WAVELET PACKETS FOR TIME-FREQUENCY ANALYSIS OF MULTISPECTRAL IMAGERY

JOHN J. BENEDETTO, WOJCIECH CZAJA, AND MARTIN EHLER

Abstract. Multispectral geospatial image sets retain the scene's spatial and spectral information. To jointly use both of them for analysis purposes, we propose to extend the concept of wavelet packets, by introducing a new integrated multispectral entropy function. Each spectral band is individually decomposed by the wavelet packets transform, and then the entropy term is jointly guided by information from all bands, simultaneously. Finally, the wavelet packets coefficients undergo a dimension reduction process. We present examples of this theory applied to hyperspectral satellite imagery.

1. Introduction

In geosciences, remote sensing is understood as collecting data about objects without coming in direct contact with them. Typically, the setting for remote sensing is the Earth's surface and atmosphere and the concept is often associated with gathering information dealing with various aspects of the classification and detection problems associated with both man-made and natural issues. Remote sensing technology is phenomenally varied and it includes many different modalities. In this paper we focus on multi- and hyperspectral remote sensing imagery. Multispectral imaging is nowadays successfully used to detect, classify, and quantify objects and materials in a wide range of subjects, ranging from geophysics to biomedicine. The multispectral image sets contain both spatial and spectral information of the scene. This leads to a multitude of novel and effective methods to make use of spatial and spectral components in a joint fashion. For an example of spatial-spectral fusion in hyperspectral satellite imagery we refer the interested reader, e.g., to [5] and references therein.

The purpose of this paper is to provide a very different point of view on the spatial-spectral fusion problem, that is inspired by recent developments at the intersection of harmonic analysis and wavelet theory with machine learning and nonlinear dimension reduction. Novel dimension reduction techniques help us to overcome the ever increasing complexity of hyperspectral data, due to the presence of multiple spectral bands, as well as improving pixel resolution of the imagery. Principal component analysis (PCA) has been widely used in this regard, and the few significant principal components are typically used to induce a linear transformation maximizing the captured variance within the dimension reduced data. If the data set lies on a linear manifold, then PCA optimally recovers this simple manifold structure. In recent years, however, arguments have been raised that indicate that data acquired from natural sources typically does not reside on a linear manifold. This led to replacement of PCA with more sophisticated methods. These include, among others, diffusion maps [14], multiscale approaches [13, 27], and nonlinear dimension reduction [4, 6, 37, 38, 42]. Wavelet dimension reduction

of hyperspectral image data has also been applied, see, e.g., [1]. Wavelets and PCA were jointly used in [29], which motivated the present work. Further references to the applications of wavelets in multi- and hyperspectral imagery include [8, 30, 35, 39, 40, 41, 45] and references therein. Despite its shortcomings, we do note that PCA's simplicity and robustness to noise are behind its widespread use and, thus, provide a good argument to incorporate PCA as a subroutine in more comprehensive analysis schemes.

Wavelet methods and the underlying multiscale decompositions received significant attention from both mathematics and applied sciences. Because of the adaptability of the multiresolution scheme, it has proven to be well-suited for the analysis of spatial characteristics, such as those typically observed in remote sensing imagery, see, e.g., [2, 24, 31]. The associated Discrete Wavelet Transform (DWT) is an iterative scheme that splits the signal into approximation and detail coefficients. Each level is computed by passing through only the previous approximation coefficients. The Wavelet Packets Transform (WPT), on the other hand, further decomposes both the approximation and the detail parts [16]. Contrary to the DWT, this yields a full wavelet tree decomposition that offers more flexibility. The best basis algorithm of Coifman and Wickerhauser finds the optimal subtree for reconstruction, i.e., the best coefficient set with respect to a chosen entropy measure.

To jointly use spatial and spectral information in multispectral image sets, we propose an extension of the classical entropy measure within the concept of wavelet packets. The wavelet characteristics efficiently capture spatial correlations, and thus, by using an extended entropy term, we are also able to utilize the spectral information at the same time. We decompose each spectral band separately by using the WPT. Spatial information can be transferred from one band into the wavelet decomposition of another by using a new joint entropy function. Hence, spectral and spatial components are intertwined in the wavelet packets coefficients. The joint entropy preserves correlations between frequency bands in the wavelet domain. Wavelet coefficients are thresholded to remove noise and artifacts. The best joint subtree is determined by minimizing the joint entropy over all wavelet coefficients and across all spectral bands. Each node of the joint subtree is now a coefficient vector whose entries refer to different spectral bands. Next, PCA is used in the wavelet domain to reduce the data dimension. Pseudo-spectral images are reconstructed by applying the inverse WPT. We verify that this proposed scheme for fusion of spectral and spatial information significantly improves the output of classification schemes in multispectral satellite imaging, by applying it to AVIRIS, ROSIS, and HYDICE data.

The outline of the paper is as follows. In Section 2 we introduce wavelet packets with the underlying WPT and we specify the concept of the joint entropy for multispectral image sets. We apply our proposed wavelet packets scheme to analyze hyperspectral satellite images in Section 3. Conclusions are given in Section 4.

2. From wavelets to wavelet packets

Wavelet packets were introduced by Coifman, Meyer, and Wickerhauser in the early nineties [16, 12]. The underlying transform is a descendant of the Discrete Wavelet Transform, but it provides an enormous flexibility, which stems from the redundancy associated with analyzing the complete impulse response tree. To exploit this flexibility, appropriate entropy functions need to be designed and proper

filters must be selected. As such, we start this Section with a description of the basics of biorthogonal multiresolution analyses, and then proceed to explain the construction of the resulting wavelet packets scheme with joint entropy and dimension reduction.

2.1. **Biorthogonal wavelets.** Let M denote an *expansive dilation matrix*, understood, for example, as an integer matrix whose eigenvalues are greater than one in modulus. We refer the interested reader to [10] for other ways in which expansive dilations can be defined. Given such M and $f : \mathbb{R}^d \to \mathbb{C}$, let

$$f_{j,k}(x) := m^{\frac{j}{2}} f(M^j x - k), \quad \text{for } j \in \mathbb{Z}, k \in \mathbb{Z}^d,$$

denote a *wavelet system*, with $m := |\det(M)|$. A collection $\psi^{(1)}, \ldots, \psi^{(n)}$ of $L_2(\mathbb{R}^d)$-functions is called a *wavelet Riesz basis* if $\{\psi^{(\mu)}_{j,k} : j \in \mathbb{Z},\ k \in \mathbb{Z}^d,\ \mu = 1 \ldots, n\}$ is a Riesz basis for $L_2(\mathbb{R}^d)$. Mallat and Meyer proposed the concept of multiresolution analysis, building the foundation of fast wavelet algorithms: an increasing sequence of closed subspaces $(V_j)_{j\in\mathbb{Z}}$ in $L_2(\mathbb{R}^d)$ is called a *multiresolution analysis* if the following hold:

- **(M-1):** $f \in V_j$ if and only if $f(M^{-j}\cdot) \in V_0$, for all $j \in \mathbb{Z}$,
- **(M-2):** $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L_2(\mathbb{R}^d)$,
- **(M-3):** $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$,
- **(M-4):** there is a function $\varphi \in V_0$, called the *generator*, whose integer shifts constitute a Riesz basis for V_0.

