hypotheses, the relation that links the distortion between $\mathbf{X}$ and $\widehat{\mathbf{X}}$ to the quantizers distortions, when the distortion is the mean square error:

$$D_a(\mathbf{X},\widehat{\mathbf{X}}) = \frac{1}{NL}\|\mathbf{X}-\widehat{\mathbf{X}}\|^2 \quad \text{with} \quad \|\mathbf{X}-\widehat{\mathbf{X}}\|^2 = \sum_{i=1}^{N}\sum_{k=1}^{L}(X_i(k)-\widehat{X}_i(k))^2. \tag{1}$$

We begin by recalling the solution of the problem in a simple general case (Gersho & Gray, 1992; Taubman & Marcellin, 2002).

3.1 A simple general case

Lemma 3.1. *Let $\mathbf{X}$ be a real random vector with N components and $\mathcal{A}$ be an invertible matrix of order N. The transformed vector $\mathbf{Y} = \mathcal{A}\mathbf{X}$ is quantized and dequantized in $\mathbf{Y}^q$. The original vector $\mathbf{X}$ is approximated by $\widehat{\mathbf{X}} = \mathcal{A}^{-1}\mathbf{Y}^q$ and let $\mathbf{b} = \mathbf{Y} - \mathbf{Y}^q$ be the quantization noise. Then, the end-to-end distortion $D = \frac{1}{N}\,\mathrm{E}(\|\mathbf{X}-\widehat{\mathbf{X}}\|^2)$, where E denotes the mathematical expectation, satisfies the relation $D = \frac{1}{N}\,\mathrm{tr}\left[\mathrm{E}(\mathbf{b}\mathbf{b}^T)\mathcal{A}^{-T}\mathcal{A}^{-1}\right]$.*

Proof: We have $\mathbf{X}-\widehat{\mathbf{X}} = \mathcal{A}^{-1}\mathbf{b}$ and $\|\mathcal{A}^{-1}\mathbf{b}\|^2 = \mathbf{b}^T\mathcal{A}^{-T}\mathcal{A}^{-1}\mathbf{b} = \mathrm{tr}[\mathcal{A}^{-1}\mathbf{b}\mathbf{b}^T\mathcal{A}^{-T}] = \mathrm{tr}[\mathbf{b}\mathbf{b}^T\mathcal{A}^{-T}\mathcal{A}^{-1}]$, therefore $D = \frac{1}{N}\,\mathrm{E}(\|\mathcal{A}^{-1}\mathbf{b}\|^2) = \frac{1}{N}\,\mathrm{tr}\left[\mathrm{E}(\mathbf{b}\mathbf{b}^T)\mathcal{A}^{-T}\mathcal{A}^{-1}\right]$. ■

Further, we may need the following assumption, that can be deduced from high resolution quantization hypothesis (Gersho & Gray, 1992) (this point is recalled in Subsection 3.2).

$\mathcal{H}_1$: *The components of the quantization noise are zero mean and uncorrelated.*

Theorem 1. *1. With the hypotheses of Lemma 3.1 and assuming $\mathcal{H}_1$, the distortion becomes*

$$D = \frac{1}{N}\sum_{i=1}^{N}\alpha_i D_i, \tag{2}$$

where $D_i = \mathrm{E}(b_i^2)$ is the quantizer distortion of the i^{th} component $\mathbf{Y}_i$ of $\mathbf{Y}$ and, with $\mathbf{e}_i$ the i^{th} canonical vector of $\mathbb{R}^N$ and $(\mathcal{A}^{-1})_{ij}$ the element of $\mathcal{A}^{-1}$ located on row i and column j, we have

$$\alpha_i = \sum_{j=1}^{N}(\mathcal{A}^{-1})_{ji}^2 = \|\mathcal{A}^{-1}\mathbf{e}_i\|^2. \tag{3}$$

2. The assertion 1. holds without the assumption $\mathcal{H}_1$ if $\mathcal{A}^{-T}\mathcal{A}^{-1}$ is diagonal, e.g. if $\mathcal{A}$ is orthogonal.

Proof: The assumptions in 1. or 2. state that at least one of the two matrices $\mathcal{A}^{-T}\mathcal{A}^{-1}$ and $\mathrm{E}(\mathbf{b}\mathbf{b}^T)$ is diagonal. Hence the trace of their product is equal to the sum of the products of their diagonal elements. ■

3.2 Justification of the assumption $\mathcal{H}_1$

We recall here well known results that can be found e.g. in (Gersho & Gray, 1992). The assumption $\mathcal{H}_1$ can be justified under the following conditions. C_1: the random vector $\mathbf{Y} = (Y_1,\ldots,Y_N)^T$ has a continuous probability density function (pdf) $f_{\mathbf{Y}}$; C_2: separable high-rate quantization is achieved, meaning that the quantization steps $\mathbf{h} = (h_i)_{1\le i\le N}$ of the N components are small with respect to the variations of $f_{\mathbf{Y}}$ (i.e. $f_{\mathbf{Y}}(\mathbf{y}+\mathbf{h}) \simeq f_{\mathbf{Y}}(\mathbf{y})$, $\forall \mathbf{y} \in \mathbb{R}^N$) and C_3 : for any cell S of the separable N-D quantizer, the dequantized value $\mathbf{Y}^q$

associated with S is the iso-barycenter of S. Indeed, if the three conditions C_1, C_2 and C_3 hold, then the pdf $f_{\mathbf{Y}}$ can be considered as quasi constant in the hypercube $\mathbf{Y}^q + \prod_{i=1}^{N}[-h_i/2\,,\, h_i/2]$. Further, the conditional law of the quantization noise $\mathbf{b} = \mathbf{Y} - \mathbf{Y}^q$ knowing the dequantized value $\mathbf{Y}^q$ satisfies $f_{\mathbf{b}|\mathbf{Y}^q}(\mathbf{u}) \simeq 1/\prod_{i=1}^{N} h_i$ if $\mathbf{u} \in \prod_{i=1}^{N}[-h_i/2\,,\, h_i/2]$, 0 otherwise. We see that the conditional pdf $f_{\mathbf{b}|\mathbf{Y}^q}$ does not depend on the quantized value $\mathbf{Y}^q$, hence it is equal to $f_{\mathbf{b}}$, the pdf of $\mathbf{b}$. Further the components of $\mathbf{b}$ are zero mean and (quasi) independent since their joint density is approximatively equal to the product of their marginal densities.

3.3 The separable subband scheme

In the following, the symbols $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Y}^q$ refer again to the matrices defined in Section 2 and $\mathcal{A}$ denotes the matrix of the linear transform that associates $\mathbf{Y}$ with $\mathbf{X}$. We are going to apply the formulae of the general simplified case to the separable scheme. The actual distortion D_a given in relation (1) is an estimation of the distortion

$$D(\mathbf{X}, \widehat{\mathbf{X}}) = \mathrm{E}[D_a(\mathbf{X}, \widehat{\mathbf{X}})] = \frac{1}{NL}\,\mathrm{E}[\|\mathbf{X} - \widehat{\mathbf{X}}\|^2]. \tag{4}$$

Now, in order to express the relation (3) in terms of the DWT $\mathbf{W}$ and the spectral transform $\mathbf{A}$, it is important to note first that the canonical basis of the space of matrices of dimension $N \times L$ is the family of matrices $\mathbf{e}_{i,k} = \mathbf{e}_i \mathbf{e}_k^{'T}$ ($1 \le i \le N$, $1 \le k \le L$), with $\mathbf{e}_i$ (resp. $\mathbf{e}_k^{'}$) the i^{th} (resp. k^{th}) vector of the canonical basis of $\mathbb{R}^N$ (resp. $\mathbb{R}^L$). Therefore, the weighting factor α_i in relation (3) depends here on the two indices i and k: $\alpha_{ik} = \|\mathcal{A}^{-1}\mathbf{e}_{i,k}\|^2$. Then, let

$$w_i = \|\mathbf{A}^{-1}\mathbf{e}_i\|^2 \qquad (1 \le i \le N), \tag{5}$$

we have $\mathcal{A}^{-1}\mathbf{e}_{i,k} = \mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_k^{'T}\mathbf{W}^{-T}$ and $\|\mathcal{A}^{-1}\mathbf{e}_{i,k}\|^2 = \mathrm{tr}[\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_k^{'T}\mathbf{W}^{-T}\mathbf{W}^{-1}\mathbf{e}_k^{'}\mathbf{e}_i^T\mathbf{A}^{-T}] = \mathbf{e}_i^T\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_k^{'T}\mathbf{W}^{-T}\mathbf{W}^{-1}\mathbf{e}_k^{'}$ and finally

$$\alpha_{ik} = \|\mathcal{A}^{-1}\mathbf{e}_{i,k}\|^2 = w_i\|\mathbf{W}^{-1}\mathbf{e}_k^{'}\|^2. \tag{6}$$

Therefore, according to Theorem 1, under assumption $\mathcal{H}_1$ we have

$$D(\mathbf{X}, \widehat{\mathbf{X}}) = \frac{1}{NL}\sum_{i=1}^{N}\sum_{k=1}^{L} w_i\|\mathbf{W}^{-T}\mathbf{e}_k^{'}\|^2\,\mathrm{E}[(Y_i(k) - Y_i^q(k))^2]. \tag{7}$$

Now, for any subband m ($1 \le m \le M$), let K_m be the number of columns in $\mathbf{Y}$ corresponding to that subband and let

$$\pi_m = \frac{K_m}{L} \qquad (1 \le m \le M) \tag{8}$$

be the ratio of wavelets coefficients that belong to it. If k ($1 \le k \le L$) refers to a column indice of the matrix $\mathbf{Y}$ located in that subband and if we assume that

> $\mathcal{H}_2$: *for any component i ($1 \le i \le N$), the distortion* $\mathrm{E}[(Y_i(k) - Y_i^q(k))^2] = D_i^{(m)}$ *does not depend on the spatial position k in the subband m,*

(which is the case under high resolution quantization hypothesis), then equation (7) becomes

$$D(\mathbf{X}, \widehat{\mathbf{X}}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M} \pi_m w_i \omega_m D_i^{(m)} \quad \text{with} \quad \omega_m = \frac{1}{K_m}\sum_{k}\|\mathbf{W}^{-T}\mathbf{e}_k^{'}\|^2, \tag{9}$$

where in the last summation, the range of k consists in the columns of $\mathbf{Y}$ with the subband m. Or, by adopting a different perspective, if we assume that

$\mathcal{H}_3$: *the weight* $\|\mathbf{W}^{-1}\mathbf{e}_k\|^2 = \omega_m$ *does not depend on the spatial position* k *in the subband* m,

then equation (7) becomes

$$D(\mathbf{X},\widehat{\mathbf{X}}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\pi_m\omega_m w_i D_i^{(m)} \quad \text{with} \quad D_i^{(m)} = \frac{1}{K_m}\sum_{k}\mathrm{E}[(Y_i(k)-Y_i^q(k))^2], \tag{10}$$

where in the last summation, the range of k consists in the columns of $\mathbf{Y}$ with the subband m.

Remark 1. *The condition* $\mathcal{H}_3$ *is satisfied by dyadic wavelets having Finite Impulse Response (FIR) synthesis filters, when edge effects are neglected (for more details see e.g. (Usevitch, 1996; Woods & Naven, 1992)).*

Lastly, we can notice that the actual distortion D_a given in equation (1) satisfies

$$D_a(\mathbf{X},\widehat{\mathbf{X}}) = \frac{1}{NL}\,\mathrm{tr}[(\mathbf{X}-\widehat{\mathbf{X}})(\mathbf{X}-\widehat{\mathbf{X}})^T] = \frac{1}{NL}\,\mathrm{tr}[\mathbf{A}^{-1}(\mathbf{Y}-\mathbf{Y}^q)\mathbf{W}^{-T}\mathbf{W}^{-1}(\mathbf{Y}-\mathbf{Y}^q)^T\mathbf{A}^{-T}],$$

therefore if we assume

$\mathcal{H}_4$: *the DWT is orthogonal, i.e.* $\mathbf{W}\mathbf{W}^T = \mathbf{I}_L$, *with* $\mathbf{I}_L$ *the identity matrix of dimension* L,

then $D_a(\mathbf{X},\widehat{\mathbf{X}}) = \frac{1}{NL}\,\mathrm{tr}[\mathbf{A}^{-1}(\mathbf{Y}-\mathbf{Y}^q)(\mathbf{Y}-\mathbf{Y}^q)^T\mathbf{A}^{-T}] = \frac{1}{NL}\sum_{m=1}^{M}\mathrm{tr}[\mathbf{A}^{-1}(\mathbf{Y}^{(m)}-\mathbf{Y}^{q(m)})(\mathbf{Y}^{(m)}-\mathbf{Y}^{q(m)})^T\mathbf{A}^{-T}]$.

