

GENERALIZED RELEVANCE LEARNING VECTOR QUANTIZATION FOR CLASSIFICATION-DRIVEN FEATURE EXTRACTION FROM HYPERSPECTRAL DATA

Michael J. Mendenhall

Erzsébet Merényi

Department of Electrical and Computer Engineering

Rice University

Houston, TX 77005

mendenmi@rice.edu

erzsebet@rice.edu

ABSTRACT

Remotely sensed hyperspectral images provide the fine details needed to delineate many diverse material classes. However, the richness of the data comes at the expense of large volume. A natural question is whether one can select specific spectral channels, while discarding the rest, without adversely affecting the discovery or classification potential of the imagery. In a supervised classification scenario, the classes of interest are known ahead of time, and the opportunity for feature extraction arises. We must demand that classification accuracy go uncompromised as a result of the feature extraction process – a difficult task given the highly correlated nature of hyperspectral data, and the potentially large number of classes. Well established analytical tools based on PCA and wavelets give information regarding signal content. However, feature extraction models based on these methods often judge feature importance by the magnitude of their coefficients, rarely leading to an appropriate set of features for classification. We investigate an improved version of the recent neural paradigm Generalized Relevance Learning Vector Quantization (GRLVQ) (Hammer, 2002). GRLVQ extends Learning Vector Quantization (LVQ) (Kohonen, 2001) to optimally select input dimensions relevant for classification. Our improvements to GRLVQ (GRLVQI) stabilize it for high-dimensional data and increase classification accuracy and convergence rate (Mendenhall, 2006). We couple the power of GRLVQI with wavelet analysis to produce as good or higher classification accuracy as with GRLVQI alone applied in the spectral domain, while significantly reducing the required number of features. We demonstrate this by classifying 224-band AVIRIS spectra into 23 material classes.

Disclaimer: 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.

INTRODUCTION

Remotely sensed hyperspectral images, by virtue of instrument design, have hundreds of spectral bands that provide the fine spectral resolution needed to discriminate among a vast variety of surface materials. This poses a challenge for many-class classifications with high accuracy demand. Hundreds of spectral bands, many of which are often highly correlated, make it difficult to produce high-quality classification for a large number of surface covers. In addition, the compositional detail is at the cost of high data volume: hyperspectral images are often greater than 100 Mbytes in size. For supervised classification where classes are predetermined (as opposed to the discovery of clusters in a data mining situation), extraction of an appropriate feature set brings with it the potential of improving classification accuracy if important signal information is preserved. Other possible advantages of feature extraction can be processing efficiency, reduced storage requirements, and better utilization of bandwidth for image transmission.

Several previous works address feature extraction and classification performance utilizing wavelets as a feature space. Some methods select features based on the largest magnitude wavelet coefficients, a method that minimizes the distortion of the reconstructed signal often seen in wavelet-based image compression. Moon and Merényi (Moon, 1995) use this form of wavelet feature extraction for remotely sensed hyperspectral images where the features are in turn used as the input to a hybrid neural network classifier. In the real-world 13-class AVIRIS (Airborne Visible/Infrared Imaging Spectrometer, NASA/JPL) (Green, 1996) image they evaluate, they find no

consistent relationship between the number of retained wavelet coefficients and the achieved classification accuracy. A reasonable question to ask is whether or not these results were due to the classifier or the inappropriateness of the feature set for classification. Presumably it is the feature set since the hybrid neural architecture performs well with all available (158) spectral channels for the above image (Moon, 1995), and for other hyperspectral classifications (Merényi, 2000; Merényi, 1996).

Zhang et al. (Zhang, 2005) use wavelet sub-band energy to characterize hyperspectral signatures. They cascade the energy features to a maximum likelihood classifier for a 3- and 12-class soil texture problem. This combination of feature extraction and classification is successful for classifying the easier 3-class problem of major soil texture types, but has a difficult time with the 12-class problem, where each major soil texture type has four subclasses. Degradation of classification performance likely indicates that sub-band energy as a feature set is inadequate to distinguish the subclass structure found in their 12-class problem.