Let $(V_j)_{j\in\mathbb{Z}}$ and $(\widetilde{V}_j)_{j\in\mathbb{Z}}$ be two multiresolution analyses with compactly supported generators φ and $\widetilde{\varphi}$, respectively. In addition, we assume that their integer shifts are biorthogonal to each other, i.e., for all $k, k' \in \mathbb{Z}^d$, we require that

$$\langle \varphi(\cdot - k), \widetilde{\varphi}(\cdot - k') \rangle = \delta_{k,k'}. \tag{1}$$

Moreover, let us assume that $\{\psi^{(\mu)}_{j,k} : k \in \mathbb{Z}^d,\ \mu = 1 \ldots, n\}$ and $\{\widetilde{\psi}^{(\mu)}_{j,k} : k \in \mathbb{Z}^d,\ \mu = 1 \ldots, n\}$ constitute Riesz bases for spaces W_j and $\widetilde{W}_j$, respectively, such that

$$V_{j+1} = V_j \oplus W_j, \quad \widetilde{V}_{j+1} = \widetilde{V}_j \oplus \widetilde{W}_j, \quad \text{and} \quad \langle \psi^{(\mu)}_{j,k}, \widetilde{\psi}^{(\mu')}_{j',k'} \rangle = \delta_{j,j'}\delta_{k,k'}\delta_{\mu,\mu'}. \tag{2}$$

We note here that W_0 and $\widetilde{W}_0$ are algebraic complements and not necessarily orthogonal to V_0 and $\widetilde{V}_0$, respectively, but are assumed to be related by the following relationship:

$$W_j \perp \widetilde{V}_j, \quad \widetilde{W}_j \perp V_j. \tag{3}$$

According to the theory of shift invariant spaces, the number of wavelets is determined as $n = m - 1$, cf., [44]. Condition (M-3) and the relations (2) yield decompositions of the signal space up to level j. Indeed, we have

$$V_j = V_{j_0} \oplus \bigoplus_{j'=j_0}^{j-1} W_{j'} = \bigoplus_{j'=-\infty}^{j-1} W_{j'} \qquad \text{and} \qquad \widetilde{V}_j = \widetilde{V}_{j_0} \oplus \bigoplus_{j'=j_0}^{j-1} \widetilde{W}_{j'} = \bigoplus_{j'=-\infty}^{j-1} \widetilde{W}_{j'},$$

where $j_0 < 0$ is some fixed integer. For simplicity, suppose that f is contained in V_0. Then, there exists the expansion,

$$f = \sum_{k\in\mathbb{Z}^d} \langle f, \widetilde{\varphi}_{0,k} \rangle \, \varphi_{0,k}. \tag{4}$$

Since $V_0 = V_{j_0} \oplus \bigoplus_{j=j_0}^{-1} W_j$, the biorthogonality relations (3) yield

$$f = \sum_{k\in\mathbb{Z}^d} \langle f, \widetilde{\varphi}_{j_0,k}\rangle\, \varphi_{j_0,k} + \sum_{j=j_0}^{-1}\sum_{\mu=1}^{m-1}\sum_{k\in\mathbb{Z}^d} \langle f, \widetilde{\psi}_{j,k}^{(\mu)}\rangle \psi_{j,k}^{(\mu)}. \tag{5}$$

The fast wavelet transform enables us to switch between the two representations (4) and (5) by computing the coefficients of one from the other. We shall now illustrate this fact, after some necessary preparations. According to (M-1) and (M-4), the collection,

$$\left\{\varphi_{j,k} : k \in \mathbb{Z}^d\right\},$$

is a Riesz basis for V_j. Since the spaces V_j are increasing, φ is contained in V_1. Thus, there exists a sequence $(a_k)_{k\in\mathbb{Z}^d} \in \ell_2(\mathbb{Z}^d)$ such that φ satisfies the *refinement equation*,

$$\varphi(x) = \sum_{k\in\mathbb{Z}} a_k^{(0)} \varphi(Mx - k).$$

For simplicity, we assume that the sequence $(a_k^{(0)})_{k\in\mathbb{Z}^d}$ has only finitely many nonzero entries. This implies that

$$a^{(0)}(\xi) = \frac{1}{m}\sum_{k\in\mathbb{Z}^d} a_k^{(0)} e^{-2\pi i k\cdot\xi}$$

is a trigonometric polynomial, called the *symbol* of φ. The biorthogonality relation (1) is equivalent to

$$\sum_{\gamma\in\Gamma_M} a^{(0)}(\xi+\gamma)\overline{b^{(0)}(\xi+\gamma)} = 1, \quad \text{for all } \xi \in \mathbb{R}^d, \tag{6}$$

where Γ_M is a complete set of representatives of $M^{-\top}\mathbb{Z}^d/\mathbb{Z}^d$ with $0 \in \Gamma_M$, cf. [17]. Since $\{\varphi_{1,k} : k \in \mathbb{Z}^d\}$ and $\{\widetilde{\varphi}_{1,k} : k \in \mathbb{Z}^d\}$ are Riesz bases for V_1 and $\widetilde{V}_1$, respectively, the inclusions $W_0 \subset V_1$ and $\widetilde{W}_0 \subset \widetilde{V}_1$ provide that there exist sequences $\left(a_k^{(\mu)}\right)_{k\in\mathbb{Z}^d} \in \ell_2(\mathbb{Z}^d)$ and $\left(b_k^{(\mu)}\right)_{k\in\mathbb{Z}^d} \in \ell_2(\mathbb{Z}^d)$ such that

$$\psi^{(\mu)}(x) = \sum_{k\in\mathbb{Z}^d} a_k^{(\mu)} \varphi(Mx-k) \qquad \text{and} \qquad \widetilde{\psi}^{(\mu)}(x) = \sum_{k\in\mathbb{Z}^d} b_k^{(\mu)} \widetilde{\varphi}(Mx-k).$$

Again, we assume that both sequences, $\left(a_k^{(\mu)}\right)_{k\in\mathbb{Z}^d}$ and $\left(b_k^{(\mu)}\right)_{k\in\mathbb{Z}^d}$, have only finitely many entries. Let $a^{(\mu)}$ and $b^{(\mu)}$ denote their symbols, respectively. The geometrical conditions (1), (3) with the complement property (2) imply

$$\sum_{\gamma\in\Gamma_M} a^{(\mu)}(\xi+\gamma)\overline{b^{(\nu)}(\xi+\gamma)} = \delta_{\mu,\nu}, \quad \mu,\nu = 0,\dots,m-1, \tag{7}$$

see [17] for details. Note that (7) includes the duality relations (6).