Remark 2. *The hypothesis* $\mathcal{H}_4$ *is roughly satisfied with the approximately orthogonal Daubechies 9/7 DWT (indeed, a simulation shows that the infinity norm of the diagonal, and respectively the off diagonal, elements of* $\mathbf{W}^T\mathbf{W} - \mathbf{I}_L$ *is worth* 0.42 *and* 0.16, *for five levels of decomposition on a 1-D signal of length 512).*

Now, $\frac{1}{K_m}(\mathbf{Y}^{(m)}-\mathbf{Y}^{q(m)})(\mathbf{Y}^{(m)}-\mathbf{Y}^{q(m)})^T$ is the actual autocorrelation matrix of the m-th subband quantization noise. If we assume

$\mathcal{H}_1'$: in each subband, the actual autocorrelation matrix of the quantization noise is diagonal, i.e., $\frac{1}{K_m}(\mathbf{Y}^{(m)}-\mathbf{Y}^{q(m)})(\mathbf{Y}^{(m)}-\mathbf{Y}^{q(m)})^T = \mathrm{diag}(D_1^{(m)},\ldots,D_N^{(m)})$ $(1 \le m \le M)$,

then we have

$$\mathrm{tr}[\mathbf{A}^{-1}\,\mathrm{diag}(D_1^{(m)},\ldots,D_N^{(m)})\mathbf{A}^{-T}] = \sum_{i=1}^{N} w_i D_i^{(m)}$$

$$D_a(\mathbf{X},\widehat{\mathbf{X}}) = \frac{1}{N}\sum_{m=1}^{M}\pi_m\,\mathrm{tr}[\mathbf{A}^{-1}\,\mathrm{diag}(D_1^{(m)},\ldots,D_N^{(m)})\mathbf{A}^{-T}]$$

$$D_a(\mathbf{X},\widehat{\mathbf{X}}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\pi_m w_i D_i^{(m)}. \tag{11}$$

Moreover, if we assume $\mathcal{H}_4$ and

$\mathcal{H}_5$: *the spectral transform* $\mathbf{A}$ *is orthogonal, i.e.* $\mathbf{A}\mathbf{A}^T = \mathbf{I}_N$,

then

$$D_a(\mathbf{X},\widehat{\mathbf{X}}) = D_a(\mathbf{Y},\mathbf{Y}^q). \tag{12}$$

Let us state these results in the following theorem.

Theorem 2. *With the notations of Section 2.2, the end-to-end distortion of the separable scheme is given by:*

- *equation (9) under the assumptions $\mathcal{H}_1$ and $\mathcal{H}_2$;*
- *equation (10) under the assumptions $\mathcal{H}_1$ and $\mathcal{H}_3$;*
- *equation (11) under the assumptions $\mathcal{H}_1^{'}$ and $\mathcal{H}_4$;*
- *equation (12) under the assumptions $\mathcal{H}_4$ and $\mathcal{H}_5$.*

Remark 3. *1. The assumptions $\mathcal{H}_1$ and $\mathcal{H}_1^{'}$ are consequences of high resolution quantizations (see Subsection 3.2). They can also be deduced from the condition of statistical independence of the transformed components, since if the components of $\mathbf{Y}$ are independent, then the components of the quantization noise $\mathbf{Y} - \mathbf{Y}^q$, which is generally centered, are uncorrelated.*

2. *A method for the computation of the weighting wavelet coefficients ω_m $(1 \leq m \leq M)$ can be found in (Usevitch, 1996; Woods & Naven, 1992).*
3. *Since the assumptions $\mathcal{H}_1^{'}$, $\mathcal{H}_1$, ..., $\mathcal{H}_4$, are only approximatively satisfied, the equalities (9–13) are only approximations. However, we observed on many experiments that these approximations are very good for bit-rates greater than 0.25 bits per pixel and per band.*

We search the optimal spectral transform (that is the one which minimizes the total bit-rate for a given end-to-end distortion) which adapts to the data, assuming high resolution quantizations hypotheses and 2-D DWT with fixed coefficients, i.e., which do not adapt to the data. As already mentioned, in our tests we always used the Daubechies 9/7 DWT. First, we derive the criterion minimized by an optimal spectral transform. We emphasize the fact that we do not assume Gaussian data and that generally in the literature this assumption is made in order to clarify the criterion (coding gain) maximized by the optimal transform. However, the Bennett's formula and the optimal bit allocation between quantizers formula on which our criteria are based are well-known and therefore it is straightforward to deduce these criteria from well-known results. Our major innovation consists especially in the computation of the optimal transforms, since this computation is generally presented as a difficult task in classical transform coding and has never been done in the case of the separable scheme which is JPEG2000 compatible.

4. Criteria for optimal transforms under high resolution quantizations

We recall the extension of the Bennett's formula which can be stated as follows: if X is a real random variable quantized under the high resolution hypothesis, then the bit-rate of quantized variable X^q is well approximated by $H(X) - \frac{1}{2}\log_2(cD)$, where $H(X)$ is the differential entropy of X, D is the distortion (expected mean square error) introduced by the quantization and c is a constant depending on the quantization, e.g., for uniform scalar quantization $c = 12$ (Gray & Neuhoff, 1998). Hence, if $R_i^{(m)}$ denotes the quantizer bit-rate associated with component i and subband m, the Bennet's approximation gives

$$R_i^{(m)} \simeq H(Y_i^{(m)}) - \frac{1}{2}\log_2(cD_i^{(m)})$$

and the total bit-rate $R = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\pi_m R_i^{(m)}$ satisfies

$$R \simeq \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\pi_m\left[H(Y_i^{(m)}) - \frac{1}{2}\log_2(cD_i^{(m)})\right]. \tag{13}$$

The problem now consists in minimizing R under the constraint (given by Theorem 2)

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\pi_m\omega_m w_i D_i^{(m)} \leq D_t \tag{14}$$

for a given end-to-end distortion D_t. In other words, for a target end-to-end distortion D_t, how can the quantizer distortions $D_i^{(m)}$ be distributed in each subband of each component in order to minimize the total bit-rate? It is a classical problem in compression, called optimal bit allocation (Gersho & Gray, 1992), that can be solved as follows. According to relation (13), when the spectral and spatial transforms $\mathbf{A}$ and $\mathbf{W}$ are given, the differential entropies $H(Y_i^{(m)})$ and the factors w_i and ω_m are given. Then, the total bit-rate is minimized if and only if $\prod_{i=1}^{N}\prod_{m=1}^{M}(D_i^{(m)})^{\frac{\pi_m}{N}}$ is maximized, that is if and only if

$$\left[\prod_{i=1}^{N}\prod_{m=1}^{M}\left(D_i^{(m)}\right)^{\frac{\pi_m}{N}}\right]\left[\prod_{i=1}^{N}w_i\right]^{\frac{1}{N}}\left[\prod_{m=1}^{M}\omega_m^{\pi_m}\right]=\prod_{i=1}^{N}\prod_{m=1}^{M}\left(\omega_m w_i D_i^{(m)}\right)^{\frac{\pi_m}{N}}$$

is maximized. Now the mean inequality states the last expression (which is a geometric mean) is not greater than the arithmetic mean corresponding to the left member of inequality (14), with equality if and only if all the terms in the summation are equal. Hence, the minimization holds when $D_i^{(m)} = D_t\omega_m^{-1}w_i^{-1}$ for all m and i. That leads to

$$R \simeq \sum_{m=1}^{M}\pi_m\left[\frac{1}{N}\sum_{i=1}^{N}\left\{H(Y_i^{(m)})+\frac{1}{2}\log_2 w_i\right\}+\frac{1}{2}\log_2\omega_m\right]-\frac{1}{2}\log_2(cD_t)$$

and since w_i is the i^{th} diagonal element of $\mathbf{A}^{-T}\mathbf{A}^{-1}$, the other terms ω_m do not depend on $\mathbf{A}$, we obtain the following theorem.

Theorem 3. *For the separable scheme when the 2-D DWT has fixed coefficients, if high resolution quantizations hypotheses are assumed, then the optimal spectral transform* $\mathbf{A}$ *is an* $N \times N$ *matrix that minimizes the criterion:*

$$C_2(\mathbf{A}) = \sum_{j=1}^{N}\sum_{m=1}^{M}\pi_m H(Y_j^{(m)}) + \frac{1}{2}\log_2\det\operatorname{diag}(\mathbf{A}^{-T}\mathbf{A}^{-1}). \tag{15}$$

Remark 4. *Since* $\sum_{m=1}^{M}\pi_m = 1$, *the criterion* $C_2(\mathbf{A})$ *can be expressed as*

$$\begin{aligned} C_2(\mathbf{A}) &= \sum_{m=1}^{M}\pi_m\left[\sum_{i=1}^{N}H(Y_i^{(m)}) - \log_2|\det\mathbf{A}|\right] + \frac{1}{2}\log_2\left[\frac{\det\operatorname{diag}\left(\mathbf{A}^{-T}\mathbf{A}^{-1}\right)}{\det(\mathbf{A}^{-T}\mathbf{A}^{-1})}\right] \\ &= \sum_{m=1}^{M}\pi_m C_{ICA}^{(m)}(\mathbf{A}) + C_O(\mathbf{A}), \end{aligned} \tag{16}$$

where, for $1 \leq m \leq M$, $C_{ICA}^{(m)}(\mathbf{A}) = \sum_{i=1}^{N}H(Y_i^{(m)}) - \log_2|\det\mathbf{A}|$ *is the criterion to minimize when performing only ICA to the N components of the transformed coefficients that belong to the subband m. Pham (Pham, 2004) used that criterion to perform the algorithm* `ICAinf`.

Remark 5. *It results of Hadamard's inequality, that the term* $C_O(\mathbf{A}) = \frac{1}{2}\log_2 \frac{\det \operatorname{diag}(\mathbf{A}^{-T}\mathbf{A}^{-1})}{\det(\mathbf{A}^{-T}\mathbf{A}^{-1})}$ *is always positive or null (Narozny et al., 2008) and vanishes if and only if* $\mathbf{A}$ *is a matrix whose columns are pairwise orthogonal, therefore it can be seen like a kind of measure of deviation to orthogonality.*

The relation (16) shows that the criterion $C_2(\mathbf{A})$ takes into consideration the fact that one quantizer per subband and per component is allocated. It is also important to notice that the criterion $C_2(\mathbf{A})$ involves the transformed coefficients $\mathbf{Y}$. Therefore, even for the separable scheme (where the order of processing between the 2-D DWT and the spectral transform does not matter), the search of the optimal spectral transform must be done *after* the 2-D DWT.
Note that the separable compression scheme does not take into account the difference of statistics between subbands, since the same spectral transform is applied to all the subbands. Moreover it is well known that after a DWT some redundancies remain between adjoining wavelets coefficients. In (Akam Bita et al., 2010a), the authors introduced the subband compression scheme, that uses as many optimal spectral transforms as subbands in order to capture the difference of statistics between subbands, and the mixed subband compression scheme, that captures both redundancies between adjacent wavelet coefficients and the difference of statistics between subbands. Their experiments on hyperspectral images showed that these variants of the separable scheme, which are not JPEG2000 compatible, perform finally worse than the separable scheme because of the increasing of memory size occupied by the optimal spectral transforms in the bit stream.
Lastly, note that the algorithm that computes a KLT is customarily applied first to the image before the DWT, but this would be equivalent to applying it after the DWT (i.e. to the DWT coefficients) if the DWT is orthogonal (as is often the case or at least nearly so in practice[1]). But then it will not distinguish subbands: the DWT coefficients are considered as coming from a same (Gaussian) distribution, regardless of the subband they belong to. We feel that the higher performance — shown in § 7 — of criterion (15) over the criterion $\frac{1}{2}\log_2 \prod_{j=1}^{N} \operatorname{var}(Y_j)$, which leads to the KLT, is due primarily to the fact that it treats each subband separately rather than that treating the distribution in each subband as non Gaussian. This is logical since, after any DWT, the energy in each subband depends on the power spectrum of the input signal. It is important to notice that there is no contradiction in the fact that the criterion (15) treats each subband separately, while the same spectral transform $\mathbf{A}$ is applied to all the subbands. The idea is then to introduce the distinction between subbands but retain the (approximate) Gaussian assumption used by the KLT. The distribution of all the wavelet coefficients (with no distinction between subbands) is a mixture of distributions of the coefficients in the subbands. It can be shown that the kurtosis of the mixed distribution is higher than the average kurtosis of the individual distributions. In particular, mixture of Gaussian distributions has always a positive kurtosis, unless all the individual distributions are the same. Thus the wavelet coefficients, regardless of the subband they belong to, have a positive kurtosis even if in each subband their distribution is Gaussian. The above consideration suggests modifying the criterion (15) by treating the transformed coefficients in each subband m as having a Gaussian distribution *with differing variance for different* m. The transformation minimizing this modified criterion is no longer optimal, but can be nearly so if the distribution in each subband is not too far from Gaussian. This is not an unrealistic situation: the wavelet coefficients in a subband is the (decimated) output of a bandpass filter which tends to produce more Gaussian output than input, due to the reasoning (given e.g. in (Papoulis, 1984) section 8-5) that yields to the proof of the Central Limit Theorem. The advantage of the modified criterion is that it avoids

[1] See Remark 2.

the entropy estimation and uses only second order statistics. Thus its minimization requires much less computer resources than using (15).

