Relevance-based Feature Extraction from Hyperspectral Images in the ℂomplex Wavelet Domain

Michael J. Mendenhall[1]
Department of Electrical and Computer Engineering
Rice University, Houston, TX
e-mail: mendenmi@rice.edu

Erzsébet Merényi,
Department of Electrical and Computer Engineering
Rice University, Houston, TX
e-mail: erzsebet@rice.edu

***Abstract*— Generalized Relevance Learning Vector Quantization (GRLVQ) [1] is a "double action" supervised neural learning machine that simultaneously adapts classification boundaries and a weighting of the input dimensions to reflect the relevance of each dimension for the given classification. It is thus a joint classification and feature extraction technique. In [2] we developed an improved version (GRLVQI) to handle intricate high-dimensional data. GRLVQI makes significant headway of feature reduction for hyperspectral images without compromising classification accuracy. However, the number of features to which the data can be reduced in the original (reflectance data) domain is naturally limited by higher order correlations. Here we investigate GRLVQI processing on wavelet coefficients because of the approximately decorrelated nature and the sparsity of those coefficients. We investigate the Dual-Tree ℂomplex Wavelet Transform (DTℂWT) [3] for its reduced oscillatory effects because spectral data often have discontinuities due to data fallout. We demonstrate that GRLVQI on the DTℂWT coefficients achieves better classification with fewer features than using the Critically Sampled Discrete Wavelet Transform (CSDWT), which was already shown to yield better results with far fewer features than GRLVQI applied in the original data space.**

I. Introduction

Accurate classification of remotely sensed hyperspectral images (described in Section I-C) is an important task for a wide variety of applications. The complexity and high dimensionality of the data makes it difficult to classify. If we can extract a set of meaningful features while suppressing features that confuse the classifier, it is possible to increase both classification accuracy and speed. However, meaningful feature extraction that allows one to retain a high quality classification of intricate data is a difficult task. Many methods approach feature extraction independent of classification, often yielding a reduced feature set but poorer classification than by using all available features. This end result is often unacceptable. In our work, we use a jointly optimized classifier and feature extractor: Generalized Relevance Learning Vector Quantization (GRLVQ) [1].

A. Relevance learning

The term relevance learning describes a supervised neural machine learning paradigm that simultaneously performs a) adaptive Learning Vector Quantization (LVQ) [4] for optimal placement of the prototype vectors to define class boundaries, and b) an adaptation of the so-called relevance vector, to reflect the importance of the input dimensions for discrimination of the given classes. The feedback between classifier training and feature extraction ensures that the latter is optimized for the particular classification problem. GRLVQ in its original form [1] was not robust enough for high-dimensional data (such as hyperspectral images). Our improvements in [2] remedied that, and all results here were produced by our improved version of GRLVQ, called GRLVQI(mproved). Both will be described briefly in Section II-A.

B. The novelty of a relevance-wavelet model for classification

Wavelet-based signal processing, such as compression, takes advantage of the signal compaction achieved by the wavelet transform. Significant information is contained in a *few* sparsely located coefficients with the largest magnitudes. Coupled with this powerful property, very successful image compression methods make use of the multi-scale aspects of the transform; significant wavelet coefficients persist across wavelet scales. Jerome Shapiro's embedded zero-tree wavelet compression algorithm is a well known example [5], which takes advantage of the multi-scale and sparsity aspects of the wavelet transform allowing for progressively better signal reconstruction as more significant coefficients are retained. There are a host of wonderful properties which make the wavelet transform a powerful signal processing tool.

The same success compression algorithms have with largest magnitude coefficient selection have not been shared by classification algorithms (see, e.g., Moon and Merényi [6]). Experiments in our previous work indicated that the largest magnitude wavelet coefficients have relatively poor class discrimination capability [7]. Similarly, classification using wavelet subband energy has largely been unsuccessful (see, e.g., Zhang et al. [8]).

Our relevance-wavelet model selects wavelet coefficients based on their GRLVQ-determined *importance* for the given classification which is a paradigm shift from more typical largest magnitude selection of coefficients. This paradigm shift results in very high classification accuracies with only a handful of wavelet coefficients [2].

[1]The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

C. Hyperspectral images

Hyperspectral images are coveted data that offer the necessary spectral details to distinguish virtually any material. They are a stack of (co-registered) images where each image is the measured light response in a narrow frequency band. This collection of hundreds of images can be thought of as an image cube (Fig. 1 left), where each pixel has associated with it a spectral curve (Fig. 1 right).