Given the coefficients of f in V_0, i.e.,

$$H_0^{(0)}(k) := \langle f, \widetilde{\varphi}_{0,k}\rangle, \quad k \in \mathbb{Z}^d,$$

the fast wavelet transform allows us to efficiently compute the coefficients of the decomposition (5), i.e.,

$$H_{j_0}^{(0)}(k) = \left\langle f, \widetilde{\varphi}_{j_0,k}\right\rangle \qquad \text{and} \qquad H_j^{(\mu)}(k) = \left\langle f, \widetilde{\psi}_{j,k}^{(\mu)}\right\rangle, \quad k \in \mathbb{Z}^d,$$

for $1 \le \mu \le m-1,\ j_0 \le j \le -1$. The inverse transform allows us to derive the coefficients in (4) from the coefficients in (5), cf. Algorithm 1. Similar redundant wavelet systems have been studied, e.g., in [22, 36].

(a) Decomposition:
Input: $H_0^{(0)}$
for $j = -1, -2, \dots, j_0$ **do**
  for $\mu = 0, \dots, m-1$ **do**

$$H_j^{(\mu)} := \left(b^{(\mu)} * H_{j+1}^{(0)}\right) \downarrow_M$$

  end
end
Output: $H_{j_0}^{(0)}$ and $H_j^{(\mu)}$, for $\mu = 1, \dots, m-1$, $j = -1, \dots, j_0$ with

$$H_{j_0}^{(0)}(k) = \left\langle f, \widetilde{\varphi}_{j_0,k} \right\rangle, \quad H_j^{(\mu)}(k) = \left\langle f, \widetilde{\psi}_{j,k}^{(\mu)} \right\rangle$$

(b) Reconstruction:
Input: $H_{j_0}^{(0)}$ and $H_j^{(\mu)}$, $\mu = 1, \dots, m-1$, $j = -1, -2, \dots, j_0$
for $j = j_0, \dots, -1$ **do**

$$H_{j+1}^{(0)} := \sum_{\mu=0}^{m-1} \overline{a^{(\mu)}} * (H_j^{(\mu)} \uparrow_M)$$

end
Output: $H_0^{(0)}$

Algorithm 1: The fast wavelet transform. The terms $a^{(\mu)}$ and $b^{(\mu)}$ are identified with their coefficient sequence, so that the convolution makes sense, and the symbols $\downarrow_M$ and $\uparrow_M$ denote down- and up-sampling, i.e., $(a_k)_{k\in\mathbb{Z}^d} \downarrow_M = (a_{Mk})_{k\in\mathbb{Z}^d}$ and $(a_k)_{k\in\mathbb{Z}^d} \uparrow_M = \begin{cases} a_{M^{-1}k}, & k \in M\mathbb{Z}^d, \\ 0, & \text{otherwise.} \end{cases}$

2.2. **Wavelet Packets.** The DWT is based on splitting each approximation space V_{j+2} into an approximation V_{j+1} at lower scale and a detail space W_{j+1}, which itself has the natural decomposition,

$$W_{j+1} = \bigoplus_{\mu=1}^{m-1} W_{j+1}^{(\mu)}, \quad \text{where} \quad W_{j+1}^{(\mu)} := \operatorname{span}\{\psi_{j+1,k}^{(\mu)} : k \in \mathbb{Z}^d\}.$$

Wavelet packets allow us to further split the spaces $W_{j+1}^{(\mu)}$, as needed. Indeed, for $\mu = 1 \dots, m-1$ and $\nu = 0, \dots, m-1$, we define the wavelet packets system to be:

$$\psi^{(\mu,\nu;j)}(x) = \sum_{k\in\mathbb{Z}^d} a_k^{(\nu)} \psi_{j+1,k}^{(\mu)} \qquad \text{and} \qquad \widetilde{\psi}^{(\mu,\nu;j)}(x) = \sum_{k\in\mathbb{Z}^d} b_k^{(\nu)} \widetilde{\psi}_{j+1,k}^{(\mu)}. \tag{8}$$

Then, the subspaces,

$$W_j^{(\mu,\nu)} := \operatorname{span}\{\psi_{0,k}^{(\mu,\nu;j)} : k \in \mathbb{Z}^d\} \qquad \text{and} \qquad \widetilde{W}_j^{(\mu,\nu)} := \operatorname{span}\{\widetilde{\psi}_{0,k}^{(\mu,\nu;j)} : k \in \mathbb{Z}^d\},$$

satisfy

$$W_{j+1}^{(\mu)} = \bigoplus_{\nu=0}^{m-1} W_j^{(\mu,\nu)} \qquad \text{and} \qquad \widetilde{W}_{j+1}^{(\mu)} = \bigoplus_{\nu=0}^{m-1} \widetilde{W}_j^{(\mu,\nu)}.$$

Thus, the collections,

$$\bigcup_{\mu=1}^{m-1} \bigcup_{\nu=0}^{m-1} \{\psi_{0,k}^{(\mu,\nu;j)} : k \in \mathbb{Z}^d\} \qquad \text{and} \qquad \bigcup_{\mu=1}^{m-1} \bigcup_{\nu=0}^{m-1} \{\widetilde{\psi}_{0,k}^{(\mu,\nu;j)} : k \in \mathbb{Z}^d\},$$

are biorthogonal Riesz bases for W_{j+1} and $\widetilde{W}_{j+1}$, respectively, cf. [11]. Next, we can split the spaces $W_j^{(\mu,\nu)}$ and $\widetilde{W}_j^{(\mu,\nu)}$ analogously, and, after k iterations, we derive a sequence of subspaces $W_j^{(\mu_1,\ldots,\mu_k)}$ and $\widetilde{W}_j^{(\mu_1,\ldots,\mu_k)}$ such that

$$W_{j+1}^{(\mu_1,\ldots,\mu_k)} = \bigoplus_{\mu_{k+1}=0}^{m-1} W_j^{(\mu_1,\ldots,\mu_{k+1})} \qquad \text{and} \qquad \widetilde{W}_{j+1}^{(\mu_1,\ldots,\mu_k)} = \bigoplus_{\mu_{k+1}=0}^{m-1} \widetilde{W}_j^{(\mu_1,\ldots,\mu_{k+1})}.$$

This splitting trick is completely based on the relations (7), cf. [11]. If we keep the number of splits of the detail spaces bounded, for instance, by fixing k over all scales, then we can again derive a Riesz basis [11].