5. A simplified criterion using only second order statistics

Let $H^-(Z) = \log_2 \sqrt{\mathrm{var}(Z)2\pi e} - H(Z)$ denote the negentropy of Z (which is the difference of entropy between a Gaussian distribution with variance $\mathrm{var}(Z)$ and the distribution of Z), it is non negative and vanishes if and only if Z is Gaussian. The criterion (15) can be rewritten for orthogonal[2] matrices

$$C_\perp(\mathbf{A}) = -\sum_{i=1}^{N}\sum_{m=1}^{M} \pi_m H^-(Y_i^{(m)}) + \frac{1}{2}\sum_{i=1}^{N}\sum_{m=1}^{M} \pi_m \log_2[\mathrm{var}(Y_i^{(m)})2\pi e]. \tag{17}$$

An analysis of criterion (17) shows that it takes into account two phenomena: 1) the non Gaussianity of the transformed coefficients $Y_i^{(m)}$ for $1 \leq m \leq M$ and $1 \leq i \leq N$ — this is controlled by the first term — and 2) the inhomogeneity of the variances in the subbands — this is controlled by the second term. It is natural to explore the case where the second phenomenon is the most important, since the DWT tends to render the variables more Gaussian. In practice, this condition is generally roughly satisfied, except in the LL subband (a subband of lowest resolution) for which the weighting coefficient π_m is generally small. Thus, if we neglect the variation, induced by the spectral transform $\mathbf{A}$, of the first term in the right member of equation (17), and if we consider only orthogonal matrices $\mathbf{A}$, then the optimal transform minimizes the new criterion

$$C'(\mathbf{A}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{m=1}^{M} \pi_m \log_2[\mathrm{var}(Y_i^{(m)})]. \tag{18}$$

Furthermore if we assume in each component the transformed coefficients have all the same variance, regardless of the subband they belong to, then the criterion (18) becomes $\frac{1}{2}\log_2\left[\prod_{i=1}^{N}\mathrm{var}(Y_i)\right]$, leading to the KLT.

In the following, we express criterion (18) in terms of the covariance matrices of the wavelets coefficients $\mathbf{X}\mathbf{W}^T = \left[\,(\mathbf{X}\mathbf{W}^T)^{(1)}\ (\mathbf{X}\mathbf{W}^T)^{(2)}\ \cdots\ (\mathbf{X}\mathbf{W}^T)^{(M)}\,\right]$ located in the same subband. The matrix $(\mathbf{X}\mathbf{W}^T)^{(m)}$ is of dimension $N \times \pi_m L$. Its columns can be considered as different realizations of a random vector of dimension N whose covariance matrix is denoted $\mathbf{C}^{(m)}$. Now, $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{W}^T$ can be written $\mathbf{Y} = [\mathbf{Y}^{(1)} \cdots \mathbf{Y}^{(M)}]$, where $\mathbf{Y}^{(m)} = (\mathbf{A}\mathbf{X}\mathbf{W}^T)^{(m)}$ is a matrix whose columns can also be considered as different realizations of a random vector having $\mathbf{A}\mathbf{C}^{(m)}\mathbf{A}^T$ as covariance matrix. With these notations, we have $\prod_{j=1}^{N}\mathrm{var}(Y_j^{(m)}) = \det\mathrm{diag}(\mathbf{A}\mathbf{C}^{(m)}\mathbf{A}^T)$ and hence the new criterion becomes

$$C'(\mathbf{A}) = \frac{1}{2}\sum_{m=1}^{M} \pi_m \log_2 \det\mathrm{diag}(\mathbf{A}\mathbf{C}^{(m)}\mathbf{A}^T) \tag{19}$$

to be minimized with respect to $\mathbf{A}$, under the constraint that it is orthogonal.

[2] The orthogonality constraint will be justified in § 7 in which we find that minimizing (15) with and without this constraint yields almost the same performances. With the orthogonality constraint, the second term in (15) vanishes.

The FG algorithm in (Flury & Gautschi, 1986) can be used to minimize the above criterion. We have developed a slightly different algorithm (called JADO) which is briefly described in Appendix 6.3.

6. Minimization of the criteria for the separable scheme

We explain now three algorithms that minimize the criterion (15), one with no constraint but invertibility, another with the constraint of orthogonality and the third with the constraints of orthogonality and Gaussian data. To simplify some mathematical expressions we shall use the Neperian logarithm instead of the base two logarithm until the end of this section.

6.1 The algorithm OST

As in (Pham, 2004) and (Narozny et al., 2008), the algorithms of minimization are based on a quasi-Newton method with the relative gradient and a simplified relative Hessian. Starting with a current estimator $\mathbf{A}$, the method consists of expanding $C_2(\mathbf{A}+\mathcal{E}\mathbf{A})$ with respect to the matrix $\mathcal{E}=[\mathcal{E}_{ij}]$ up to the second order, in a neighborhood of $\mathcal{E}=\mathbf{0}_N$ (the null matrix), and then minimizing the resulting quadratic form in $\mathcal{E}$ to obtain a new estimate. Using the results of (Pham, 2005) it is straightforward to deduce that the Taylor expansion up to the second order of $C_{ICA}^{(m)}(\mathbf{A}+\mathcal{E}\mathbf{A})$ can be approximated as follows

$$C_{ICA}^{(m)}(\mathbf{A}+\mathcal{E}\mathbf{A}) = C_{ICA}^{(m)}(\mathbf{A}) + \sum_{1\le i\ne j\le N} \mathrm{E}[\psi_{Y_i^{(m)}}(Y_i^{(m)})Y_j^{(m)}]\mathcal{E}_{ij} + \frac{1}{2}\sum_{1\le i\ne j\le N}\{\mathrm{E}[\psi^2_{Y_i^{(m)}}(Y_i^{(m)})]\,\mathrm{E}[Y_j^{(m)2}]\mathcal{E}_{ij}^2 + \mathcal{E}_{ij}\mathcal{E}_{ji}\} + \cdots, \tag{20}$$

where the function $\psi_{Y_i^{(m)}}$ is equal to the derivative of $-\log p(y_i^{(m)})$ — $p(y_i^{(m)})$ denoting the probability density function of $Y_i^{(m)}$ — and is known as the score function. Let $\mathbf{M}=\mathbf{A}^{-T}\mathbf{A}^{-1}$. In (Narozny et al., 2008), the Taylor expansion of $C_O(\mathbf{A}+\mathcal{E}\mathbf{A})$ is given up to the second order, however it is quite involved and it is simplified into

$$C_O(\mathbf{A}+\mathcal{E}\mathbf{A}) \approx C_O(\mathbf{A}) - \sum_{1\le i\ne j\le N}\frac{M_{ji}}{M_{ii}}\mathcal{E}_{ji} + \frac{1}{2}\sum_{1\le i\ne j\le N}\left[\frac{M_{jj}}{M_{ii}}\mathcal{E}_{ji}^2 + \mathcal{E}_{ji}\mathcal{E}_{ij}\right] + \cdots \tag{21}$$

by neglecting the non diagonal elements of $\mathbf{M}=[M_{ij}]$ in the second order terms of the Taylor expansion.
Using the approximation (21), the equality (20) and the relation (16) we obtain

$$C_2(\mathbf{A}+\mathcal{E}\mathbf{A}) = C_2(\mathbf{A}) + \sum_{1\le i\ne j\le N}\left[\sum_{m=1}^{M}\pi_m E\left[Y_j^{(m)}\psi_{Y_i^{(m)}}\left(Y_i^{(m)}\right)\right] - \frac{\mathbf{M}_{ij}}{\mathbf{M}_{jj}}\right]\mathcal{E}_{ij} + \frac{1}{2}\sum_{1\le i\ne j\le N}\left[\sum_{m=1}^{M}\pi_m\mathcal{E}_{ij}^2 E\left[Y_j^{(m)2}\right]E\left[\psi^2_{Y_i^{(m)}}\left(Y_i^{(m)}\right)\right] + \frac{M_{ii}}{M_{jj}}\mathcal{E}_{ij}^2 + 2\mathcal{E}_{ij}\mathcal{E}_{ji}\right]. \tag{22}$$

The quadratic form associated to this last expansion is positive definite. One iteration of the algorithm is first to solve the following equation

$$\begin{bmatrix}\Psi_{ij} & 2\\ 2 & \Psi_{ji}\end{bmatrix}\begin{pmatrix}\mathcal{E}_{ij}\\ \mathcal{E}_{ji}\end{pmatrix} = \begin{pmatrix}\Phi_{ij}\\ \Phi_{ji}\end{pmatrix}, \tag{23}$$

with $\Phi_{ij} = \frac{M_{ij}}{M_{jj}} - \sum_{m=1}^{M} \pi_m E[\psi_{Y_i^{(m)}}(Y_i^{(m)}) Y_j^{(m)}]$ and $\Psi_{ij} = \sum_{m=1}^{M} \pi_m E[\psi^2_{Y_i^{(m)}}(Y_i^{(m)})] E[Y_i^{(m)2}] + \frac{M_{ii}}{M_{jj}}$ and then to replace the current solution $\mathbf{A}$ with $\mathbf{A} + \mathcal{E}\mathbf{A}$. Since the diagonal elements of $\mathcal{E}$ are undetermined, they are arbitrarily fixed to zero. For the practical computation of the algorithm, we replace $\psi_{Y_i^{(m)}}$ with its estimator $\widehat{\psi}_{Y_i^{(m)}}$ that is described in (Pham, 2005) as well as the estimator of the differential entropy. The mathematical expectations are replaced with simple empirical means. We call `OST` (*Optimal Spectral Transform*) the algorithm described above and OST the optimal transform returned by this algorithm.

6.2 The algorithm `OrthOST`

To minimize the criterion (15) with the constraint that the solution is an orthogonal matrix, it is important to note, as in (Narozny et al., 2008), that if $\mathbf{A}$ is orthogonal, then $\mathbf{A} + \mathcal{E}\mathbf{A}$ remains orthogonal when $\mathbf{I} + \mathcal{E}$ is also orthogonal. This condition is satisfied up to the first order if $\mathcal{E}$ is an antisymmetrical matrix, since then $(\mathbf{I} + \mathcal{E})^T(\mathbf{I} + \mathcal{E}) = \mathbf{I} + \mathcal{E}^T\mathcal{E}$. Using that condition, the expansion (22) becomes

$$C(\mathbf{A} + \mathcal{E}\mathbf{A}) = C(\mathbf{A}) + \sum_{m=1}^{M} \sum_{1 \le i < j \le N} \pi_m \left\{ E[Y_j^{(m)} \psi_{Y_i^{(m)}}(Y_i^{(m)})] - E[Y_i^{(m)} \psi_{Y_j^{(m)}}(Y_j^{(m)})] \right\} \mathcal{E}_{ij} +$$

$$\frac{1}{2} \sum_{1 \le i < j \le N} \left[\sum_{m=1}^{M} \pi_m \left\{ E[Y_j^{(m)2}] E[\psi^2_{Y_i^{(m)}}(Y_i^{(m)})] + E[Y_i^{(m)2}] E[\psi^2_{Y_j^{(m)}}(Y_j^{(m)})] \right\} - 2 \right] \mathcal{E}_{ij}^2. \quad (24)$$

The matrix $\mathcal{E}$ is calculated in that case according to

$$\mathcal{E}_{ij} = \frac{\sum_{m=1}^{M} \pi_m \left\{ E[Y_i^{(m)} \psi_{Y_j^{(m)}}(Y_j^{(m)})] - E[Y_j^{(m)} \psi_{Y_i^{(m)}}(Y_i^{(m)})] \right\}}{\sum_{m=1}^{M} \pi_m \left\{ E[Y_j^{(m)2}] E[\psi^2_{Y_i^{(m)}}(Y_i^{(m)})] + E[Y_i^{(m)2}] E[\psi^2_{Y_j^{(m)}}(Y_j^{(m)})] \right\} - 2}. \quad (25)$$

Actually, $\mathbf{A} + \mathcal{E}\mathbf{A}$ obtained in this way is not a true orthogonal matrix. This can be overcome by replacing $\mathbf{A} + \mathcal{E}\mathbf{A}$ with $e^{\mathcal{E}}\mathbf{A} = (\mathbf{I} + \mathcal{E} + \mathcal{E}^2/2! + \cdots)\mathbf{A}$, which is orthogonal and differs from $\mathbf{A} + \mathcal{E}\mathbf{A}$ only by second order terms. We call `OrthOST` (*Orthogonal Optimal Spectral Transform*) this algorithm and OrthOST the orthogonal transform returned by the algorithm. The case where the spectral transform is constrained to be orthogonal is particularly interesting because the weightings which depend on the linear transform are all equal to one.