There are numerous methods for feature extraction and classification. However, many of them (including those described thus far) suffer from the same problem: feature extraction is not optimized for classification. Oehler and Gray (Oehler, 1995) use Learning Vector Quantization (LVQ) to find the optimal tradeoff between a Bayes risk and a squared error term. This method requires a parametric model for the posterior class probabilities. Since parametric models for remotely sensed hyperspectral images are not readily available, we must look elsewhere. Hammer and Villmann use an adaptive diagonal metric in their Generalized Relevance Learning Vector Quantization (GRLVQ) to learn which input dimensions are most important for classification (Hammer, 2002). We improve the classification accuracy and convergence rate of GRLVQ with our GRLVQ-Improved (GRLVQI) in an earlier work (Mendenhall, 2006). GRLVQ, and consequently GRLVQI too, offers the power of an adaptive diagonal metric that indicates which dimensions are important for classification combined with the robustness an LVQ classifier has towards noisy and incomplete data.

LEARNING VECTOR QUANTIZATION

Kohonen's Learning Vector Quantization (LVQ) (Kohonen 2001) is a supervised neural learning paradigm that iteratively adjusts the quantization prototypes to define class boundaries. Variants of LVQ based on LVQ2.1 *differentially shift* a best matching prototype with the same label as the input sample, and a best matching prototype with a different label than the input sample, to define classification boundaries while minimizing the Bayes risk.

Generalized Learning Vector Quantization (GLVQ) developed by Sato and Yamada (Sato, 1996) tackles a serious divergence problem with LVQ2.1. Specifically, the LVQ2.1 formulation allows prototype vectors to migrate to, and then drift from, their optimal locations with subsequent training. Sato and Yamada solve this issue by incorporating classification accuracy in the cost function which is minimized via gradient decent.

Hammer and Villmann (Hammer, 2002) expand further upon GLVQ by incorporating an adaptive diagonal metric which learns an importance weighting of each input dimension. This importance weighting is the so-called relevance. One may describe GRLVQ as a classification-driven feature extraction algorithm. Two constraints are placed on the relevance factors: they must be non-negative and they must sum to one. These constraints may allow relevance to have a nice interpretation as a probability.

Our analysis of GRLVQ reveals two critical issues. The first is the single winning prototype problem. The second is the potential for diverging prototype vectors. We address the first issue by incorporating an equiprobabilistic winner selection strategy ensuring that all prototype vectors are utilized in the learning process. In the closing remarks of (Hammer, 2002) Hammer and Villmann suggest that such an equiprobabilistic (or maximum entropy) approach could bring improvements to GRLVQ. We address the second issue by an intuitive modification to the prototype update rule. Our changes improve the generalization capability as well as the convergence rate of the classifier. We dub our improved form of GRLVQ as GRLVQ-Improved (GRLVQI). The interested reader is referred to (Mendenhall, 2006) for further algorithmic details and experimental results. In (Mendenhall, 2006) we point out that although GRLVQI shows superior classification performance for real-world hyperspectral imagery, the highly correlated spectral bands appear to limit the achievable dimensionality reduction.

CONSIDERING A TRANSFORM DOMAIN

Given that the highly correlated nature of the spectral bands may be a limiting factor for further feature reduction using GRLVQI, one might consider a transformation of the data. The goal is not to extract features prior to

GRLVQI processing, but rather to transform the spectral data into a different feature space and do GRLVQI processing in that feature space. There are several good transforms, each with special properties. When considering a transform, we should choose one that decorrelates the transform coefficients. A sparse transform may also be desirable given that GRLVQI chooses specific coefficients for classification. Further, an efficient transform is desired so as not to hinder the potential near real-time uses of GRLVQI processing. An obvious choice, given these requirements, is the wavelet transform. The wavelet transform decorrelates transform coefficients, provides a sparse representation of the data, and it is computable in linear time. What is unclear at this time, based on the limited success of earlier works, is the quality of the wavelet feature space for classification.

The remainder of this publication studies the feasibility of using GRLVQI to select important wavelet coefficients, rather than the coefficients with the largest magnitude, for the task of classification. The unique aspect of this wavelet feature extraction method is that wavelet coefficient selection is linear. More typical wavelet processing algorithms select as features the largest magnitude wavelet coefficients which is a non-linear process.