As described above, the concepts of wavelet packets and its associated transform are a direct generalization of the DWT. When decomposing approximation and detail spaces simultaneously down to scale j_0, we obtain a full wavelet tree. Let $\mathcal{T}$ denote the collection of those subtrees of a full wavelet tree, whose nodes have either m or zero descendants, see Figure 2. Thus, each node represents a basis of a particular subspace, and each subtree represents a different basis of the entire V_0.

Our next step is to modify the wavelet coefficients in order to reduce noise and unwanted artifacts. This step will ensure the stability of minimization of the entropy functional and hence will impact the selection of the best basis. For this purpose, we shall *shrink* the wavelet coefficients associated with each node. A large variety of shrinkage strategies has been proposed in the literature. Here, we recall a few common shrinkage rules, see also Figure 1. *Soft-* and *hard-shrinkage*,

$$\varrho_s(x,\alpha) = (x - \operatorname{sign}(x)\alpha)\mathbf{1}_{\{|x|>\alpha\}} \qquad \text{and} \qquad \varrho_h(x,\alpha) = x\mathbf{1}_{\{|x|>\alpha\}},$$

respectively, are most prominent in wavelet schemes [9, 19], but many other techniques have proven useful, and a common theoretical foundation is given in [23]. While soft-shrinkage still modifies large x, hard-shrinkage retains large coefficients, but introduces a discontinuity. The *(nonnegative) garotte-shrinkage* rule,

$$\varrho_g(x,\alpha) = (x - \frac{\alpha^2}{x})\mathbf{1}_{\{|x|>\alpha\}},$$

is continuous, and large coefficients are left almost unaltered. It has been successfully applied to image denoising in [25]. Bruce and Gao proposed *firm-shrinkage*,

$$\varrho_{f,\beta}(x,\alpha) = x\mathbf{1}_{\{|x|>\beta\}} + \operatorname{sign}(x)\frac{\beta(|x|-\alpha)}{\beta-\alpha}\mathbf{1}_{\{\alpha\le|x|\le\beta\}}, \quad \beta > 0,$$

in [26] as a piecewise linear method.

After shrinking the wavelet coefficients associated to each node, we aim to find the optimal subtree with respect to some entropy E (or cost function) of the wavelet packets decomposition, that is defined as a nonnegative map

$$E : \mathcal{T} \to \mathbb{R}.$$

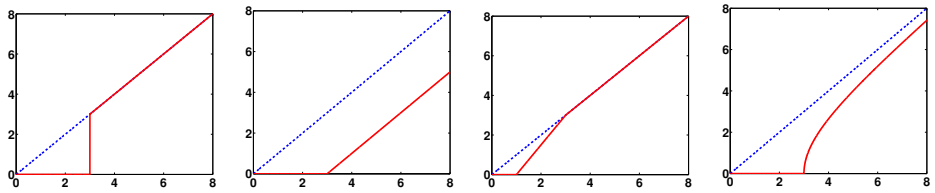
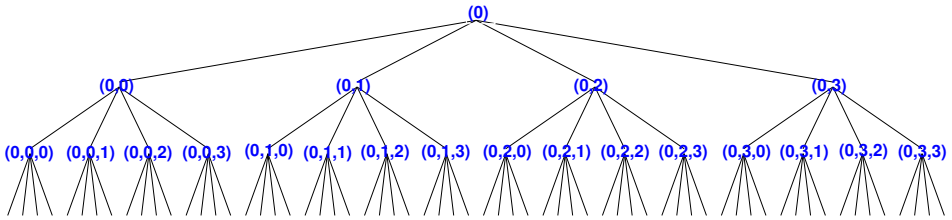
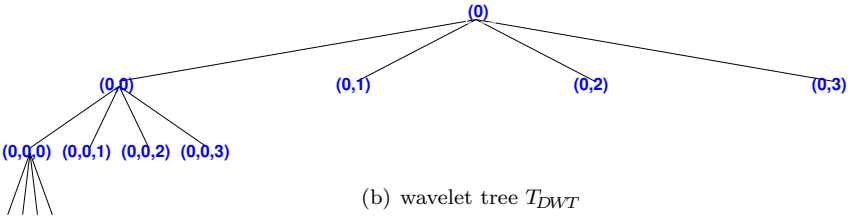


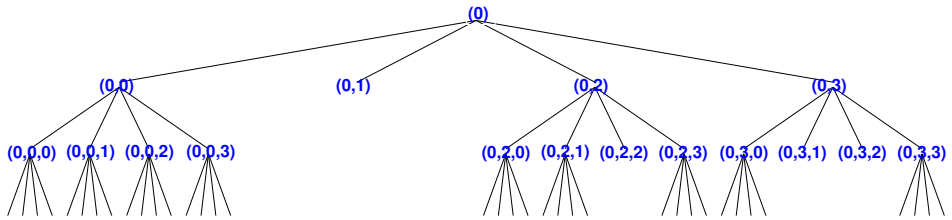
FIGURE 1. The shrinkage rules: hard-, soft-, firm with $\beta = 1$, and garotte



(a) full wavelet packets tree T_{full}



(b) wavelet tree T_{DWT}



(c) possible wavelet packets subtree $\hat{T}$

FIGURE 2. The minimization of the entropy term determines the subtree $\hat{T}$ that is chosen for reconstruction. From left to right, the four child nodes correspond to $a^{(0)}$, $a^{(1)}$, $a^{(2)}$, and $a^{(3)}$.

The subtree $\hat{T}$, at which E attains its minimum, allows us to select the best basis with respect to the chosen entropy. W refer the reader to Figure 2 for a visualization of the subtree concept for $M = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. The WPT is shown in Algorithm 2.