6.3 The `JADO` (Joint Approximate Diagonalization under Orthogonality constraint) algorithm

Given K positive definite (complex) matrices $\mathbf{C}_1, \ldots, \mathbf{C}_K$ associated with positive weights $w_1, \ldots, w_K$, the `JADO` algorithm aims to find a unitary matrix $\mathbf{B}$ which minimizes

$$C(\mathbf{B}) = \sum_{k=1}^{K} w_k \log\det \operatorname{diag}(\mathbf{B}\mathbf{C}_k\mathbf{B}^*) \quad (26)$$

where * denotes the hermitian operator. This algorithm differs only slightly from FG algorithm in (Flury & Gautschi, 1986). However, its derivation in (Flury & Gautschi, 1986) is complex and difficult to understand. Here we provide briefly a much simpler derivation.

The idea is to make successive Givens rotations, each time on a pair of rows of $\mathbf{B}$, the ith row $\mathbf{B}_{i\cdot}$ and the jth row $\mathbf{B}_{j\cdot}$, say:

$$\begin{bmatrix} \mathbf{B}_{i\cdot} \\ \mathbf{B}_{j\cdot} \end{bmatrix} \leftarrow \mathbf{T}_{ij} \begin{bmatrix} \mathbf{B}_{i\cdot} \\ \mathbf{B}_{j\cdot} \end{bmatrix}, \tag{27}$$

where $\mathbf{T}_{ij}$ is a 2×2 unitary matrix, chosen so that the criterion is decreased. The processing of all the $\frac{K(K-1)}{2}$ pairs is called a sweep. The algorithm consists of repeated sweeps until convergence is achieved.
The decrease of the criterion (26) induced by (27) is

$$\sum_{k=1}^{K} w_k \log \left[(\mathbf{B}_{i\cdot}\mathbf{C}_k\mathbf{B}_{i\cdot}^*)(\mathbf{B}_{j\cdot}\mathbf{C}_k\mathbf{B}_{j\cdot}^*) \Big/ \det \operatorname{diag} \left(\mathbf{T}_{ij} \begin{bmatrix} \mathbf{B}_{i\cdot} \\ \mathbf{B}_{j\cdot} \end{bmatrix} \mathbf{C}_k \begin{bmatrix} \mathbf{B}_{i\cdot}^* & \mathbf{B}_{j\cdot}^* \end{bmatrix} \mathbf{T}_{ij}^* \right) \right].$$

A natural idea is to chose $\mathbf{T}_{ij}$ to maximize this decrease, but there is no closed form formulae for that. Our idea is to maximize a lower bound of it instead. Since for $a > 0, b \geq 0, \log(a/b) \geq 1 - b/a$, the above decrease can be seen to be bounded below by

$$2(w_1 + \cdots + w_K) - \mathbf{T}_{ij;1\cdot}\mathbf{P}\mathbf{T}_{ij;1\cdot}^* - \mathbf{T}_{ij;2\cdot}\mathbf{Q}\mathbf{T}_{ij;2\cdot}^*, \tag{28}$$

where $\mathbf{T}_{ij;1\cdot}$ and $\mathbf{T}_{ij;2\cdot}$ are the first and second rows of $\mathbf{T}_{ij}$ and

$$\mathbf{P} = \sum_{k=1}^{K} \frac{w_k}{\mathbf{B}_{i\cdot}\mathbf{C}_k\mathbf{B}_{i\cdot}^*} \begin{bmatrix} \mathbf{B}_{i\cdot} \\ \mathbf{B}_{j\cdot} \end{bmatrix} \mathbf{C}_k [\, \mathbf{B}_{i\cdot}^* \;\; \mathbf{B}_{j\cdot}^* \,]; \quad \mathbf{Q} = \sum_{k=1}^{K} \frac{w_k}{\mathbf{B}_{j\cdot}\mathbf{C}_k\mathbf{B}_{j\cdot}^*} \begin{bmatrix} \mathbf{B}_{i\cdot} \\ \mathbf{B}_{j\cdot} \end{bmatrix} \mathbf{C}_k [\, \mathbf{B}_{i\cdot}^* \;\; \mathbf{B}_{j\cdot}^* \,].$$

Since $\mathbf{T}_{ij;2\cdot}$ has unit norm and is orthogonal to $\mathbf{T}_{ij;1\cdot}$, it must be of the form $e^{i\alpha}\overline{\mathbf{T}}_{ij;1\cdot}\mathbf{J}$ where α is some phase angle, $\overline{x}$ denotes the complex conjugate of x and $\mathbf{J}$ is the 2×2 matrix with 0 on the diagonal and $1, -1$ on the anti-diagonal. Thus $\mathbf{T}_{ij;2\cdot}\mathbf{Q}\mathbf{T}_{ij;2\cdot}^* = \overline{\mathbf{T}}_{ij;1\cdot}\mathbf{J}\mathbf{Q}\mathbf{J}^*\overline{\mathbf{T}}_{ij;1\cdot}^*$, but since the above left hand side is real (as $\mathbf{Q}$ is hermitian), it also equals $\mathbf{T}_{ij;1\cdot}\mathbf{J}\overline{\mathbf{Q}}\mathbf{J}^*\mathbf{T}_{ij;1\cdot}^*$. Therefore expression (28) can be rewritten as $2(w_1 + \cdots + w_K) - \mathbf{T}_{ij;1\cdot}(\mathbf{P} + \mathbf{J}\overline{\mathbf{Q}}\mathbf{J}^*)\mathbf{T}_{ij;1\cdot}^*$. Maximizing it with respect to the unitary matrix $\mathbf{T}$ thus amounts to minimizing $\mathbf{T}_{ij;1\cdot}(\mathbf{P} + \mathbf{J}\overline{\mathbf{Q}}\mathbf{J}^*)\mathbf{T}_{ij;1\cdot}^*$ with respect to the vector of unit norm $\mathbf{T}_{ij;1\cdot}$. The solution is that $\mathbf{T}_{ij;1\cdot}$ is (up to a factor of unit modulus) the normalized left eigenvector of the smallest eigenvalue of $\mathbf{P} + \mathbf{J}\overline{\mathbf{Q}}\mathbf{J}^*$. Since $\mathbf{T}_{ij;2\cdot}$ is orthogonal to $\mathbf{T}_{ij;1\cdot}$ it is the other eigenvector. Finally, $\mathbf{T}_{ij}$ is the matrix formed by the left eigenvectors of $\mathbf{P} + \mathbf{J}\overline{\mathbf{Q}}\mathbf{J}^*$. Its elements can be computed explicitly in closed form as follows.
We note that the off diagonal elements of $\mathbf{J}\overline{\mathbf{Q}}\mathbf{J}^*$ is the negative of those of $\mathbf{Q}$ while the diagonal elements are those of $\mathbf{Q}$ *in reverse order*. Thus $\mathbf{J}\overline{\mathbf{Q}}\mathbf{J}^* = \operatorname{tr}(\mathbf{Q})\mathbf{I} - \mathbf{Q}$ where tr denotes the trace. Since the addition of a multiple of the identity matrix does not change the eigenvectors, $\mathbf{T}_{ij}$ is also the matrix formed by the left eigenvectors of $\mathbf{P} - \mathbf{Q}$. One can now recognize that the rotation (27) is the same as an iteration in the G loop of the FG algorithm. However, it differs from our `JADO` algorithm in that it repeats (27) with the same pair i, j (but with the newly computed $\mathbf{B}_{i\cdot}$ and $\mathbf{B}_{j\cdot}$) until convergence (the G loop) and only then another pair i, j is considered. We feel that this is not efficient since the decrease of the criterion will be very small near the end of the G loop. We call JADOST the transform returned by the algorithm.

6.4 Computational complexity of the optimal transforms

We give here a rough estimation of the number of operations required for the computation of the two first algorithms described above, taking into account only multiplications and divisions. The differential entropies and the score functions are calculated according to a method explained in (Pham, 2005). The computational complexity of each of these quantities is O(NrL), where r is the number of bins in the binned kernel density estimation. In general $r \ll L$ and for most cases r belongs to the interval $[30\,,\,60]$. At each iteration, the criterion and the matrix $\mathcal{E}$ must be computed. The complexity of the criterion computation is O($NrL + N^3$). For the calculation of the matrix $\mathcal{E}$, we first need to compute the score function. The complexity of the matrix $\mathcal{E}$ computation (including the score function computation) is O($NrL + N^2L$). Finally, the complexity of one iteration is O($NrL + N^2L + N^3$). In practice, the convergence of the algorithm is usually obtained after p iterations, $p \in [20\,,\,60]$. Generally $N \ll L$ and the total computational complexity is O($p(NrL + N^2L)$). The computational complexities of `OrthOST` is the same. We recall that for the computation of a KLT, this complexity is $O(LN^2)$. The JADOST and KLT computation complexities are roughly the same.

7. Experimental results

In this section we present the performances in image compression of the optimal transforms described in the previous sections.

7.1 Description of the tests

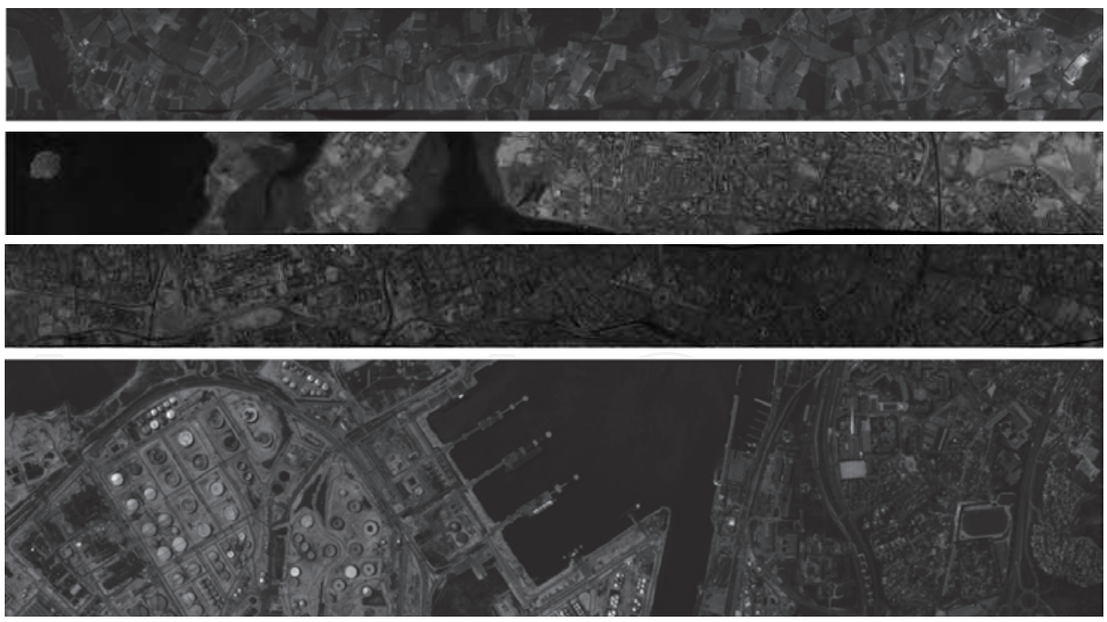

Fig. 1. From up to down *Moissac*, *Vannes*, *Toulouse*, *Port-de-Bouc*

We tested two kinds of multicomponent images: multispectral ones and hyperspectral ones. The multispectral images are[3] PLEIADES simulations of French cities with $N = 4$ components and coded on $N_b = 12$ bpppb: *Moissac* with $N_c \times N_r = 320 \times 3152$, *Port-de-Bouc* with

[3] These images have been given by the French Space Agency CNES (Centre National d'Etudes Spatiales). They are described on the web site `http://smsc.cnes.fr/PLEIADES/`.