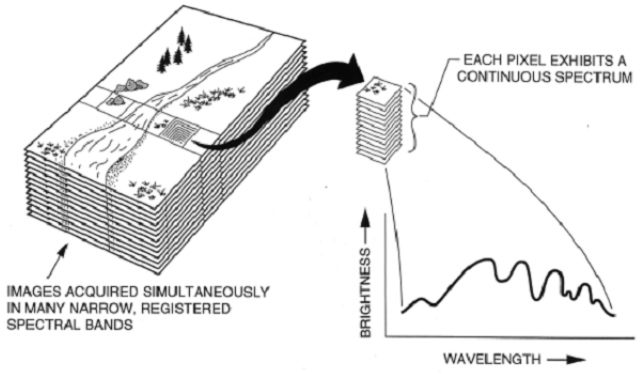

Fig. 1. A hyperspectral image cube where each pixel is an n-dimensional vector which is the spectrum of the material in that pixel. Figure with permission from Campbell [9].

Precise classification of hyperspectral data has historically proven very difficult. Yet, one wants to discriminate material differences that cutting-edge instrumentation records. Sophisticated classifiers and sophisticated feature extractors are required to realize the full potential of these data sets.

D. Discontinuities in hyperspectral data

Data fallout often causes discontinuities in spectra. The most common example in hyperspectral images is irreparable damage to data in two wavelength windows (see Fig. 2 top) where the atmospheric water vapor saturates the instrument response. These two regions are deleted, and a piecewise signal with discontinuities at the boundaries (Fig. 2 bottom) results.

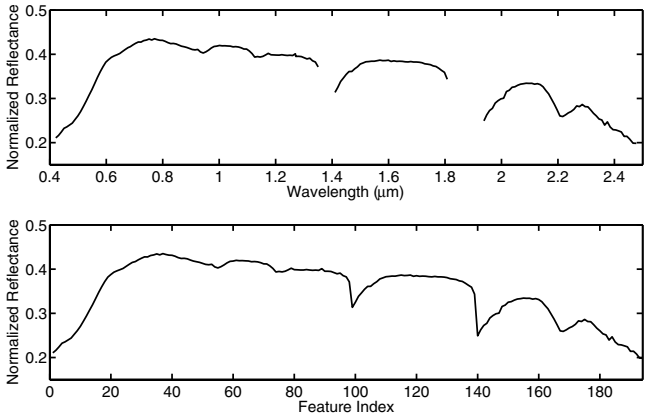

Fig. 2. Spectral discontinuities caused by data fallout. **Top:** Normalized reflectance spectrum vs. wavelength with "empty" regions where data are deleted due to saturation of the atmospheric water bands. **Bottom:** Reflectance vs. feature index illustrating the band locations of the discontinuities as they appear to the wavelet transform.

The deletion of spectral bands has no ill effect on our ability to process the data using GRLVQI on the spectral features. However, the discontinuities can manifest in a (wavelet) transform domain creating a set of false features. These false features are a concern if using GRLVQI because it may waste relevance resources by learning artifacts. We can potentially mitigate the effect of the discontinuities by carefully choosing the transform.

The Dual-Tree $\mathbb{C}$omplex Wavelet Transform (DT$\mathbb{C}$WT) is known to have less oscillation due to discontinuities then other forms of the wavelet transform (e.g., the Critically Sampled Discrete Wavelet Transform (CSDWT) with Daubechies length-four orthogonal (db4) filters [10] used in our previous work [7]).

To demonstrate the reduced oscillatory effects of the DT$\mathbb{C}$WT over the CSDWT, we define a function to place discontinuities in the vicinity of the locations where spectral bands were deleted, "between" the band pairs (98,99) and (139,140) (see Fig. 2 bottom):

$$\begin{aligned} f[n] &= u[n-97]-u[n-101] \\ &\quad +u[n-138]-u[n-142]. \end{aligned} \tag{1}$$

where n indicates band index and $u[\cdot]$ stands for the unit step function. We then take the wavelet transform of $f[n]$. Fig. 3 top is the CSDWT of $f[n]$ using the CSDWT with the db4 filters and Fig. 3 bottom with the DT$\mathbb{C}$WT from Selesnick et al. [3]. Fig. 3 shows that the DT$\mathbb{C}$WT results in less oscillation of the wavelet coefficients.