2.3. **Joint entropy functional and dimension reduction.** Before the application of the entropy and selection of the best basis, wavelet coefficients were thresholded to remove noise and artifacts. This procedure additionally enables a more rapid entropy minimization and smooths the reconstruction [18]. The best basis

(a) Decomposition:
Input: $H^{(\mu_1)}$, $\mu_1 = 0 \ldots, m-1$
for $i = 1, \ldots, -j_0$ **do**
  for $\mu_1, \ldots, \mu_{i+1} = 0, \ldots, m-1$ **do**

$$H^{(\mu_1,\ldots,\mu_{i+1})} := \left(b^{(\mu_{i+1})} * H^{(\mu_1,\ldots,\mu_i)}\right) \downarrow_M$$

  end
end
Output: $H^{(\mu_1,\ldots,\mu_i)}$, $\mu_1, \ldots, \mu_i = 0, \ldots, m-1$, $i = 1, \ldots, -j_0$.

(b) Best basis selection:
Input: $H^{(\mu_1,\ldots,\mu_i)}$, $\mu_1, \ldots, \mu_i = 0, \ldots, m-1$, $i = 1, \ldots, -j_0$.
choose admissible subtree

$$\hat{T} := \arg\min_{T \in \mathcal{T}} E(T)$$

Output: $\hat{T}$ and $H^{(\mu_1,\ldots,\mu_i)}$, $(\mu_1, \ldots, \mu_i) \in \hat{T}$

(c) Reconstruction:
Input: $\hat{T}$ and $H^{(\mu_1,\ldots,\mu_i)}$, $(\mu_1, \ldots, \mu_i) \in \hat{T}$
for $i = -j_0, \ldots, 1$ **do**
  if $(\mu_1, \ldots, \mu_i, 0) \in \hat{T}$ **then**

$$H^{(\mu_1,\ldots,\mu_i)} := \sum_{\mu_{i+1}=0}^{m-1} \overline{a^{(\mu_{i+1})}} * (H^{(\mu_1,\ldots,\mu_{i+1})} \uparrow_M)$$

  end
end
Output: $H^{(\mu_1)}$, $\mu_1 = 0 \ldots, m-1$

Algorithm 2: The wavelet packets transform. Let $f = \sum_{k\in\mathbb{Z}^d} H_k^{(0)} \varphi_{0,k} + \sum_{\mu_1=1}^{m-1} \sum_{k\in\mathbb{Z}^d} H_k^{(\mu_1)} \psi_{0,k}^{(\mu_1)}$. If $f \in V_0$, then $H^{(\mu_1)} = 0$, $\mu = 1 \ldots, m-1$.

algorithm allows us to find a set of wavelet bases that provides the most desirable representation of the data relative to a particular cost function. Such a cost function needs to fit the particular application. Here, we are given n spectral bands, we decompose each band by means of the WPT, and we apply wavelet thresholding. We choose an entropy E_i, $i = 1, \ldots, n$, for each band, and define the joint entropy E through the weighted ℓ_p-term over $\{E_i\}_{i=1}^n$, so that the best subtree $\hat{T}$ is selected by

$$\hat{T} := \arg\min_{T\in\mathcal{T}} E = \arg\min_{T\in\mathcal{T}} \sum_{i=1}^{n} \omega_i |E_i(T)|^p, \tag{9}$$

where $0 < p \leq 2$ and $\{\omega_i\}_{i=1}^n$ represents a sequence of positive weights to be specified. As such the subtree $\hat{T}$ provides a suitable basis which incorporates both spatial and spectral information from the multispectral image sets.

The wavelet thresholding yields image smoothing and can be thought of as a local linearization step. Each node of the joint subtree is now a coefficient vector whose entries refer to different spectral bands. The vectors of wavelet coefficients likely do not lie on a linear manifold exactly, but may be approximated in a linear

regime locally. To reduce the dimension of the coefficient vectors, we shall apply PCA.

Remark 2.1. The wavelet transform and PCA are both linear transformations. However, wavelet shrinkage introduces a nonlinearity to the approximation of the multispectral data set, cf [34]. In principle, PCA can be replaced with any other dimension reduction method, such as Locally Linear Embedding (LLE) [37], Hessian LLE [20], Laplacian Eigenmaps (LE) [3, 4], or Diffusion Wavelets/Diffusion Maps [13, 15]. However, these methods are computationally expensive and numerically do not appear as stable as PCA.

3. Hyperspectral image sets

Sunlight is diffusely reflected (scattered) by materials on the Earth's surface, and hyperspectral satellite and airborne imaging records the reflected electromagnetic spectrum partitioned into spectral bands. The reflectivity of objects and materials across the electromagnetic spectrum is referred to as their spectral signatures, enabling classification of a scene from a set of hyperspectral images. The recording procedure of hyperspectral sensors yields a certain number of bands, pixels, and lines, all of which depend on the sensor. Such data can be arranged into a collection of images, each single one being associated with one spectral band.

To compare our proposed mathematical analysis approach with other methods, we shall apply several wavelet type schemes to analyze publicly available hyperspectral data sets. For simplicity, we restrict our considerations to the 2-dimensional tensor Haar-wavelet, so that M is two times the identity matrix. We shall use the Shannon entropy functional [43], and we select constant weights $\omega_i = 1$, for all $i = 1, \ldots, n$. Garotte shrinkage is applied, setting 70% of the wavelet coefficients to zero. The number of principal components is chosen so that 95% of the variance in the wavelet domain is captured.

3.1. Indian Pines, Pavia University, Salinas, Salinas-A.

For the quantitative evaluation of our proposed method, we have chosen 4 data sets with publicly available ground truth. The first collection of hyperspectral images is Indian Pines from the AVIRIS sensor with 200 bands, and 145 pixels by 145 lines. The Pavia University collection consists of 103 spectral bands, and 610 pixels by 610 lines and was gathered by the ROSIS sensor. Both data sets were analyzed in [7, 21]. The third data set is Salinas collected by the AVIRIS sensor with 202 bands, and 217 pixels by 512 lines. Salinas-A is a subscene of Salinas with 86×83 pixel images per spectral band. All 4 data sets include ground truth, see Figure 3, and are available as Matlab files at http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

We shall use several wavelet type schemes to preprocess the data sets and then apply the Spectral Angle Mapper (SAM) [28, 32, 33] to derive the final classification. For SAM analysis, we label randomly selected 20% of each classes' pixels. We only count the number of correctly classified pixels, hence, prioritizing the minimization of false-negatives and ignoring false-positives. Indeed, the percentage of correctly classified pixels for the Indian Pines and Pavia University data sets appears consistently higher than in [7, 21]. We do note however that such direct comparisons are rather delicate and may also depend on selected wavelet approaches, pre- and post-processing, as well as other factors.