$N_c \times N_r = 320 \times 1376$, *Toulouse* with $N_c \times N_r = 352 \times 3816$, *Vannes* with $N_c \times N_r = 352 \times 3736$, ... The hyperspectral images are[4] AVIRIS images (*Moffett*, *Cuprite* and *Jasper*) with $N = 224$ components from the visible to the infrared and coded on $N_b = 16$ bpppb. They are originally acquired with $N_r \times N_c = 512 \times 624$, but for the simulations we kept only the 512 leftmost columns. Some images used in our tests are shown in figures 1 and 2. As already

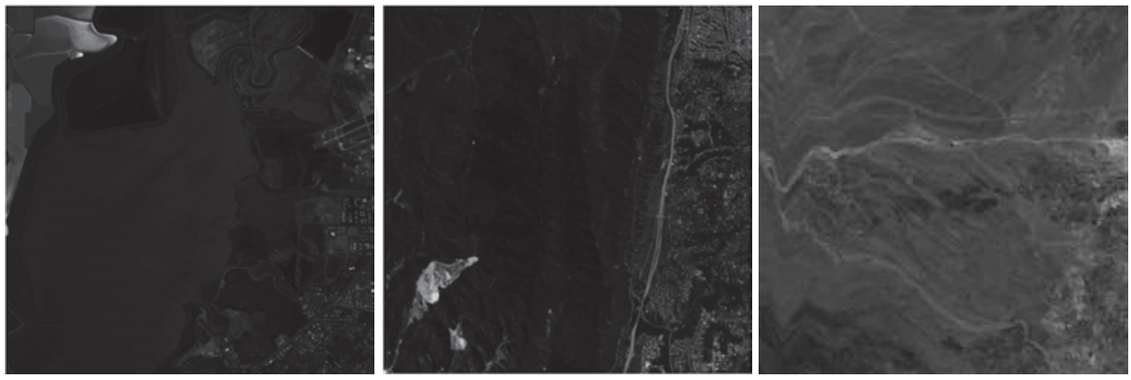

Fig. 2. From left to right: *Moffett*, *Jasper* and *Cuprite*

mentioned, the 2-D DWT used in all our experiments is the Daubechies 9/7 which proved to be efficient in lossy image compression (Antonini et al., 1992; Taubman & Marcellin, 2002). For simplicity, we used only uniform scalar quantizers with a dead zone twice as large as the quantization step. The performances are evaluated in terms of bit-rate versus end-to-end distortion. For hyperspectral images, we considered four distortions. A first one is the mean square error (MSE) expressed in terms of the Signal to Noise Ratio, $\text{SNR} = 10\log_{10}(\sigma^2/D)$ where D is the actual end-to-end MSE distortion and $\sigma^2 = \sum_{i=1}^{N}\sum_{n=1}^{L}(X_i(n)-\mu)^2/(NL)$ is the empirical variance of the initial image with the empirical mean of the image $\mu = \sum_{i=1}^{N}\sum_{n=1}^{L}X_i(n)/(NL)$. A second distortion is the maximal absolute difference ($\text{MAD} = \max\{|X_i(n)-\widehat{X}_i(n)| : 1 \le i \le N \text{ and } 1 \le n \le L\}$), a third one is the maximum spectral angle $\text{MSA} = \max\left\{\text{acos}\left(\frac{\sum_{i=1}^{N}X_i(n)\widehat{X}_i(n)}{\sqrt{\sum_{i=1}^{N}X_i^2(n)\sum_{i=1}^{N}\widehat{X}_i^2(n)}}\right) : 1 \le n \le L\right\}$ and the last one is the mean absolute error ($\text{MAE} = \sum_{i=1}^{N}\sum_{n=1}^{L}|X_i(n)-\widehat{X}_i(n)|/(NL)$). With these four distortions, one can estimate the performances of a codec on usual applications of hyperspectral images, like classifications and targets detections (Christophe et al., 2005). For multispectral images, we considered only the MAD and the MSE distortions, the last one being expressed in terms of Peak of Signal to Noise Ratio, $\text{PSNR} = 10\log_{10}[(2^{N_b}-1)^2/D]$, where D is the actual end-to-end MSE distortion and N_b is the number of bits per pixel and per band (bpppb) of the initial image. The bit-rate, expressed in bpppb, was measured on the actual bit stream obtained with the JPEG2000 coder EBCOT (Taubman, 2000) and its PCRD optimizer applied across components for optimal bit allocation. We used the Verification Model version 9.1 (VM9 (JPEG2000, 2001)) codec developed by the JPEG2000 group. The coefficients of $\mathbf{A}^{-1}$ (the inverse matrix of the optimal spectral transform) and the mean of each component are stored in the bitstream as `float32` data (this costs $32(N+1)/L$ bpppb). A difference exists between the aimed bit-rate and the actual bit-rate obtained with the VM9. In our tests, this difference does not exceed ± 0.001 bpppb and thus the precision of the PSNR is about ± 0.05 dB.

[4] These images have been downloaded from the NASA web site `http://aviris.jpl.nasa.gov/`.

7.2 Bit-rate versus distortion performances

In this subsection, we discuss and compare the bit-rate versus distortion performances of different spectral transforms. Table 1 presents the bit-rate of different transforms versus the

	PSNR (dB)								MAD							
bit-rate	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00
	Moissac															
Id	36.37	39.59	41.93	43.89	47.22	50.16	52.93	55.66	**691**	366	253	187	108	68	49	38
KLT	38.61	42.39	45.24	47.63	51.51	54.49	56.98	59.44	716	381	**214**	**135**	79	48	32	25
JADOST	38.54	42.30	45.14	47.52	51.39	54.42	56.98	59.47	700	**357**	298	137	79	**46**	**31**	24
OrthOST	38.67	42.50	45.35	47.72	51.55	54.55	57.11	59.60	818	399	229	145	**78**	47	33	24
OST	**38.69**	**42.55**	**45.43**	**47.80**	**51.62**	**54.59**	**57.15**	**59.65**	745	496	215	138	**78**	48	32	**23**
	Port-de-Bouc															
Id	30.36	33.68	36.14	38.25	41.93	45.27	48.43	51.52	1198	653	544	361	198	135	85	64
KLT	**33.47**	37.74	40.88	43.45	47.53	50.89	53.82	56.53	922	513	**297**	**230**	139	**74**	**50**	35
JADOST	33.26	37.56	40.73	43.32	47.47	50.91	53.93	56.68	922	**504**	324	256	135	75	52	32
OrthOST	33.42	37.80	41.05	43.71	47.90	51.31	54.28	56.99	885	513	305	237	**122**	77	51	**31**
OST	33.46	**37.85**	**41.12**	**43.78**	**48.00**	**51.40**	**54.36**	**57.06**	**866**	557	351	256	129	82	53	35
	Vannes															
Id	39.25	42.89	45.67	47.99	51.77	54.80	57.51	60.11	603	269	178	109	63	42	29	21
KLT	41.36	45.71	48.78	51.11	54.38	56.82	59.24	61.79	482	219	148	86	51	33	**24**	18
JADOST	41.83	46.15	49.16	51.42	54.61	57.09	59.53	62.06	368	214	**134**	**84**	48	**29**	**24**	19
OrthOST	41.90	46.27	49.29	51.54	54.71	57.18	59.62	62.16	**354**	**190**	135	91	46	30	25	18
OST	**41.94**	**46.34**	**49.35**	**51.59**	**54.74**	**57.22**	**59.68**	**62.20**	393	204	138	88	**45**	33	25	**16**
	Strasbourg															
Id	30.82	34.19	36.73	38.91	42.70	46.09	49.20	52.13	1357	877	546	353	205	118	86	60
KLT	**32.51**	**36.59**	39.77	42.49	46.99	50.58	53.51	56.08	1041	927	**438**	403	184	90	52	**38**
JADOST	32.47	36.51	39.65	42.33	46.78	50.36	53.33	55.92	1082	**872**	543	371	189	85	56	45
OrthOST	**32.51**	**36.59**	39.78	42.50	47.01	50.61	53.55	56.11	**1010**	948	449	404	178	87	**50**	**38**
OST	32.49	**36.59**	**39.79**	**42.53**	**47.07**	**50.67**	**53.60**	**56.17**	1149	904	455	**289**	**162**	**81**	55	42
	Montpellier															
Id	32.17	35.23	37.59	39.62	43.17	46.30	49.17	51.95	1216	630	406	292	168	117	77	54
KLT	34.09	37.75	40.60	43.03	47.20	50.69	53.63	56.18	747	488	340	248	143	75	49	34
JADOST	34.09	37.72	40.55	42.99	47.15	50.62	53.55	56.13	782	501	**323**	245	124	81	51	36
OrthOST	34.08	37.90	40.92	43.46	47.72	51.16	54.01	56.55	**681**	**454**	338	255	**127**	**68**	47	**32**
OST	**34.14**	**37.99**	**41.01**	**43.56**	**47.79**	**51.21**	**54.06**	**56.60**	704	483	332	**239**	**127**	**68**	**46**	33
	Perpignan															
Id	33.71	36.90	39.34	41.43	45.04	48.17	51.04	53.78	984	526	332	230	158	89	62	42
KLT	36.51	40.44	43.29	45.60	49.33	52.36	54.99	57.52	726	435	245	172	**84**	**54**	41	29
JADOST	36.55	40.51	43.37	45.69	49.43	52.44	55.07	57.60	715	388	234	172	90	58	39	30
OrthOST	36.59	40.59	43.48	45.83	49.61	52.66	55.30	57.82	721	**371**	**232**	165	94	**54**	**37**	30
OST	**36.60**	**40.60**	**43.49**	**45.84**	**49.62**	**52.67**	**55.32**	**57.85**	**645**	383	292	**164**	94	55	38	**28**

Table 1. Bit-rate (in bpppb) versus PSNR (in dB) and versus MAD of different spectral transforms on multispectral images (best results are bolded). The bit-rate was computed with the VM9.

two distortions PSNR and MAD on six multispectral images and Tables 2 and 3 show the bit-rate of different transforms versus the four distortions SNR (in dB), MAE, MAD and MSA (expressed in degree °) on three hyperspectral images. All the 2-D DWT was applied with five levels of decomposition. We observe the well-known fact that spectral transforms perform significantly better than the identity matrix (i.e., no spectral transform), especially for hyperspectral images. Indeed, on six multispectral images (see Table 1) the average gains

	SNR (dB)								MAE						
bit-rate	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	0.50	0.75	1.00	1.50	2.00	2.50	3.00
	Moffett														
Id	25.45	30.37	33.97	36.94	41.78	45.76	49.15	52.01	24.32	16.93	12.52	7.62	5.02	3.48	2.51
KLT	44.21	47.68	50.08	51.97	54.76	57.10	59.21	61.04	3.83	3.03	2.49	1.82	1.39	1.07	0.85
JADOST	45.13	48.39	50.70	52.50	55.17	57.47	59.53	61.30	3.54	2.83	2.35	1.74	1.33	1.03	0.82
OrthOST	45.31	**48.57**	**50.87**	52.61	55.28	57.57	59.62	61.37	**3.47**	**2.78**	**2.32**	**1.72**	1.31	1.02	0.81
OST	**45.32**	48.56	**50.87**	**52.62**	**55.30**	**57.64**	**59.77**	**61.63**	**3.47**	**2.78**	**2.32**	**1.72**	**1.30**	**1.00**	**0.79**
	Cuprite														
Id	29.99	33.48	36.12	38.41	42.44	45.99	49.19	52.11	26.07	19.85	15.60	10.13	6.89	4.83	3.47
KLT	47.79	50.46	52.55	54.16	56.76	59.07	61.26	63.27	3.96	3.23	2.73	2.04	1.55	1.19	0.92
JADOST	48.22	50.85	52.86	54.42	56.97	59.27	61.44	63.43	3.80	3.13	**2.65**	1.99	1.51	1.16	0.90
OrthOST	48.25	50.88	**52.89**	**54.44**	56.99	59.29	61.46	63.44	**3.79**	**3.12**	**2.65**	**1.98**	1.51	1.16	0.90
OST	**48.26**	**50.89**	**52.89**	**54.44**	**57.01**	**59.34**	**61.56**	**63.60**	**3.79**	**3.12**	**2.65**	**1.98**	**1.50**	**1.14**	**0.88**
	Jasper														
Id	21.34	24.83	27.56	29.92	34.01	37.67	41.09	44.33	64.84	34.39	26.82	17.23	11.52	7.89	5.49
KLT	42.93	46.49	48.61	50.37	53.18	55.56	57.72	59.66	4.04	3.27	2.72	1.99	1.51	1.16	0.91
JADOST	43.56	46.89	48.97	50.67	53.43	55.78	57.91	59.83	3.87	3.15	2.63	1.94	1.47	1.13	0.89
OrthOST	43.66	46.94	49.02	50.73	53.47	55.81	57.94	59.85	3.85	3.13	2.62	1.93	1.46	1.13	0.88
OST	**43.70**	**46.96**	**49.05**	**50.74**	**53.50**	**55.87**	**58.03**	**60.01**	**3.84**	**3.12**	**2.61**	**1.92**	**1.45**	**1.11**	**0.87**

Table 2. Bit-rate (in bpppb) versus SNR (in dB) and versus MAE of different spectral transforms on hyperspectral images. The bit-rate was computed with the VM9.

of the KLT, JADOST, OrthOST and OST on Identity are respectively 3.6 dB, 3.6 dB, 3.8 dB and 3.8 dB. On three hyperspectral images (see Table 2) the average gains of the KLT, JADOST, OrthOST and OST on Identity are respectively 15.9 dB, 16.3 dB, 16.3 dB and 16.4 dB. Moreover, we can notice that the optimal transforms OrthOST and OST perform always a little better than the KLT at medium and high bit-rates: on six multispectral (resp. three hyperspectral) images the average gains of OrthOST and OST on KLT are about 0.23 dB and 0.28 dB (resp. 0.43 dB and 0.49 dB). On the multispectral images, we observed that JADOST performs roughly as the KLT for MSE distortion, sometimes slightly better, sometimes slightly worse, at any rate. On six images, the average gain of JADOST on the KLT is negligible (about 0.02 dB) at medium and high bit-rates (from 0.25 to 3 bpppb), whereas the average gain of OrthOST on JADOST is about 0.21 dB at the same rates. Nevertheless, on hyperspectral images, JADOST performs slightly but significantly better than the KLT for the four distortions tested at medium and high bit-rates (see Tables 2 and 3) and nearly reaches the OrthOST scores with a significantly lower computational complexity. The average gain of JADOST on the KLT (resp. OrthOST on JADOST) is 0.37 dB (resp. 0.07 dB) on the range [0.25 bpppb , 3 bpppb]. Further, we can remark that there is an insignificant difference of performances between OrthOST and OST. This can be explained by the fact that transforms minimizing the criterion (16) must have a small value for $C_O(\mathbf{A})$, i.e., they must be close to orthogonality (see Remark 5). Therefore there is no advantage to use OST rather than the orthogonal transform OrthOST. In examining the MAD distortion we observe that on the multispectral images tested, at medium bit-rates (i.e. between 0.25 and 1.5 bpppb), OrthOST performs worse than the KLT (see Table 1). On the other hand, on the three hyperspectral (AVIRIS) images tested, for all the distortions measured, at medium and high bit-rates, JADOST and OrthOST perform *always* better than the KLT (see Tables 2 and 3). This is a nice finding, since the optimality of OrthOST is justified only for the MSE distortion and at high bit-rates.