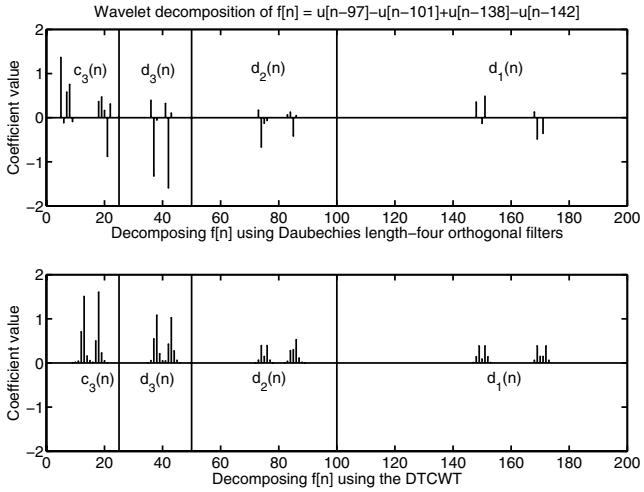

Fig. 3. Wavelet transform of $f[n]$ in Eq. 1 in the vicinity of the spectral discontinuities. **Top:** Three scales of the CSDWT of $f[n]$ using the db4 filters. **Bottom:** The magnitude of three scales of the DT$\mathbb{C}$WT of $f[n]$.

II. Background on GRLVQ(I) & the DT$\mathbb{C}$WT

A. GRLVQ and GRLVQ-Improved

Learning Vector Quantization 2.1 (LVQ2.1), the third in a series of progressively more sophisticated paradigms defined by Kohonen [4], is an adaptive neural learning paradigm used for supervised classification. Prototype vectors are adapted to define classification boundaries while minimizing the Bayes

risk. This is accomplished by *differentially shifting* a best matching in-class prototype vector w^J (with distortion d^J) which has the same class label as the input sample x^m, and a best matching out-of-class prototype vector w^K (with distortion d^K) which has a different label than the input sample x^m, at each iteration (Fig. 4). The distortion measure used is the squared Euclidean distance.

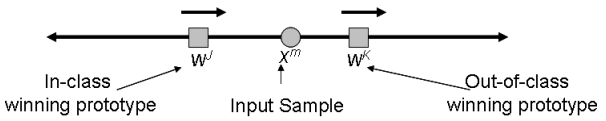

Fig. 4. Differential shifting of the in-class (w^J) and out-of-class (w^K) winning prototype vectors.

In LVQ2.1, prototype vectors can drift from their optimal locations once placed, resulting in a degradation of classification accuracy. Sato and Yamada [11] addressed this divergence issue in their Generalized LVQ (GLVQ). GLVQ descends a *cost function* C which is a sum where each term is a sigmoid-modulated measure of the *classification success* $\mu(\cdot)$, of one input sample. $f(\cdot)$ is called the loss function (Eqn. 2).

$$C = \sum_{m=1}^{M} f(\mu(x^m)) \tag{2}$$

$$\begin{aligned} f(\mu(x^m)) &= \frac{1}{1+e^{-\mu(x^m)}}, \\ &= \frac{1}{1+e^{\left(-\frac{d^J-d^K}{d^J+d^K}\right)}} \end{aligned} \tag{3}$$

$$\mu(x^m) = \left(\frac{d^J-d^K}{d^J+d^K}\right) \tag{4}$$

Using GLVQ as the starting point, Hammer and Villmann [1] incorporate the learning of an importance weighting of the input dimensions for classification. This adaptive diagonal metric is the so-called relevance (a vector $\lambda = (\lambda_1,\ \lambda_2,\ ...,\ \lambda_n)$ where n is the dimensionality of the data), and gives Generalized Relevance LVQ (GRLVQ). Adaptation of the prototype vectors is accomplished via gradient descent [1], [11]:

$$\Delta w^J = \frac{4\epsilon(t)^J f'|_{\mu(x^m)} d_\lambda^K}{(d_\lambda^J + d_\lambda^K)^2} \Lambda(x^m - w^J), \tag{5}$$

$$\Delta w^K = -\frac{4\epsilon(t)^K f'|_{\mu(x^m)} d_\lambda^J}{(d_\lambda^J + d_\lambda^K)^2} \Lambda(x^m - w^K), \tag{6}$$

where Λ is a diagonal matrix with the λ_i as its diagonal elements, f' is the derivative of f, and $\epsilon(t)^J$ and $\epsilon(t)^K$ are (potentially) time-varying in-class and out-of-class learn rates, respectively.