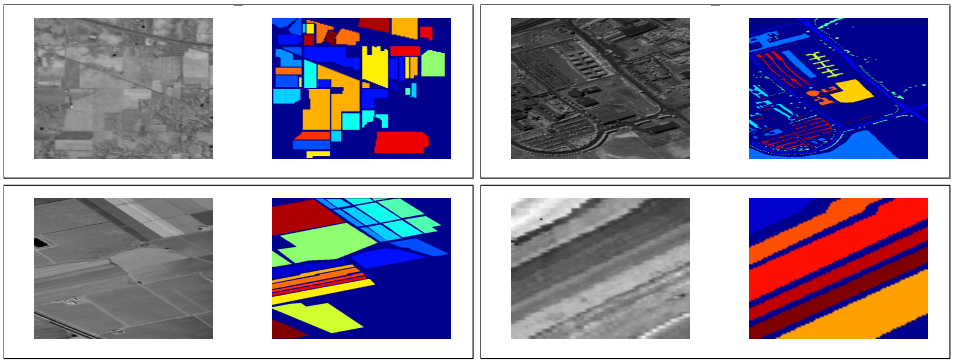


FIGURE 3. 4 hyperspectral data sets with associated ground truth: Indian Pines, Pavia University, Salinas, and Salinas-A.

The results of our analysis are shown in Table 1. First, SAM was applied directly to the raw data. In another experiment we applied PCA and then SAM. Next, the DWT was used, with and without shrinkage and PCA. The standard WPT for each spectral band separately with separate entropy terms was also used. The WPT with joint entropy term appears to yield better classification results. It seems that PCA improves the classification results more than shrinkage. To us, the latter seemed a bit counterintuitive because small wavelet coefficients are considered of less importance so that they would interfere with the classification scheme. So, we can claim that PCA is an even more important contributor. In summary, the WPT with joint entropy and the choice $p = 1$, combined with shrinkage and PCA, appears to provide the best results among the tested methods in our numerical experiments, cf. Table 1.

	% correct			
Method	Indian Pines	Pavia U.	Salinas	Salinas-A
raw data	62 %	60 %	59 %	58 %
PCA	65 %	69 %	64 %	66 %
DWT	71 %	67 %	69 %	70 %
DWT+shrinkage+PCA	81 %	76 %	79%	78 %
WPT(sep)	70 %	65 %	66 %	68 %
WPT(sep)+shrinkage+PCA	79 %	75 %	80 %	77 %
WPT($p = 2$)	74 %	72 %	71 %	73 %
WPT($p = 2$)+shrinkage+PCA	84 %	81 %	84 %	86 %
WPT($p = 1$)	73 %	75 %	72 %	71 %
WPT($p = 1$)+shrinkage+PCA	87 %	84 %	91 %	92%

TABLE 1. Comparison of different methods for 4 data sets. *Raw data* - unprocessed, *WPT(sep)* - entropy is used separately for each band, *WPT($p = i$)* - joint entropy is used with $p = i$, $i = 1, 2$. The wavelet packets transform combined with shrinkage and principal component analysis consistently outperforms the other methods.

Since wavelet packets with joint entropy, shrinkage, and PCA outperform the other methods, we list the detailed results for this particular experiment for each

class in each data set, in order to provide further insight into the performance of our proposed scheme. Tables 2, 3, 4, and 5 show the number of correctly classified samples per class and the rate of success.

Class	#Samples	# correct	% correct
Alfalfa	46	33	72 %
Corn-notill	1428	1313	92 %
Corn-mintill	830	689	83 %
Corn	237	178	75 %
Grass-pasture	483	386	80 %
Grass-trees	730	628	86 %
Grass-pasture-mowed	28	20	71 %
Hay-windrowed	478	425	89 %
Oats	20	19	95 %
Soybean-notill	972	894	92 %
Soybean-mintill	2455	2136	87 %
Soybean-clean	593	498	84 %
Wheat	205	162	79 %
Woods	1265	1151	91 %
Buildings-Grass-Trees-Drives	386	317	82 %
Stone-Steel-Towers	93	86	92 %
Total	10249	8935	87 %

TABLE 2. AVIRIS sensor, Indian Pines Scene, WPT+shrinkage+PCA.

Class	#Samples	# correct	% correct
Asphalt	6631	5437	82 %
Meadows	18649	16971	91 %
Gravel	2099	1847	88 %
Trees	3064	2482	81 %
Painted metal sheets	1345	1036	77 %
Bare Soil	5029	4224	84 %
Bitumen	1330	1237	93 %
Self-Blocking Bricks	3682	3166	86 %
Shadows	947	572	60 %
Total	42776	36072	84 %

TABLE 3. ROSIS sensor, Pavia University Scene, WPT+shrinkage+PCA.

3.2. **Copperas Cove data set.** In the preceding section, we have validated the performance of our proposed wavelet packets transform with joint entropy for $p = 1$ combined with shrinkage and PCA. To add additional dimension to our experimental comparisons, we also analyze the HYDICE sensor imagery data set, a collection of 210 spectral bands, each 307 pixels by 307 lines, taken over Copperas Cove, TX. It is publicly available at http://www.agc.army.mil/hypercube/. As before, we shall use garotte shrinkage with setting 70% of the wavelet coefficients to zero and we retain 12 principal components capturing more than 95% of the variance of

Class	#Samples	# correct	% correct
Brocoli-green-weeds-1	2009	1804	90 %
Brocoli-green-weeds-2	3726	3316	89 %
Fallow	1976	1817	92 %
Fallow-rough-plow	1394	1255	90 %
Fallow-smooth	2678	2276	85 %
Stubble	3959	3682	93 %
Celery	3579	3436	96 %
Grapes-untrained	11271	10657	95 %
Soil-vinyard-develop	6203	5521	89 %
Corn-senesced-green-weeds	3278	2917	95 %
Lettuce-romaine-4wk	1068	897	84 %
Lettuce-romaine-5wk	1927	1657	86 %
Lettuce-romaine-6wk	916	669	73 %
Lettuce-romaine-7wk	1070	973	91 %
Vinyard-untrained	7268	6796	94 %
Vinyard-vertical-trellis	1807	1554	86 %
Total	54129	49257	91 %

TABLE 4. AVIRIS sensor, Salinas Scene, WPT+shrinkage+PCA.