	MSA (°)								MAD							
bit-rate	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00
	Moffett															
Id	12.12	6.82	3.94	2.66	1.29	0.85	0.52	0.36	1676	781	492	1259	183	62	32	20
KLT	1.43	0.87	0.57	0.37	0.20	0.15	0.12	0.10	392	211	119	67	24	14	8	7
JADOST	1.15	0.59	0.42	0.27	0.19	**0.14**	**0.11**	**0.09**	279	120	67	44	18	12	8	**6**
OrthOST	0.96	**0.47**	**0.31**	**0.25**	**0.18**	**0.14**	**0.11**	**0.09**	261	**77**	49	**33**	**18**	**10**	8	**6**
OST	**0.86**	0.50	0.32	**0.25**	**0.18**	**0.14**	**0.11**	**0.09**	**207**	101	**46**	37	19	12	**7**	**6**
	Cuprite															
Id	5.30	2.81	2.20	1.57	1.01	0.59	0.40	0.26	659	360	253	185	110	62	61	40
KLT	0.42	0.25	0.22	0.15	0.12	**0.08**	**0.07**	0.06	154	135	100	54	26	16	10	8
JADOST	0.33	0.25	**0.16**	**0.14**	**0.10**	**0.08**	**0.07**	**0.05**	**112**	**109**	61	39	20	**11**	**9**	**7**
OrthOST	**0.32**	0.25	0.17	**0.14**	**0.10**	**0.08**	**0.07**	**0.05**	113	110	61	**37**	22	**11**	**9**	**7**
OST	0.35	**0.24**	**0.16**	**0.14**	**0.10**	**0.08**	**0.07**	**0.05**	113	**109**	**58**	42	**17**	**11**	**9**	**7**
	Jasper															
Id	18.20	12.53	7.88	5.70	3.87	2.14	1.41	1.01	1907	1220	732	559	241	160	84	55
KLT	0.91	0.53	0.43	0.34	0.26	0.20	0.15	0.12	225	151	82	57	30	15	10	7
JADOST	0.87	**0.51**	0.44	0.33	**0.24**	0.19	0.15	0.12	157	91	56	51	**20**	**11**	9	7
OrthOST	0.83	**0.51**	**0.40**	0.33	**0.24**	0.19	0.15	0.12	157	**84**	**46**	**34**	23	13	9	7
OST	**0.79**	**0.51**	0.41	**0.32**	**0.24**	**0.18**	**0.14**	**0.11**	**156**	86	48	**34**	22	14	**8**	**6**

Table 3. Bit-rate (in bpppb) versus MSA (in degree °) and versus MAD of different spectral transforms on hyperspectral images for the separable scheme. The bit-rate was computed with the VM9.

As already mentioned, the main drawback of the OrthOSTs returned by `JADO` and `OrthOST` algorithms is their heavy computational costs. In the next section we introduce quasi-optimal orthogonal spectral transforms.

8. Performances of exogenous quasi-optimal spectral transforms

8.1 Exogenous quasi-optimal spectral transforms

When one gets a set of images coming from one (and only one) spectrometer sensor, it is possible to compute an exogenous OrthOST from a learning basis extracted from this set. Generally, images from one spectrometer have the same number of bands and the same number of rows. However, the number of columns may vary. To compute an *exogenous* OrthOST, we first split the set of all images in two disconnected sets, one consisting of several images and which becomes the learning basis $\mathcal{L}$, the other constituted of the remaining images and which becomes the test subset. Then, all the images of the learning basis are connected band per band and row per row to construct a single virtual large image $\mathbf{X}$ having the same numbers of bands and rows as any image from the spectrometer and a large number of columns. This image is used as input of the OrthOST algorithm described in (Akam Bita et al., 2010a) and the output is the exogenous OrthOST associated with the learning basis $\mathcal{L}$. The exogenous KLT and exogenous JADOST are calculated similarly.

8.2 Performance comparison between exogenous and non exogenous OrthOSTs

In our tests, we used 10 images[5] (shown in Figure 3) from the imaging spectrometer MERIS on-board the satellite ENVISAT. This fifteen spectral bands spectrometer operates in the solar

[5] The images were acquired via the Data Disseminated System (http://dwlinkdvb.esrin.esa.it/DDS/) thanks to ESA/ESRIN.

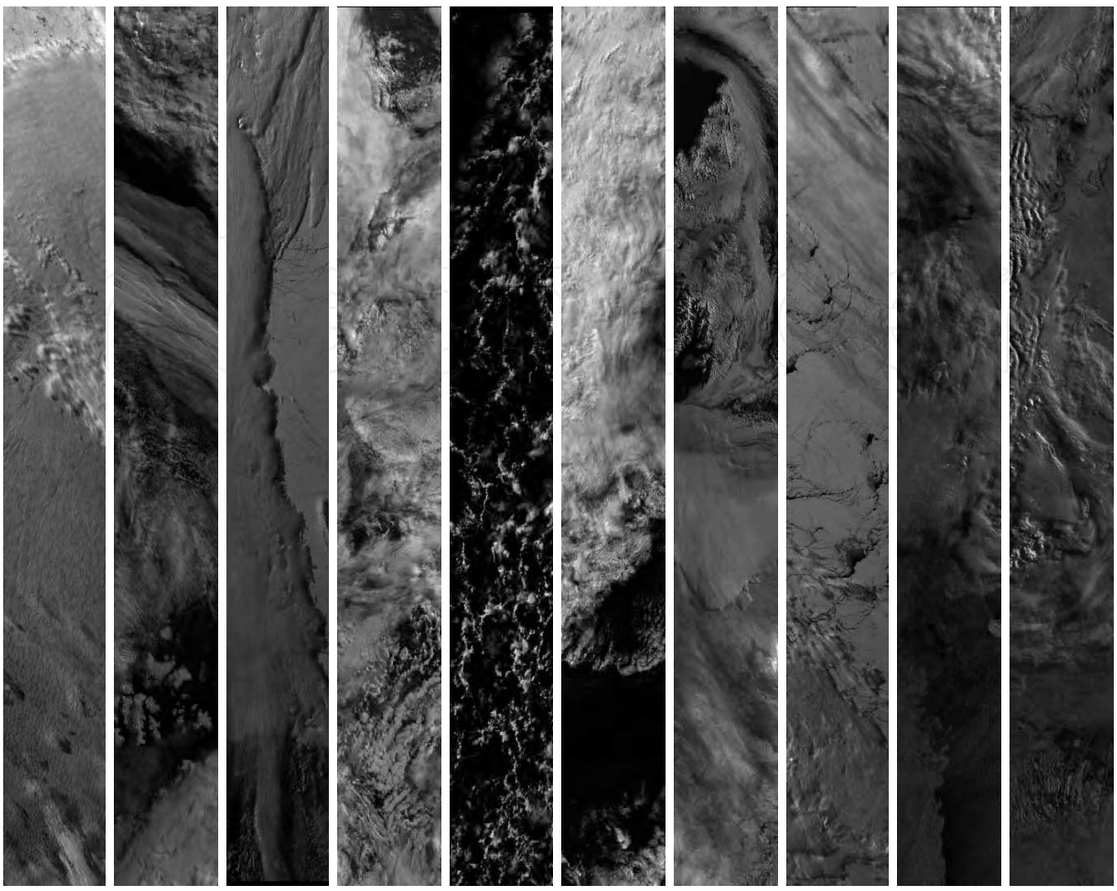

Fig. 3. Fifth component (corresponding approximately to the band [555 nm, 565 nm]) of the hyperspectral images MERIS, numbered 1–4, 6, 8, 10, 13, 15–16, from left to right

reflective spectral range of visible and near infrared light. Each band has a programmable width and a programmable location in the 390 nm to 1040 nm spectral range. As mentioned in (MERIS, 2006) the instrument scans the Earth's surface by the push-broom method, CCD arrays provide spatial sampling in the across-track direction, while the satellite's motion provides scanning in the along-track direction. The scene is imaged simultaneously across the entire spectral range, through a dispersing system, onto the CCD array. Therefore there is no problem of deregistration on the MERIS images. The ten images of our tests have all the same dimensions: $N_r = 128$, $N_c = 1121$ and $N = 15$. They are originally coded on $N_b = 16$ bpppb and they were acquired with the same fifteen spectral bands. To construct exogenous KLT and exogenous OrthOST, we split the ten MERIS images in two disconnected sets, one constituted of seven images (the learning basis), the other constituted of the three remaining images (the test subset). We considered 13 various learning bases, denoted $\mathcal{L}_i$ $(1 \leq i \leq 13)$ which are presented in Table 4. The bit-rate is computed with the Verification Model version 9 (VM9 (JPEG2000, 2001)). Note that the exogenous transforms are fixed (i.e., they do not adapt to the encoded image), hence they are known by the decoder and have not to be transmitted. However, in the lossy compression results given in Table 5 with the VM9, the inverse of the spectral transform is coded in the bit-stream (it costs less than 0.001 bpppb).

	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$	$\mathcal{L}_4$	$\mathcal{L}_5$	$\mathcal{L}_6$	$\mathcal{L}_7$	$\mathcal{L}_8$	$\mathcal{L}_9$	$\mathcal{L}_{10}$	$\mathcal{L}_{11}$	$\mathcal{L}_{12}$	$\mathcal{L}_{13}$
Meris1	×	×		×		×	×	×		×		×	×
Meris2	×	×		×		×		×	×	×	×		×
Meris3	×	×			×	×	×		×	×	×		×
Meris4	×	×	×	×	×		×	×		×	×	×	
Meris6	×		×	×	×	×	×			×	×	×	
Meris8	×		×	×	×			×	×	×		×	×
Meris10	×		×		×	×	×	×	×		×	×	
Meris13		×	×			×	×	×	×	×	×	×	×
Meris15		×	×	×	×	×	×		×		×		×
Meris16		×	×	×	×	×		×	×			×	×

Table 4. Various learning bases, denoted $\mathcal{L}_i$ ($1 \leq i \leq 13$) and constituted of seven MERIS images each

In Table 5 we present the performances obtained with two images when the learning basis varies. Among all the tests we made, we chose to show the best and worst cases obtained with an exogenous OrthOST. For this, the PSNR of a spectral transform is compared to that obtained with the KLT by subtracting, and we considered that difference of PSNR at 1 bpppb. The best and worst cases correspond respectively to the tested images *MERIS2* and *MERIS8*. We can see that for *MERIS2*, the exogenous OrthOST performs significantly better than the KLT at all rates for both MSE and MAE global distortions. Whereas for both MAD and MSA local distortions, exogenous OrthOST and KLT have roughly the same performance, the winner depending on the bit-rate. A more interesting result is the worst case: at bit-rates not greater than 1 bpppb, the worst exogenous OrthOST performs worse than the worst exogenous KLT and this trend is reversed for bit-rates larger than 1.0 dB. Moreover, the loss of PSNR compared to the KLT is 4.3 dB at 1 bpppb, however, the difference of PSNR between the KLT and identity (i.e., no spectral transform) is particularly high here (30 dB). For the other eight tested images, the loss of PSNR of the **worst** exogenous OrthOST with respect to the KLT, is not greater that 2.5 dB, at all bit-rates. Moreover, it is always smaller than the loss of PSNR of the **best** exogenous KLT with respect to the KLT. An example is shown in Fig. 4, where the bit-rate is computed either with the VM9 or the Bit Plane Encoder (BPE) (CCSDS-1, 2007) recommended by the CCSDS (Yeh et al., 2005). In order to compute the bit-rate with the BPE, we proceeded as follows: first, for each transformed component we computed a few hundred points of the graph that links mean square error to bit-rate, then we applied the algorithm by Shoham and Gersho (Shoham & Gersho, 1988) to optimally allocate distortions between components for given maximal total bit-rates.
In average on the 10 images, the loss of PSNR of the **worst** exogenous OrthOST with respect to the KLT is significantly smaller than the one of the **best** exogenous KLT (see Table 6). It is no longer the case for exogenous JADOST. We observed the importance of the learning basis, whose influence can range from 0 dB to -4 dB. In other words, when the learning basis is well chosen (depending on the scene and not only on the spectrometer), one can expect a loss of PSNR of an exogenous OrthOST with respect to the KLT not greater than 0.4 dB. Whereas, when it is badly chosen, the same loss of PSNR should be limited to 4 dB. However, these values are only indicative and should not be considered definitive, because they were obtained on a set of 10 MERIS images, which was not proven to be statistically significant.
We observed good performances of exogenous OrthOST used with the VM9 and in (Akam Bita et al., 2010c) the authors observed that, associated with the BPE and the optimal bit allocation