Similarly, adaptation of the relevance factors is derived as [1]:

$$\begin{aligned} \Delta\lambda_i &= -\frac{2\epsilon(t)^\lambda f'|_{\mu(x_i^m)} d_\lambda^K (x_i^m - w_i^J)}{(d_\lambda^J + d_\lambda^K)^2} \\ &\quad + \frac{2\epsilon(t)^\lambda f'|_{\mu(x_i^m)} d_\lambda^J (x_i^m - w_i^K)}{(d_\lambda^J + d_\lambda^K)^2} \end{aligned} \tag{7}$$

For stability reasons, relevances must be scaled such that $\|\lambda\|_p = 1$ [1] and $\lambda_i \geq 0 \quad \forall i \in \{1, ..., n\}$. We choose $p = 1$ because relevances then have a nice numerical interpretation as percentage of a whole.

In our earlier analysis and experimentation with GRLVQ [2], we find it still suffers from the potential for diverging prototype vectors, and in addition, from poor prototype utilization.

We remedied the first problem by a change to the learning rule. Our *in-class conditional update* learning rule updates both the in-class and out-of-class winning prototype vectors *only* if the input sample is misclassified. If the input sample *is* classified correctly, we only update the in-class prototype vector [2].

We addressed the problem of poor prototype utilization by adapting DeSieno's conscience learning [12] for in-class prototype selection, separately for each class. The idea behind conscience learning is to bias the Euclidean distance between the input sample x^m and the prototype vector w^p with the bias B^p. The bias B^p is calculated based on the winning frequency of w^p and alters the chance of w^p becoming the winner, encouraging the selection of infrequent winners and discouraging the selection of frequent winners. This equiprobablistic approach ensures all prototypes receive an opportunity to learn during the training process, and achieves an information theoretically optimum quantization of the data.

Our improvements to GRLVQ (dubbed GRLVQI) result in better classification and decreased training time. Specifics of the GRLVQ algorithm and our improved GRLVQI, along with experimental results, can be found in [2].

B. The Dual-Tree ℂomplex Wavelet Transform

The Critically Sampled Discrete Wavelet Transform (CSDWT) represents the signal $f(t)$ as the sum of its scaling coefficients $c(n)$ and wavelet coefficients $d_k(n)$:

$$f(t) = \sum_{n=-\infty}^{\infty} c(n)\phi(t-n) \tag{8}$$

$$+ \sum_{k=0}^{\infty} \sum_{n=-\infty}^{\infty} d_k(n) 2^{k/2} \psi(2^k t - n) \tag{9}$$

where $\phi(t)$ is the scaling function and $\psi(2^k t - n)$ the wavelet function at scale k.

To perform a k-level CSDWT of the function $f(t)$, one simply takes the inner product of $f(t)$ with scaling function $\phi(t)$ and wavelet function $\psi(2^k t - n)$:

$$c(n) = \int_{-\infty}^{\infty} f(t)\phi(t-n)dt, \tag{10}$$

$$d_k(n) = \int_{n=-\infty}^{\infty} f(t) 2^{k/2} \psi(2^k t - n), \tag{11}$$

where $2^{k/2}$ is a normalizing term.

One can efficiently compute the wavelet and scaling functions using a filter bank. Here, the discrete input signal $f[n]$ is filtered with low-pass scaling filter $H(z)$ and high-pass wavelet filter $G(z)$, iterating on the low-pass scaling

coefficients at each scale. This process is demonstrated in Fig. 5 for a three-level wavelet transform.

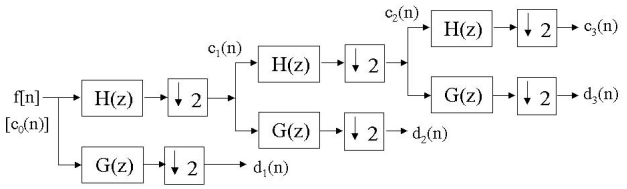

Fig. 5. Analysis filter bank for the Critically Sampled Discrete Wavelet Transform. The filters $H(z)$ and $G(z)$ are the z-transforms of the high-pass wavelet filters and low-pass scaling filters respectively. The symbol $\downarrow 2$ denotes a down sampling operation by a factor of two.