APPLYING GRLVQI IN THE SPECTRAL AND WAVELET DOMAINS

We compare the selection of spectral features and wavelet coefficients using GRLVQI on real-world hyperspectral data obtained by AVIRIS). We demonstrate the quality of the reduced feature set by comparing classification accuracy, produced by an independent classifier, using only those features discovered by GRLVQI versus using all available spectral features.

Data

We use carefully verified selected spectra from the Lunar Crater Volcanic Field (LCVF) scene acquired by AVIRIS in 1994 (Figure 1). These samples, which served as a training set in an earlier classification study (Merényi, 2000) are divided into training and test sets for this work as described in the “Design of Experiments” section below. After atmospheric correction and elimination of the saturated water bands containing irrecoverable data, our hyperspectral data has 194 spectral bands. Each spectrum (194-dimensional vector) is then normalized to unit length by dividing the vector elements by the Euclidean vector norm. This cancels linear effects such as shading resulting from viewing geometry, making the spectral classes much more uniform while retaining the differences in spectral characteristics. It also eliminates the differences in geometric albedo for materials that have the same spectral signature and only differ in their albedo (Merényi, 1996), which is an undesirable effect. Fortunately this is rarely the case, but one should be aware of the possibility and, if necessary, do post-classification processing to separate such materials.

These data contain 23 surface cover classes, a description of which is provided in Table 1, along with the class labels and the number of training samples per class. Figure 1 shows sample locations for each surface material. Mean spectra of the 23 classes under evaluation are shown in Figure 2. Further details of this 23-class data set can be found in (Merényi, 2000).

Design of Experiments

Three splits of the data shown in Table 1 are generated using two-thirds for training (621 samples) and one-third for testing (310 samples). Each split is a random selection of the known training samples. Results are reported as the average of the classification accuracies on the *test* set from each of three jack-knife runs. In each run, classification accuracies are calculated as the mean of the individual class accuracies. Table 2 lists the learning parameters including the conscience parameters we used in these GRLVQI experiments.

Initial prototypes are drawn randomly from a uniform distribution on $[0.4, 0.6]$, data are scaled to the interval $[0, 1]$ and each class is assigned five prototypes.

In our experiments, we use Daubechies 4-tap orthogonal filters. We iterate the wavelet transform three times generating a three-level wavelet decomposition of the spectral data. The goal of this paper is to prove a concept. As such, we do not set out to determine the optimum set of filters (orthogonal or bi-orthogonal, even-length or odd-length, long filters or short filters) nor do we attempt to find the optimum number of decomposition levels for classification. Filter selection and the number of decomposition levels of the wavelet transform are both important considerations which can potentially have an effect on our ability to classify the data, and should be explored in follow up work.