	PSNR (dB)							MAD						
bit-rate	0.25	0.50	0.75	1.00	1.50	2.00	3.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00
	MERIS2													
Id	39.40	42.13	44.20	45.99	49.22	52.26	58.26	5878	3271	2369	2226	1258	850	443
KLT	54.87	63.81	69.60	73.66	79.29	83.02	88.39	1407	564	229	140	65	27	13
JADOST	55.65	65.04	70.53	74.49	80.02	83.66	88.92	1309	465	197	134	56	26	13
OrthOST	55.65	65.41	71.15	75.23	80.66	84.17	89.33	1354	397	189	119	48	22	12
exo3_KLT	54.11	62.36	67.79	*71.85*	77.77	81.86	87.53	1392	511	256	159	72	37	16
exo3_JADOST	55.11	64.01	69.53	73.46	78.82	82.78	88.23	1313	434	228	121	191	37	15
exo3_OrthOST	55.35	64.48	70.15	**74.18**	79.71	83.40	88.72	1267	476	240	134	53	28	13
exo5_KLT	54.08	62.37	67.89	**72.04**	78.05	82.11	87.69	1326	536	325	154	68	41	15
exo5_JADOST	55.12	64.01	69.54	73.48	79.08	82.88	88.30	1289	398	228	**117**	58	28	13
exo5_OrthOST	55.36	64.46	70.11	*74.13*	79.70	83.42	88.74	1314	497	257	134	49	26	12
exo7_KLT	54.69	62.92	68.09	72.00	77.79	81.84	87.52	1226	602	265	*183*	66	43	16
exo7_JADOST	55.22	63.97	69.14	*72.88*	78.32	82.26	87.86	1236	524	244	*165*	64	39	14
exo7_OrthOST	55.61	64.94	70.32	**74.18**	79.57	83.25	88.60	1418	504	238	**128**	56	28	13
exo12_KLT	54.13	62.37	67.83	71.95	77.99	82.07	87.66	1432	513	312	**152**	68	42	15
exo12_JADOST	55.13	64.02	69.54	**73.50**	79.08	82.87	88.30	1289	435	258	136	52	30	14
exo12_OrthOST	55.35	64.48	70.14	74.15	79.72	83.42	88.73	1305	481	221	*139*	47	30	13
	MERIS8													
Id	35.22	38.02	40.17	42.11	45.67	49.06	55.71	9723	5652	4123	3068	2050	1313	626
KLT	53.85	62.69	68.09	72.11	77.71	81.48	87.18	2638	1192	411	228	87	37	17
JADOST	54.21	63.68	69.39	73.44	78.83	82.35	87.91	2838	709	262	148	50	31	15
OrthOST	54.2	63.86	69.61	73.7	79.07	82.58	88.12	2773	705	305	130	51	26	15
exo2_KLT	52.22	59.49	64.16	68	73.95	78.34	84.79	3501	1504	608	*400*	149	51	24
exo2_JADOST	51.19	58.64	63.66	**67.83**	73.93	78.83	85.2	3536	1608	646	321	501	59	21
exo2_OrthOST	50.54	58.19	63.46	*67.8*	74.39	78.95	85.29	3446	1471	592	326	133	46	20
exo6_KLT	52.62	59.99	64.57	**68.16**	73.81	78.24	84.77	2813	973	471	**255**	115	59	23
exo6_JADOST	52.3	59.46	63.86	*67.4*	73.16	77.43	84.44	2670	1077	547	312	138	229	23
exo6_OrthOST	51.57	59.7	64.95	**68.96**	75.01	79.36	85.6	2871	1195	482	**290**	91	51	20
exo7_KLT	52.49	59.68	64.18	*67.83*	73.63	78.06	84.59	2887	1013	469	287	126	70	23
exo7_JADOST	52.28	59.45	63.85	67.44	73.31	77.92	84.6	2689	980	543	*328*	135	70	22
exo7_OrthOST	51.71	59.29	64.16	68.03	74.14	78.69	85.11	2768	1147	451	326	115	51	22
exo11_KLT	52.56	59.87	64.43	68.06	73.76	78.16	84.67	2885	876	459	**255**	104	68	24
exo11_JADOST	52.28	59.49	63.94	67.6	73.49	78.08	84.7	2610	961	525	**290**	126	67	23
exo11_OrthOST	51.89	59.59	64.64	68.73	75.01	79.44	85.68	2909	1130	441	*346*	131	47	19
	MSA (in $^{\circ}$)							MAE						
	MERIS2													
Id	34.94	24.67	19.57	14.91	10.70	8.05	4.37	503.5	377.8	302.7	248.5	173.7	123.3	62.17
KLT	8.63	2.91	1.62	1.36	0.42	0.23	0.14	87.50	31.02	15.88	10.03	5.41	3.58	1.93
JADOST	10.41	2.66	1.54	0.83	0.38	0.22	0.11	80.46	27.18	14.45	9.22	4.99	3.33	1.82
OrthOST	9.47	2.55	1.57	1.07	0.44	0.21	0.11	80.22	25.87	13.43	8.48	4.65	3.15	1.73
exo3_KLT	10.26	4.19	2.05	**1.16**	0.54	0.28	0.13	95.35	36.85	19.79	*12.44*	6.43	4.09	2.13
exo3_JADOST	8.42	3.33	1.75	1.19	0.51	0.30	0.13	85.27	30.56	16.17	10.35	5.66	3.69	1.97
exo3_OrthOST	9.07	3.27	1.63	1.10	0.38	0.25	0.11	83.01	28.91	15.06	9.52	5.17	3.44	1.86
exo5_KLT	10.12	3.89	2.06	*1.38*	0.54	0.30	0.13	95.60	36.72	19.50	**12.18**	6.24	3.98	2.10
exo5_JADOST	8.61	3.59	1.96	**1.00**	0.49	0.28	0.12	85.21	30.49	16.12	10.31	5.55	3.64	1.95
exo5_OrthOST	9.03	3.25	1.49	1.08	0.41	0.24	0.12	82.71	28.94	15.10	*9.56*	5.18	3.43	1.86
exo7_KLT	8.59	3.79	1.89	1.37	0.53	0.32	0.13	89.55	34.46	19.04	12.20	6.40	4.10	2.14
exo7_JADOST	9.72	2.95	1.67	*1.29*	0.47	0.27	0.13	84.59	30.68	16.86	*11.01*	6.03	3.91	2.06
exo7_OrthOST	10.33	2.78	1.56	**0.90**	0.43	0.24	0.12	80.62	27.32	14.70	**9.49**	5.24	3.49	1.89
exo12_KLT	9.41	4.17	2.46	1.23	0.54	0.29	0.13	95.20	36.77	19.67	12.30	6.28	3.99	2.10
exo12_JADOST	8.45	3.23	1.89	1.25	0.45	0.26	0.13	85.00	30.51	16.15	**10.30**	5.55	3.65	1.95
exo12_OrthOST	8.97	3.20	1.57	*1.25*	0.39	0.26	0.12	82.91	28.88	15.07	9.54	5.17	3.43	1.86
	MERIS8													
Id	35.69	29.24	28.75	19.75	15.56	10.93	6.22	796.1	584.5	461.4	373.3	250.6	170.83	80.01
KLT	8.34	2.81	1.54	1.02	0.5	0.28	0.14	95.77	33.84	18.52	11.93	6.47	4.25	2.22
JADOST	8.39	2.28	1.29	0.74	0.36	0.24	0.13	92.49	30.45	16.06	10.28	5.71	3.86	2.04
OrthOST	7.71	2.96	1.25	0.77	0.36	0.25	0.11	92.67	29.82	15.68	10.02	5.56	3.76	1.99
exo2_KLT	8.2	3.92	2.53	*1.5*	0.63	0.34	0.17	113.61	46.49	27.87	18.44	9.75	6.04	2.93
exo2_JADOST	10.53	4.81	3.14	**1.44**	0.68	0.39	0.17	128.47	52.80	29.90	**18.97**	9.61	5.71	2.79
exo2_OrthOST	12.27	4.97	2.61	1.53	0.81	0.34	0.16	138.68	56.09	30.87	*19.22*	9.33	5.65	2.76
exo6_KLT	8.31	4.16	1.84	**1.29**	0.71	0.41	0.17	109.55	44.79	26.67	**18.02**	9.84	6.09	2.94
exo6_JADOST	8.75	4.22	2.17	1.46	0.85	0.45	0.18	113.05	46.89	28.35	19.25	10.47	6.55	3.05
exo6_OrthOST	10.18	3.47	1.87	**1.24**	0.55	0.35	0.17	123.22	47.09	25.79	16.51	8.67	5.39	2.67
exo7_KLT	8.84	3.97	2.06	1.35	0.71	0.46	0.19	110.39	45.8	27.74	*18.67*	10.06	6.21	3
exo7_JADOST	8.10	3.58	2.00	*1.76*	0.71	0.45	0.17	113.27	46.72	28.58	*19.26*	10.32	6.30	3.00
exo7_OrthOST	9.89	4.05	2.28	*1.76*	0.68	0.38	0.19	120.79	48.94	28.22	18.41	9.52	5.8	2.82
exo11_KLT	8.18	4.12	1.95	1.3	0.68	0.46	0.18	110.01	45.14	27.14	18.32	9.93	6.15	2.97
exo11_JADOST	8.93	3.42	2.07	1.69	0.70	0.38	0.19	113.56	46.76	28.29	**18.97**	10.13	6.20	2.96
exo11_OrthOST	9.8	4.25	1.97	1.56	0.6	0.33	0.15	118.26	47.54	27.19	**17.15**	8.69	5.34	2.64

Table 5. Bit-rate (in bpppb) vs PSNR (in dB), vs MAD, vs MSA and vs MAE of various spectral transforms on two images. The exo$i$_KLT, exo$i$_JADOST and exo$i$_OrthOST correspond respectively to exogenous KLT, JADOST and OrthOST computed with the learning basis $\mathcal{L}_i$. The best (resp. worst) results of exogenous transforms at 1.0 bpppb are bolded (resp. in italics).

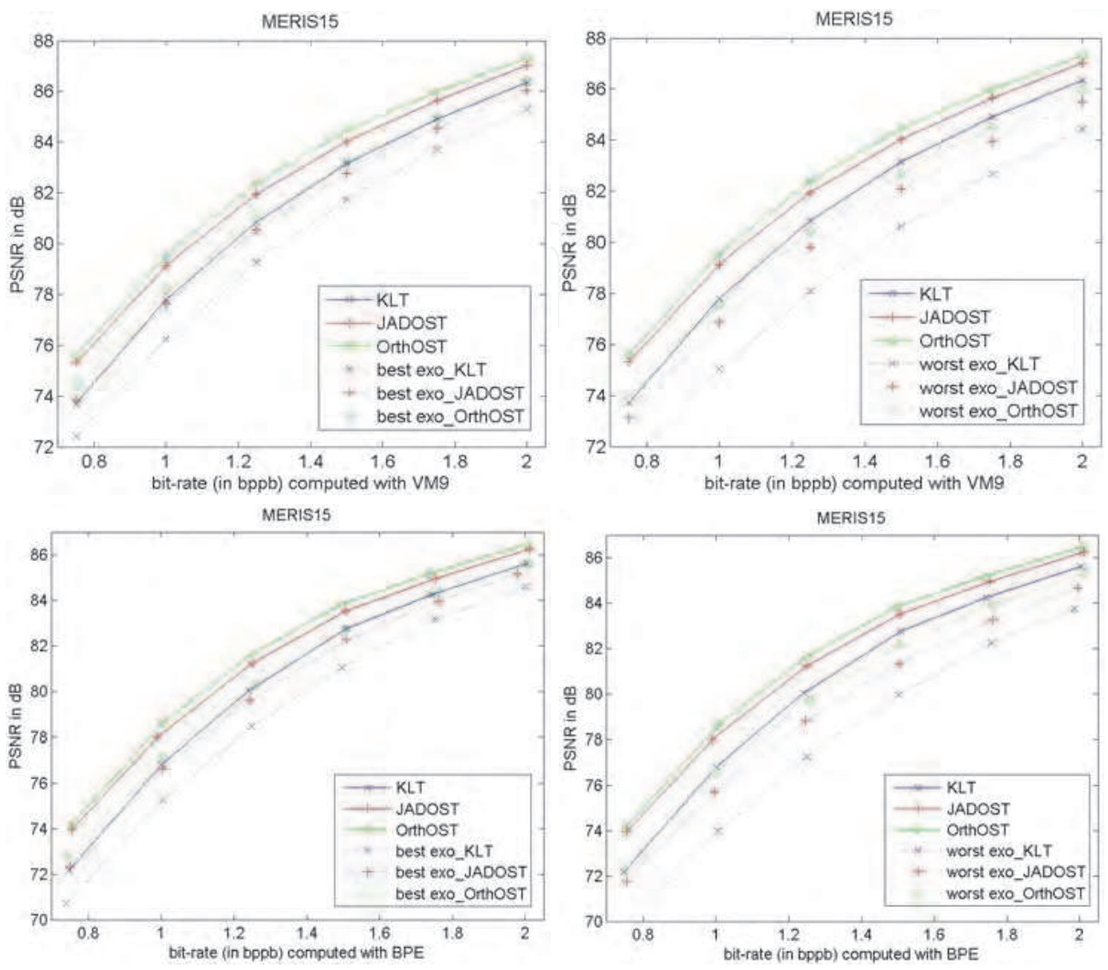

Fig. 4. PSNR (in dB) versus bit-rate (in bpppb) for various spectral transforms (KLT, JADOST OrthOST and (left) best exogenous KLT, best exogenous JADOST, best exogenous OrthOST or (right) worst exogenous KLT, worst exogenous JADOST, worst exogenous OrthOST). The image is *MERIS15* and the bit-rate is computed with first row: the VM9, second row: the BPE.