The Dual-Tree $\mathbb{C}$omplex Wavelet Transform (DT$\mathbb{C}$WT) (see Selesnick et al. [3]) extends the CSDWT by defining two trees (or filter banks): one filter bank computes the real part while a second filter bank computes the imaginary part. Each low-pass scaling coefficient and high-pass wavelet coefficient is the sum of its real and imaginary parts: $c^c(n) = c^r(n) + jc^i(n)$, $d_k^c(n) = d_k^r(n) + jd_k^i(n))$, where $j = \sqrt{-1}$.

In the case of the real component, the scaling function $\phi(t)$ and wavelet function $\psi(t)$ are both real and even (symmetric) functions [3]. Conversely, for the imaginary component, the scaling and wavelet functions are both imaginary and odd (anti-symmetric) [3]. We can similarly write the scaling and wavelet functions as the sum of their real and imaginary parts: $\psi^c(t) = \psi^r(t) + j\psi^i(t)$ and $\phi_k^c(t) = \phi_k^r(t) + j\phi_k^i(t)$.

III. Design of experiments

A. The Lunar Crater Volcanic Field data set

We use a total of 931 / 1464 verified spectra (for a 23/35-class problem respectively) from a hyperspectral image of the Lunar Crater Volcanic Field (LCVF), NV test site obtained by the NASA/JPL AVIRIS [13] sensor. Each data sample has 194 spectral features after elimination of wavelength windows where data were irrecoverably lost. We use 35 predefined classes, identified in previous studies. Mean spectra and class labels are shown for 23 of those classes (A - W) in Fig. 6 (see [14] for a more complete description). The 35-class problem uses 12 additional classes, shown in [2], which introduce more sub-class structure making it a more challenging classification problem.

B. Experiment process

We train our GRLVQI classifier on a set of training data and periodically test the quality of the learning on a set of test data. After training is finished, we record the final classification accuracy of the test data as determined by GRLVQI, and count the number of significant relevant features. We define a feature as "significantly relevant" if GRLVQI assigns a relevance factor ≥ 0.001 to it. This threshold was chosen as it represents all relevances that contribute at least one-tenth of one percent to the classification. We then use the Minimum Euclidean Distance (MED) classifier to independently assess the quality of the feature set.

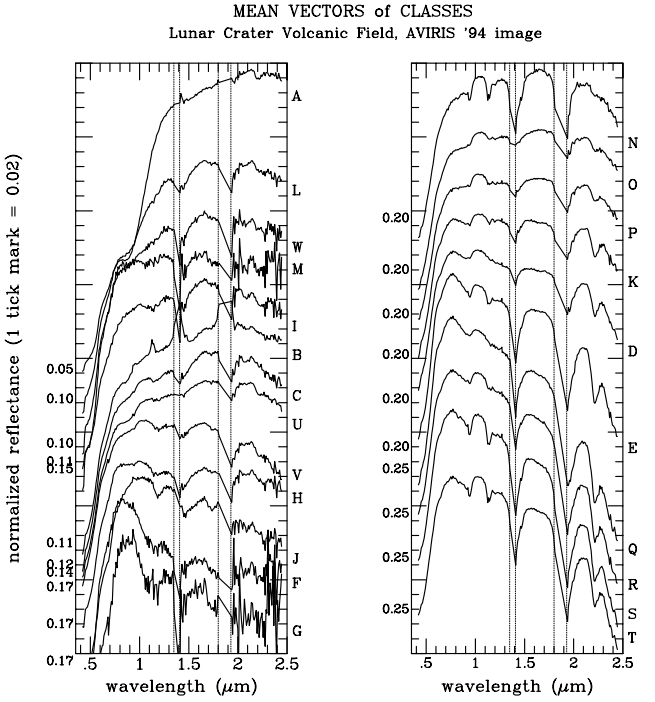

Fig. 6. Average spectra and labels for the 23-class problem. Spectra are offset for viewing convenience. Dashed vertical lines indicate data fallout due to the saturation of the atmospheric water bands.