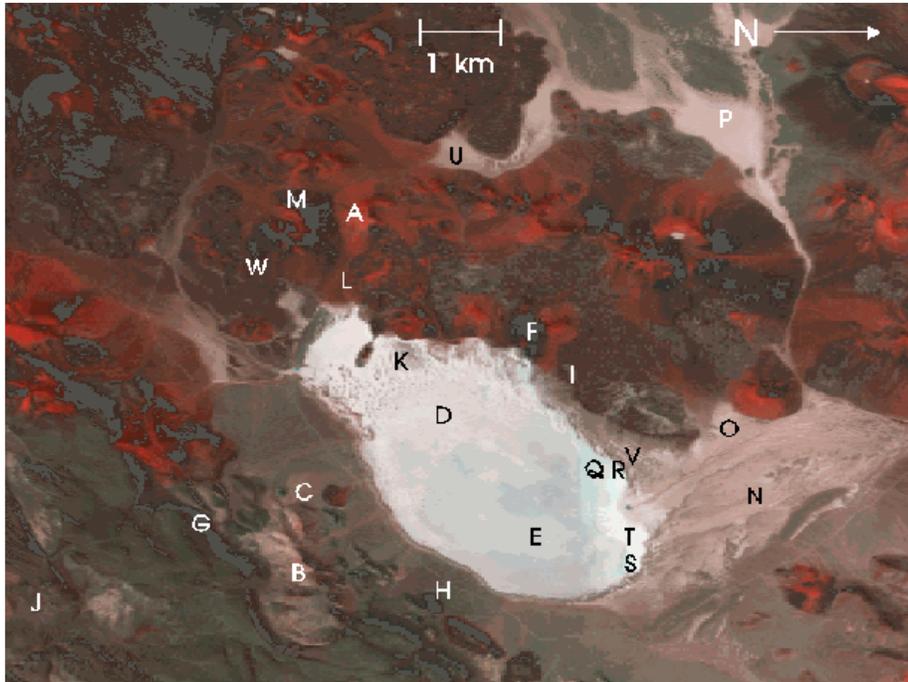


Figure 1: A natural color composite of the Lunar Crater Volcanic Field (LCVF), NV obtained in 1994 by the NASA/JPL AVIRIS imaging spectrometer. Labels indicate locations of various surface cover types described in Table 1. Further details and previous analysis of this image are given in (Merényi, 2000).

Class	Cover type description	# tr	Class	Cover type description	#tr.
A	Hematite-rich cinders	72	M	Alluvium #3 (iron rich)	14
B	Rhyolite of Big Sand Spring Valley	22	N	Dry wash #1	15
C	Alluvium #1	50	O	Dry wash #2	54
D	Dry playa	160	P	Dry wash #3	45
E	Wet playa #1	115	Q	Wet playa #2	15
F	Young basalt	21	R	Wet playa #3	14
G	Shingle Pass tuff	7	S	Wet playa #4	15
H	Alluvium #2 (with mixed scrub Brush, rocks, and soil)	50	T	Wet playa #5	18
I	Old basalt	36	U	Alluvium #4 (also rich iron)	36
J	Dense scrub brush stands	12	V	Wet playa #6	12
K	Basalt cobbles on playa	37	W	Ejecta blankets#2 (primarily unoxidized cinders with smaller Percentage of hematite-rich cinders)	33
L	Ejecta blankets #1 (mixed hematite- Rich and unoxidized cinders)	78			
				Total number of training samples	931

Table 1: Class labels, descriptions, and the number of training samples (#tr) for the 23classes of the LCVF data set.

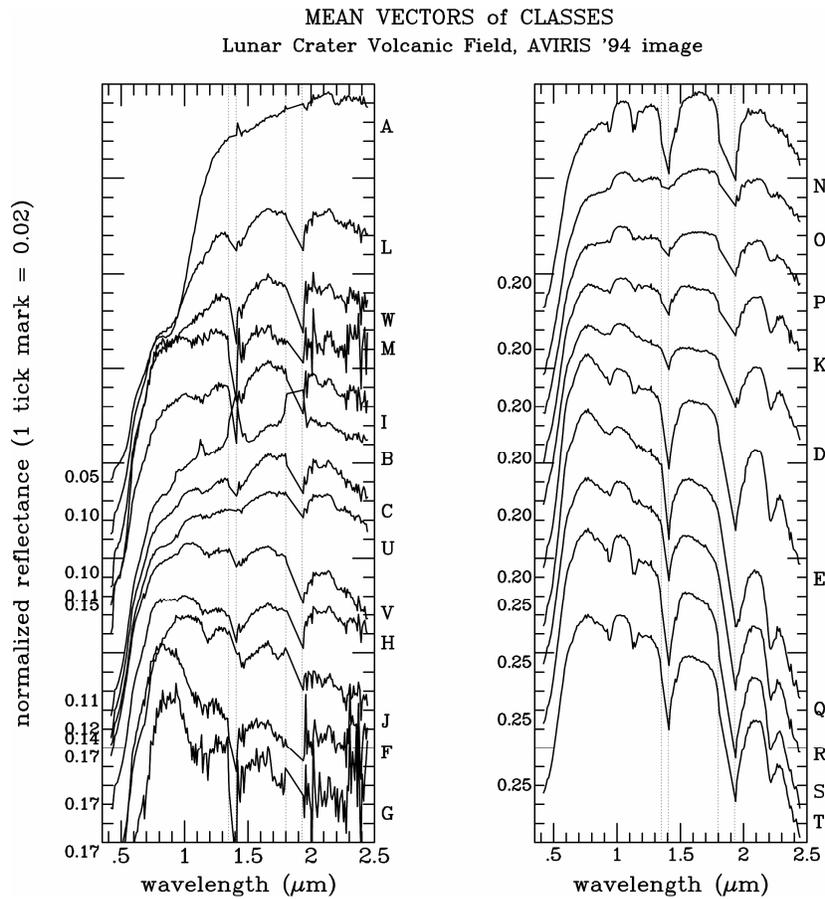


Figure 2: Means of training spectra corresponding to the classes listed in Table 1. Spectra are offset vertically for viewing convenience. Dashed vertical lines indicate data fallout due to saturation of the water bands.