Class	#Samples	# correct	% correct
Brocoli-green-weeds-1	391	325	83 %
Corn-senesced-green-weeds	1343	1289	96 %
Lettuce-romaine-4wk	616	567	92 %
Lettuce-romaine-5wk	1525	1434	94 %
Lettuce-romaine-6wk	674	593	88 %
Lettuce-romaine-7wk	799	727	91 %
Total	5348	4935	92 %

TABLE 5. AVIRIS sensor, Salinas-A Scene, WPT+shrinkage+PCA.

the remaining coefficients. Since the associated ground truth is confidential data, we only show our classification in Figure 4 that appear to recover the different classes within the scene reasonably well.

4. Conclusions

We have proposed a new method to analyze multi- and hyperspectral image sets, that takes advantage of jointly analyzing the spatial and spectral content of the available data. The new concept of joint entropy within wavelet packets enables us to fuse spatial and spectral information encoded in the wavelet coefficients. Shrinkage is applied to reduce noise and artifacts. PCA is used in the wavelet domain to reduce the dimension of the data. Our scheme performed well in numerical experiments on several hyperspectral satellite image sets. Nevertheless, more extensive experiments and more elaborate analysis are required to better understand the joint entropy, the choice of the entropy parameter p, the shrinkage parameter α, and the weights $\{\omega_i\}_{i=1}^n$. It may also be advantageous to use redundant wavelet

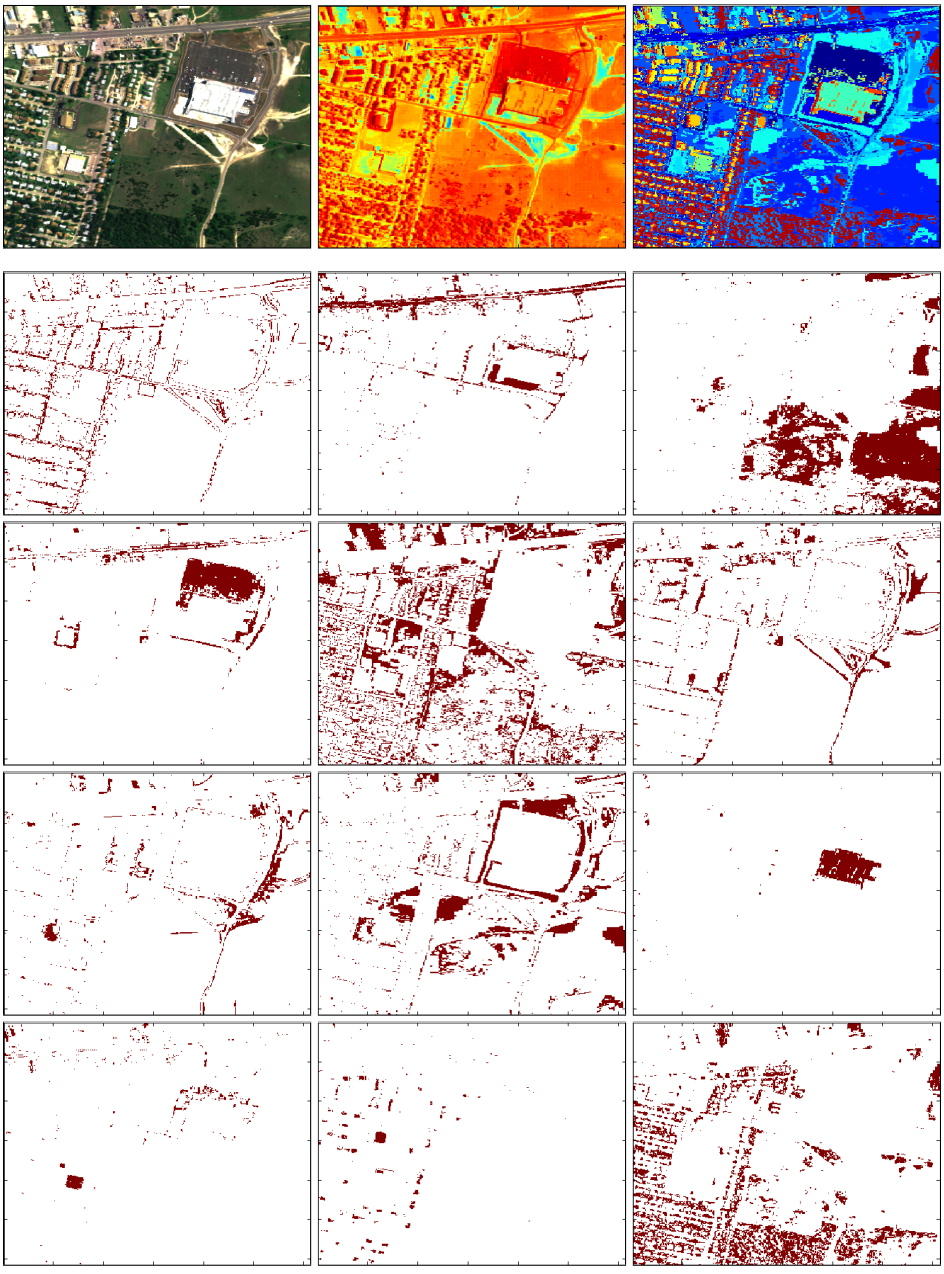


FIGURE 4. (first row) RGB image of the Urban scene, inverse transformed first principal component, and classes of our vector angle classification scheme after WPT($p = 1$)+shrinkage+PCA was used. (below) 12 detected classes shown separately.

systems, such as tight or bi-frames, offering more flexibility and better noise reduction. Finally, finding ways to take advantage of directly nonlinear dimension reduction methods, such as LLE or Laplacian EIgenmaps, is of great interest. We plan to address these topics in future research.

Acknowledgment

This work presented in this paper was supported in part by NSF (CBET 0854233), by NGA (HM 15820810009), by NIH/DFG (EH 405/1-1/575910), by WWTF (VRG 12-009), and by MURI-ARO (W911NF-09-0383).

References

[1] A. Agarwal, J. LeMoigne, J. Joiner, T. El-Ghazawi, and F. Cantonnet. Wavelet dimension reduction of AIRS infrared (IR) hyperspectral data. *Geoscience and Remote Sensing Symposium, IGARSS*, 2004.

[2] F. Bayer and M. Gutting. Spherical fast multiscale approximation by locally compact orthogonal wavelets. Int. J. Geomath., 2(1):69–85, 2011.

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *NIPS*, 14, 2002.

[4] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural. Comput.*, 15(6):1373–1396, 2003.