We use a jack-knifing strategy to assess training and testing classification accuracy during the learning process. We perform three jack-knife runs randomly selecting two-thirds of our known labeled training sample pool for training (621/976 samples for the 23/35-class problems) and one-third for testing (310/488 samples for the 23/35-class problems). We report the average of the three jack-knife results for the testing samples. Accuracy in each run is computed as the average of individual class accuracies, which allows relatively small classes to equally contribute to the assessment of the classifier.

IV. Results

A. The 23-class LCVF classification as a baseline

Our previous GRLVQI processing in the wavelet domain, using a 3-level CSDWT with the db4 filters [7], achieved an accuracy of 97.3% with 17 wavelet coefficients having relevances ≥ 0.001. We extend that experiment to a 4-level and 5-level wavelet transform using the CSDWT as well as the DT$\mathbb{C}$WT in the following sections.

B. GRLVQI & $\mathbb{C}$omplex Wavelets: An initial investigation

Our initial hypothesis was that GRLVQI processing on the magnitude of the DT$\mathbb{C}$WT would yield best results in the face of discontinuities in the data (and, consequently, artifacts in the wavelet coefficients). However, based on Table I it does not appear advantageous over the CSDWT [7]. Surprisingly, the imaginary component of the DT$\mathbb{C}$WT (especially the 4-level transform) shows superior results. We use the imaginary component of the DT$\mathbb{C}$WT and do GRLVQI processing and feature extraction on that representation for the remainder of this study.

		Real	Imaginary	Magnitude	Phase
3-level	Acc (%)	97.35%	97.44%	95.99%	74.56%
DT$\mathbb{C}$WT	# Features	55	16	51	70
4-level	Acc (%)	95.13%	98.28%	83.50%	69.97%
DT$\mathbb{C}$WT	# Features	15	11	95	75

TABLE I

ACCURACY AND NUMBER OF FEATURES FROM GRLVQI IN THE DT$\mathbb{C}$WT DOMAIN. FEATURES WITH RELEVANCES ≥ 0.001 WERE COUNTED.

As seen from Table II, the imaginary component of the 4-level decomposition from the complex wavelet transform achieves best results. For the 4-level DT$\mathbb{C}$WT, a classification accuracy of 98.0% is achieved while retaining 15 (7.2%) of the DT$\mathbb{C}$WT wavelet coefficients. Similar classification accuracy is achieved with the 4-level CSDWT requiring 24 (11.5%) of the wavelet Coefficients.

	CSDWT			DT$\mathbb{C}$WT imaginary part		
	3-level	4-level	5-level	3-level	4-level	5-level
Accuracy	97.3%	98.2%	97.0%	97.9%	98.0%	96.9%
Features	17	24	27	18	15	30

TABLE II

23-CLASS PROBLEM: ACCURACY AND NUMBER OF FEATURES FOR GRLVQI PROCESSING IN THE WAVELET DOMAIN. FEATURES WITH RELEVANCES ≥ 0.001 WERE COUNTED.

C. The strength of the DT$\mathbb{C}$WTs imaginary component

When projecting a relatively smooth function on even wavelet bases, smooth signal information is preserved which is indicated by a smoothed version (or coarse approximation) of the original signal $f(t)$. In this case, much of the detailed information (high-frequency content such as edges) are suppressed making it difficult to pinpoint the exact location of the edges. In contrast, odd wavelet bases preserve edges. The projection of a relatively smooth function onto odd wavelet bases has coefficients that are suppressed, thus accentuating the differences at the discontinuities and making singularity identification much easier.

Discriminating absorption bands in hyperspectral data are often narrow shapes with sharp boundaries that must be preserved. Symmetric (even) wavelet basis functions smooth such absorptions making it more difficult to retain critical distinguishing features. In contrast, odd wavelet basis functions preserve (and more precisely locate) them. The consequence is better classification accuracy with the imaginary component than with the real component or with the magnitude of the complex wavelet and scaling coefficients.

What about the discontinuities in the spectra due to deleting spectral information? Since the same discontinuities exist in all classes, the odd wavelet basis functions are able to accurately (and consistently) locate them, and since GRLVQI learns the differences between classes, these consistent false features are ignored.

D. A more difficult case: the 35-class problem

To further contrast the effectiveness of the imaginary component of the DT$\mathbb{C}$WT over the CSDWT with the db4 filters, we extend our study to the more difficult 35-class problem discussed briefly in Section III-A.