bit-rate (in bpppb)	0.25	0.50	0.75	1.00	1.50	2.00	3.00
mean (in dB) {PSNR(KLT)− worst exogenous PSNR(OrthOST)}	0.67	1.19	1.56	1.68	1.49	1.15	0.86
mean (in dB) {PSNR(KLT) − worst exogenous PSNR(JADOST)}	0.56	1.32	1.99	2.29	2.25	1.89	1.40
mean (in dB) {PSNR(KLT) − best exogenous PSNR(KLT)}	1.02	1.83	2.34	2.51	2.27	1.82	1.4

Table 6. Comparison of the averaged losses of PSNR with respect to the KLT for the worst exogenous OrthOST, the worst exogenous JADOST and the best exogenous KLT. The worst and best exogenous transforms are selected at 1.00 bpppb. The averages are computed on the ten images.

algorithm by Shoham and Gersho (Shoham & Gersho, 1988) for quantization and entropy coding, exogenous OrthOST still performs well (see Fig. 4). However, the VM9 and the Shoham and Gersho algorithm both have a too high computational complexity for a coder on-board a satellite. In (Gutzwiller et al., 2009), the authors propose an extension to multicomponent images of the well-known 2-D SPIHT encoder that has not the shortcoming of a high computational cost for bit-rate allocation.

9. Conclusion

In this chapter, we have studied the problem of finding optimal spectral transforms associated with fixed 2D discrete wavelet transforms in coding of multi- and hyper-spectral images, for a compression scheme that is compatible with the JPEG2000 Part 2 standard. We clarified the criterion that gives, when minimized, an optimal transform under high-rate entropy constraint scalar quantization hypothesis and when one scalar quantizer per subband and per component is applied. We showed the link between the criterion and the mutual information contrast used in independent component analysis. We derived a criterion minimized by an orthogonal optimal transform when the data are Gaussian. Then we gave three algorithms that return the spectral transforms that minimize the JPEG2000 compatible criterion, two under the constraint of orthogonality — one of which assuming Gaussian data — and the third with no constraint, but invertibility. Finally, we have tested the optimal transforms on satellite multi- and hyper-spectral images and found that for hyperspectral images the orthogonal optimal transform OrthOST and JADOST performs a little better than the KLT for four distortion measures that permit to evaluate the performances of the codec in applications of hyperspectral images like classifications or target detections. However the computational complexity of the optimal transform is too heavy for actual applications. Last we have presented the exogenous orthogonal quasi-optimal spectral transforms, that have a significantly smaller complexity, and their performances in lossy coding. In future works, we will study the problem of designing optimal spectral filters (i.e. a convolutive rather than an instantaneous mixture) in lossy compression of multi- and hyper-spectral images.

10. References

Akam Bita, I. P., Barret, M., Dalla Vedova, F. & Gutzwiller, J.-L. (2008). Onboard compression of hyperspectral images using an exogenous orthogonal quasi-optimal transform, *Proceedings of On-Board Payload Data Compression Workshop*, Noordwijk (The Netherlands).

Akam Bita, I. P., Barret, M. & Pham, D.-T. (2010a). On optimal transforms in lossy compression of multicomponent images with JPEG2000, *Signal Processing* Vol. 90(No. 3): 759–773.

Akam Bita, I. P., Barret, M. & Pham, D.-T. (2010b). On optimal orthogonal transforms at high bit-rates using only second order statistics in multicomponent image coding with JPEG2000, *Signal Processing* Vol. 90(No. 3): 753–758.

Akam Bita, I. P., Barret, M., Vedova, F. D. & Gutzwiller, J.-L. (2010c). Lossy and lossless compression of MERIS hyperspectral images with exogenous quasi-optimal spectral transforms, *Journal of Applied Remote Sensing* Vol. 4: 1–15.

Antonini, M., Barlaud, M., Mathieu, P. & Daubechies, I. (1992). Image coding using wavelet transform, *IEEE Transactions on Image Processing* Vol. 1(No. 2): 205–220.

Barret, M., Akam Bita, I. P., Gutzwiller, J.-L. & Dalla Vedova, F. (2009). Lossy hyperspectral image coding with exogenous quasi optimal transforms, *Proceedings Data Compression Conference*, Snowbird (USA), pp. 411–419.

Barret, M., Gutzwiller, J.-L. & Hariti, M. (2011). Low complexity hyperspectral image coding using exogenous orthogonal optimal spectral transform (OrthOST) and degree-2 zerotrees, *IEEE Transactions on Geoscience and Remote Sensing* Vol. 49(No. 5).

CCSDS-1 (2007). Image data compression, *Report concerning space data system standards, CCSDS 120.1-G-1, Green book* .

Chang, C.-I., Ramakrishna, B., Wang, J. & Plaza, A. (2010). Low-bit rate exploitation-based lossy hyperspectral image compression, *Journal of Applied Remote Sensing* Vol. 4: 1–24.

Christophe, E., Léger, D. & Mailhes, C. (2005). Quality criteria benchmark for hyperspectral imagery, *IEEE Transactions on Geoscience and Remote Sensing* Vol. 43(No. 9): 2103–2114.

Christophe, E., Mailhes, C. & Duhamel, P. (2006). Best anisotropic 3-D wavelet decomposition in a rate-distortion sense, *Proceedings of International Conference on Acoustic, Speech and Signal Processing*, Toulouse (France), pp. II–17–20.

Dragotti, P. L., Poggi, G. & Ragozini, A. R. P. (2000). Compression of multispectral images by three-dimensional SPIHT algorithm, *IEEE Transactions on Geoscience and Remote Sensing* Vol. 38(No. 1): 416–428.

Du, Q. & Fowler, J. E. (2007). Hyperspectral image compression using JPEG2000 and principal component analysis, *IEEE Geoscience and Remote Sensing Letters* Vol. 4: 201–205.

Du, Q. & Fowler, J. E. (2008). Low-compexity principal component analysis for hyperspectral image compression, *International Journal of High Performance Computing Applications* Vol. 22: 438–448.

Flury, B. N. & Gautschi, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form, *SIAM Journal on Scientific and Statistical Computing* Vol. 7(No. 1): 169–184.

Fowler, J. E. & Rucker, J. T. (2007). chap 14: 3D wavelet-based compression of hyperspectral imagery, *in* C.-I. Chang (ed.), *Hyperspectral Data Exploitation: Theory and Applications*, John Wiley & Sons, Hoboken.

Gersho, A. & Gray, R. M. (1992). *Vector quantization and signal compression*, Kluwer Academic Publisher.

Gray, R. M. & Neuhoff, D. L. (1998). Quantization, *IEEE Transactions on Information Theory* Vol. 44(No. 6): 2325–2384.

Gutzwiller, J.-L.and Hariti, M., Barret, M., Christophe, E., Thiebaut, C. & Duhamel, P. (2009). Extension du codeur SPIHT au codage d'images hyperspectrales, *Proceedings of Colloquium COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, Toulouse (France).
URL: *http://liris.cnrs.fr/Documents/Liris-4087.pdf*

JPEG2000 (2001). JPEG2000 verification model 9.1 (technical description), *ISO/IEC JTC 1/SC 29/WG 1 WG1 N2165* pp. 1–213.

MERIS (2006). MERIS detailed instrument description, *European Space Agency* .
URL: *http://envisat.esa.int/instruments/meris/*

Narozny, M., Barret, M., Pham, D.-T. & Akam Bita, I. P. (2005). Modified ICA algorithms for finding optimal transforms in transform coding, *Proceedings of IEEE International Symposium on Image and Signal Processing and Analysis*, Zagreb (Croatie), pp. 111–116.

Narozny, M., Barret, M. & Pham, D.-T. (2008). ICA based algorithms for computing optimal 1-D linear block transforms in variable high-rate source coding, *Signal Processing* Vol. 88(No. 2): 268–283.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill.

Penna, B., Tillo, T., Magli, E. & Olmo, G. (2007). Transform coding techniques for lossy hyperspectral data compression, *IEEE Transactions on Geoscience and Remote Sensing* Vol. 45(No. 5): 1408–1421.

Pham, D.-T. (2004). Fast algorithms for mutual information based independent component analysis, *IEEE Transactions on Signal Processing* Vol. 52(No. 10): 2690–2700.

Pham, D.-T. (2005). Entropy of random variable slightly contaminated with another, *IEEE Signal Processing Letters* Vol. 12(No. 7): 536–539.

Rucker, J. T., Fowler, J. E. & Younan, N. H. (2005). JPEG2000 coding strategies for hyperspectral data, *Proceedings of International Geoscience and Remote Sensing Symposium*, Seoul (Korea), pp. 128–131.

Said, A. & Pearlman, W. A. (1996). A new, fast and efficient image codec based on set partitioning in hierarchical trees, *IEEE Transactions on Circuits and Systems for Video Technology* Vol. 6(No. 3): 243–250.

Shoham, Y. & Gersho, A. (1988). Efficient bit allocation for an arbitrary set of quantizers, *IEEE Transactions on Acoustics, Speech, and Signal Processing* Vol. 36(No. 9): 1445–1453.

Taubman, D. S. (2000). High performance scalable compression with EBCOT, *IEEE Transactions on Image Processing* Vol. 9(No. 7): 1158–1170.

Taubman, D. S. & Marcellin, M. W. (2002). *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic, Place of publication.

Thiebaut, C., Lebedeff, D., Latry, C. & Bobichon, Y. (2006). On-board compression algorithm for satellite multispectral images, *Proceedings of Data Compression Conference*, Snowbird (USA), pp. 28–30.

Usevitch, B. (1996). Optimal bit allocation for biorthogonal wavelet coding, *Proceedings of Data Compression Conference*, Snowbird (USA), pp. 387–395.

Vaisey, J., Barlaud, M. & Antonini, M. (1998). Multispectral image coding using lattice VQ and the wavelet transform, *Proceedings of IEEE International Conference on Image Processing*, Chicago (USA), pp. 307–311.

Woods, J. W. & Naven, T. (1992). A filter based bit allocation scheme for subband compression of HDTV, *IEEE Transactions on Image Processing* Vol. 1(No. 3): 436–440.

Yeh, P. S., Armbruster, P., Kiely, A., Masschelein, B., Moury, G. & Schafer, C. (2005). The new CCSDS image compression recommendation, *Proceedings of IEEE Aerospace Conference*, Big Sky (USA), pp. 1–8.

Discrete Wavelet Transforms - Algorithms and Applications

Edited by Prof. Hannu Olkkonen

ISBN 978-953-307-482-5
Hard cover, 296 pages
Publisher InTech
Published online 29, August, 2011
Published in print edition August, 2011

The discrete wavelet transform (DWT) algorithms have a firm position in processing of signals in several areas of research and industry. As DWT provides both octave-scale frequency and spatial timing of the analyzed signal, it is constantly used to solve and treat more and more advanced problems. The present book: Discrete Wavelet Transforms: Algorithms and Applications reviews the recent progress in discrete wavelet transform algorithms and applications. The book covers a wide range of methods (e.g. lifting, shift invariance, multi-scale analysis) for constructing DWTs. The book chapters are organized into four major parts. Part I describes the progress in hardware implementations of the DWT algorithms. Applications include multitone modulation for ADSL and equalization techniques, a scalable architecture for FPGA-implementation, lifting based algorithm for VLSI implementation, comparison between DWT and FFT based OFDM and modified SPIHT codec. Part II addresses image processing algorithms such as multiresolution approach for edge detection, low bit rate image compression, low complexity implementation of CQF wavelets and compression of multi-component images. Part III focuses watermaking DWT algorithms. Finally, Part IV describes shift invariant DWTs, DC lossless property, DWT based analysis and estimation of colored noise and an application of the wavelet Galerkin method. The chapters of the present book consist of both tutorial and highly advanced material. Therefore, the book is intended to be a reference text for graduate students and researchers to obtain state-of-the-art knowledge on specific applications.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China

www.intechopen.com

51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen

IntechOpen