[5] J. J. Benedetto, W. Czaja, J. Dobrosotskaya, T. Doster, K. Duke, and D. Gillis. Integration of heterogeneous data for classification in hyperspectral satellite imagery. *Proceedings of SPIE*, Vol. 8390, 839027, 2012.

[6] J. J. Benedetto, W. Czaja, J. C. Flake, and M. Hirn. Frame based kernel methods for automatic classification in hyperspectral data. In *IEEE IGARSS*, Cape Town, South Africa, 2009.

[7] S. Bernabe, P. R. Marpu, A. Plaza, M. D. Mura, and J. A. Benediktsson. Spectral-spatial classification of multispectral images using kernel feature space representation. *to appear in IEEE Geosci. Remote Sens. Lett.*, 2013.

[8] L. M. Bruce, C. Morgan, and S. Larsen. Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms. *IEEE Trans. Geosci. Remote Sens.*, 39(10):2217–2226, 2001.

[9] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, 7:319–335, 1998.

[10] C. K. Chui, W. Czaja, M. Maggioni, and G. Weiss. Characterization of general tight wavelet frames with matrix dilation and tightness preservation oversampling. *J. Fourier Anal. and Appl.*, 8(2):173–200, 2002.

[11] A. Cohen and I. Daubechies. On the instability of arbitrary biorthogonal wavelet packets. *SIAM J. Math. Anal.*, 24(5):1340–1354, 1993.

[12] R. Coifman, Y. Meyer, and M. V. Wickerhauser. Size properties of wavelet packets. In M. B. Ruskai et al., editors, *Wavelets and Their Applications*. Jones and Bartlett, 1992.

[13] R. R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data. part ii: Multiscale methods. *Proc. Nat. Acad. Sci.*, 102:7432–7438, 2005.

[14] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. J. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data. part i: Diffusion maps. *Proc. Nat. Acad. Sci.*, 102:7426–7431, 2005.

[15] R. R. Coifman and M. Maggioni. Diffusion wavelets. *Appl. Comput. Harmon. Anal.*, 21(1):53–94, 2006.

[16] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theory*, 38(2):713–718, 1992.

[17] I. Daubechies. *Ten Lectures on Wavelets*, volume 9. SIAM, Philadelphia, 1992. CBMS-NSF Regional Conf. Ser. in Appl. Math. 61.

[18] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.

[19] D. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.
[20] D. L. Donoho and C. Grimes. Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data. *Proc. Nat. Acad. Sci.*, 100:5591–5596, 2003.
[21] I. Dopido, J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson. Semi-supervised self-learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sensing*, 51(7):4032 – 4044, 2013.
[22] M. Ehler. On multivariate compactly supported bi-frames. *J. Fourier Anal. Appl.*, 13(5):511–532, October 2007.
[23] M. Ehler. Shrinkage rules for variational minimization problems and applications to analytical ultracentrifugation. *J. Inverse Ill-Posed Probl.*, 19(4-5):593–614, 2011.
[24] W. Freeden and H. Nutz. Satellite gravity gradiometry as tensorial inverse problem. Int. J. Geomath., 2(2):177–218, 2011.
[25] H. Y. Gao. Wavelet shrinkage denoising using the non-negative garotte. *J. Comput. Graph. Statist.*, 7(4):469–488, 1998.
[26] H. Y. Gao and Andrew G. Bruce. Waveshrink with firm shrinkage. *Statistica Sinica*, 7:855–874, 1997.
[27] M. Gavish, B. Nadler, and R. R. Coifman. Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning. In *ICML*, pages 367–374, 2010.
[28] G. Girouard, A. Bannari, A. El Harti, and A. Desrochers. Validated spectral angle mapper algorithm for geological mapping: Comparative study between Quickbird and Landsat-TM. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 35(4):599–604, 2004.
[29] M. R. Gupta and N. P. Jacobson. Wavelet principal component analysis and its application to hyperspectral images. In *IEEE Conference on Image Processing*, 2006.
[30] P. Hsu. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(2):78–92, 2007.
[31] C. A. Jekeli. A wavelet approach to the terrain correction in gravimetry and gravity gradiometry. Int. J. Geomath., 3(1):139–154, 2012.
[32] F. Kruse et al. The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44:145–163, 1993.
[33] A. Mohan, G. Sapiro, and E. Bosch. Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.*, 4(2):206–210, April 2007.
[34] P. Mrázek, J. Weickert, and G. Steidl. Correspondences between wavelet shrinkage and nonlinear diffusion. *LNCS* 2695:101–116, 2003.
[35] R. Pande-Chhetri and A. Abd-Elrahman. De-striping hyperspectral imagery using wavelet transform and adaptive frequency domain filtering. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5):620–636, 2011.
[36] A. Ron and Z. Shen. Affine systems in $L_2(\mathbb{R}^d)$: the analysis of the analysis operator. *J. Funct. Anal.*, 148:408–447, 1997.
[37] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
[38] N. Saito. Data analysis and representation on a general domain using eigenfunctions of laplacian. *Appl. Comput. Harmon. Anal*, 25(1):68–97, 2008.
[39] B. Somers, K. Cools, S. Delalieux, J. Stuckens, D. Van der Zande, W. W. Verstraeten, and P. Coppin. Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards. *Remote Sensing of Environment*, 113(6):1183–1193, June 2009.
[40] X. Tang and W. A. Pearlman. Three-dimensional wavelet-based compression of hyperspectral images. *Hyperspectral Data Compression*, pages 273–308, 2006.
[41] X. Tang, W. A. Pearlman, and J. W. Modestino. Hyperspectral image compression using three-dimensional wavelet coding. *IEEE Trans. Geosci. Remote Sens.*, 3653:294–350, 2002.
[42] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–23, Dec 2000.
[43] M. V. Wickerhauser. *Adapted Wavelet Analysis: From Theory to Software*. A. K. Peters, LTD, 1996.
[44] P. Wojtaszczyk. *A Mathematical Introduction to Wavelets*. Cambridge University Press, 1997.

[45] A. C. Zelinski and V. K. Goyal. Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation. In *IEEE IGARSS*, 2006.

(J. J. Benedetto) Norbert Wiener Center, Department of Mathematics, University of Maryland, College Park, MD 20742, USA
E-mail address: jjb@math.umd.edu

(W. Czaja) Norbert Wiener Center, Department of Mathematics, University of Maryland, College Park, MD 20742, USA
E-mail address: wojtek@math.umd.edu

(M. Ehler) University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria
E-mail address: martin.ehler@helmholtz-muenchen.de