Classification accuracies and the number of retained features are tabulated in Table III for the CSDWT and the imaginary component of the DT$\mathbb{C}$WT. The 4-level wavelet decomposition has best results for the DT$\mathbb{C}$WT (allowing that 0.5% accuracy difference between the 4-level and 3-level results is "in the noise" and considering the further reduction in retained features significant).

For this problem, we achieve a 96.9% classification accuracy while retaining 16 (7.7%) coefficients. The classification accuracy improved by 1.4% with 16 coefficients, compared to 26 required by GRLVQI processing on the (4-level) CSDWT representation.

	CSDWT			DT$\mathbb{C}$WT imaginary part		
	3-level	4-level	5-level	3-level	4-level	5-level
Accuracy	95.7%	95.6%	95.4%	97.4%	96.9%	95.9%
Features	26	26	33	21	16	22

TABLE III

35-CLASS PROBLEM: ACCURACY AND NUMBER OF FEATURES FOR GRLVQI PROCESSING IN THE WAVELET DOMAIN. FEATURES WITH RELEVANCES ≥ 0.001 WERE COUNTED.

E. Independent assessment of features extracted by GRLVQI

We use a Minimum Euclidean Distance classifier for independent assessment of the discrimination capability of the features that GRLVQI deemed relevant. We take the features (wavelet coefficients) in descending order of their relevance and classify the test data, in the wavelet domain, as each new feature is added. We start with the wavelet coefficient associated with the largest relevance and classify using that single coefficient. We add to this first coefficient, the wavelet coefficient that corresponds to the second largest relevance and then classify using the two coefficients. This process continues until all wavelet coefficients are used. We record the best classification and the number of wavelet coefficients required to obtain that classification. Table IV compares, for the 23-class problem, the CSDWT and the imaginary component of the DT$\mathbb{C}$WT. As a baseline comparison, MED classification with all available features achieves 92.8% classification accuracy for the 23-class problem (not shown in Table IV).

Three general observations can be made from Table IV. First, the GRLVQI-selected features (wavelet coefficients) lead to a better classification than the use of all available features. Second, the wavelet representation by the imaginary component of the DT$\mathbb{C}$WT yields consistently similar classification results to the CSDWT with the db4 filters. Third, fewer features are required for best MED performance with the DT$\mathbb{C}$WT, than with the CSDWT coefficients.

Using the MED classifier to evaluate the quality of GRLVQI-selected features is admittedly not the best approach.

	CSDWT			DT$\mathbb{C}$WT imaginary part		
	3-level	4-level	5-level	3-level	4-level	5-level
Accuracy	95.7%	96.2%	95.8%	96.2%	95.7%	96.2%
Features	17	14	14	12	10	14

TABLE IV

CLASSIFICATION ACCURACY AND CORRESPONDING NUMBER OF FEATURES FOR THE MINIMUM EUCLIDEAN DISTANCE CLASSIFIER FOR THE 23-CLASS PROBLEM.

We use it in this study for its value of providing an easy and quick (albeit coarse) initial sanity check. An LVQ type classifier uses multiple prototype vectors per class to define class boundaries. Prototype vectors learn local class structure (the margins between adjacent classes), thus the computed relevances of features reflect more localized differences. In contrast, the MED uses one prototype vector, the mean of the training data for each class, to discriminate between classes in a more global sense. Because of this one can hardly expect the same classification accuracies from the MED and the GRLVQI, working with features extracted by the QRLVQI. Other apparent inconsistencies such as less number of features used by the MED in some cases (Table IV vs. Tables I - II), can also be ascribed to the above differences of the two classifiers. More meaningful (relevance- like) features for the MED would reflect differences in the mean between neighboring classes.

Based on the arguments above, the relevances computed by GRLVQI may not identify a set of features that can be used in any classifier with the same success (classification accuracy) as in GRLVQI itself. This in turn affects the reliability of the determination of the required number of features for peak classification accuracy. While the MED assessment in Table IV provides a first-order confirmation of the GRLVQ-derived results, a better method for independent evaluation of GRLVQI-selected features could be to use a classifier more similar to the GRLVQI, for example, a K-means classifier where K is assigned the number of GRLVQI prototype vectors used per class. We plan to address this issue as well as the approach to more principled and reliable determination of the necessary number of relevance-assigned features, in subsequent work.

V. SUMMARY AND DISCUSSION

In this paper, we investigated the application of GRLVQI in the complex wavelet domain. While investigating the DT$\mathbb{C}$WT to mitigate the effects of discontinuities introduced in the spectra, we found that the imaginary component (the wavelet bases with odd symmetry) performed better than the magnitude, real component, or phase. Since discriminating features are often sharp bumps and dips in the spectra, the odd symmetry of the imaginary component of the DT$\mathbb{C}$WT accentuates their representation in the wavelet domain making it easier for GRLVQI to focus on these important features. This results in better classification and feature reduction performance.

We found that GRLVQI more easily distinguishes difficult class structure with the odd symmetry filters than using Daubechies length-four orthogonal filters (which are neither even or odd functions) as indicated by classification accuracy gains or a reduction in number of significant relevant features, or both. Additional research is necessary to determine the effects of the odd wavelet bases over those wavelet bases that are neither even or odd. Further, we surmise that careful selection of the employed wavelet system, intelligent preprocessing of the data (see, e.g., Johnson [15]), and more principled determination of the number of required features, could yield more consistent accuracy and feature reduction.

ACKNOWLEDGMENT

The authors would like to thank Dr. Bruce Johnson for sharing his philosophies, insights, and expertise in wavelet analysis. Helpful private communications with Dr. Richard G. Baraniuk are also greatly appreciated.

MM is supported by the Air Force Institute of Technology, Wright-Patterson AFB, OH.

EM is partially supported by grant NNG05GA94G from the Applied Information Systems Research Program, NASA, Science Mission Directorate.

REFERENCES

[1] B. Hammer and T. Villmann, "Generalized relevance learning vector quantization," *Neural Networks*, no. 15, pp. 1059–1068, 2002.

[2] M. J. Mendenhall and E. Merényi, "Relevance-based feature extraction for hyperspectral images," *IEEE Trans. on Neural Networks*, Mar 2006, Submitted.

[3] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Processing Magazine*, pp. 123–151, Nov 2005.

[4] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Springer-Verlag Berlin Heidelberg, 2001.

[5] J. M. Shapiro, "Embedded image encoding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[6] T. Moon and E. Merényi, "Classification of hyperspectral images using wavelet transforms and neural networks," in *Proc. of the SPIE: Wavelet Applications in Signal and Image Processing III*, A. F. Laine, M. A. Unser, and M. V. Wickerhauser, Eds., vol. 2569, Sept. 1995, pp. 725–735.

[7] M. J. Mendenhall and E. Merényi, "Generalized relevance learning vector quantization for classification-driven feature extraction from hyperspectral data," in *Proc. American Society for Photogrammetry and Remote Sensing, Reno, NV*, May 1–5, 2006.

[8] X. Zhang, N. H. Younan, and C. G. O'Hara, "Wavelet domain statistical hyperspectral soil texture classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 615–618, Mar 2005.

[9] J. B. Campbell, *Intro. to Remote Sensing*, 2nd ed. Guilford Press, 1996.

[10] I. Daubechies, *Ten Lectures on Wavelets*, ser. CBMS-SNF Regional Conference Series on Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992, vol. 61.

[11] A. Sato and K. Yamada, "Generalized learning vector quantization," in *Advances in Neural Information Processing Systems 8: Proceedings of the 1995 Conference*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, pp. 423–429.

[12] D. DeSieno, "Adding a conscience to competitive learning," in *Proceedings of the IEEE International Conference on Neural Networks* **I**, New York, Jul 1988, pp. 117–124.

[13] R. O. Green, "Summaries of the 6th Annual JPL Airborne Geoscience Workshop, 1. AVIRIS Workshop," Pasadena, CA, Mar 4–6 1996.

[14] E. Merényi, "Precision mining of high-dimensional patterns with self-organizing maps: Interpretation of hyperspectral images," in *Quo Vadis Computational Intelligence: New Trends and Approaches in Computational Intelligence. Studies in Fuzziness and Soft Computing.*, P. Sincak and J. Vascak, Eds., vol. 54. Physica-Verlag, 2000, available: http://www.ece.rice.edu/~erzsebet/publications.html.

[15] B. Johnson, "Multiwavelet Moments and Projection Prefilters," *IEEE Trans. on Signal Processing*, vol. 48, no. 11, pp. 3100–3108, Nov 2000.