Schedule of GRLVQI Learn Parameters

Training Steps	Learn Parameters			Conscience Parameters	
	ϵ^λ	ϵ^J	ϵ^K	γ	β
$0 < TS \leq 311K$	0.005	0.025	0.025	2	0.35
$311K < TS \leq 621K$	0.0025	0.0025	0.0025	2	0.3
$621K < TS \leq 932K$	0.001	0.001	0.001	2	0.225
$932K < TS \leq 1,242K$	0.0005	0.0025	0.0025	2	0.125

Table 2: A list of the learn rates and conscience parameters for the 23-class problem. The number of training steps is rounded to the nearest thousand. The parameters ϵ^λ , ϵ^J , and ϵ^K are related to the rate of adjustment for relevance factors, in-class prototype vectors, and out-of-class prototype vectors respectively, as described in (Hammer, 2002). The parameters γ and β control the “conscience” in the selection of the winning in-class prototype vector as described in (DeSieno, 1988).

Experiments were designed with two goals in mind. First, we want to evaluate the quality of the spectral and wavelet feature spaces (processes **E** and **F** in Table 3) for the purpose of classification. Second, we want to evaluate the quality of the extracted features which we accomplish using the Minimum Euclidean distance (MED) (processes **A**, **B**, and **D** in Table 3). The MED also allows us to impartially compare more typical largest magnitude wavelet coefficient selection as a feature set (process **C** in Table 3) to GRLVQI selected spectral features and GRLVQI selected wavelet coefficients (processes **B** and **D** respectively).

Description of experiment processes
A – Benchmark MED classification with all available spectral features
B – MED classification of the GRLVQI extracted spectral features
C – MED classification of largest magnitude wavelet coefficients
D – MED classification of the GRLVQI extracted wavelet coefficients
E – Classification in the wavelet domain using GRLVQI as the classifier
F – Classification of the spectral features using GRLVQI as the classifier

Table 3: List of simulations run to compare features discovered by GRLVQI from the spectral data and from the wavelet coefficients.

Results

To facilitate the discussion of our results, we will refer to the relevance factors obtained by GRLVQI processing in the spectral domain as “spectral relevance factors.” Similarly, we will refer to the relevance factors obtained by GRLVQI processing in the wavelet domain as “wavelet relevance factors.”

Table 4 shows that although GRLVQI in the wavelet feature space (process **E**) yields only slightly better classification accuracy than GRLVQI in the spectral feature space (process **F**), executing GRLVQI in the wavelet domain requires significantly fewer features than GRLVQI in the spectral domain. This matches our expectation, due to wavelet transform properties (namely decorrelation and sparseness).

The number of spectral features required to achieve the classification accuracy reported in process **B** (Table 4) is determined by progressively adding GRLVQI selected spectral features and tabulating the MED classification accuracy on the test sample set. We start by selecting the single spectral feature corresponding to the largest spectral relevance factor, and classifying the test sample set with the MED classifier using this single spectral feature and tabulating the result. Then we add to the first spectral feature the spectral feature that corresponds to the second largest spectral relevance factor. We classify the test sample set using these two spectral features and tabulate the results. This process continues by adding spectral features one at a time in descending order of spectral relevance factors and report the number spectral features that gives the maximum classification accuracy with the MED classifier. The number of wavelet features required to achieve the classification reported for process **D** in Table 4 is obtained in the same manner using wavelet relevance factors to select the wavelet coefficients. We take the inverse wavelet transform on the retained wavelet coefficients prior to MED classification.

The number of largest magnitude wavelet coefficients required to achieve the classification accuracy reported in Table 4 for process **C** is determined in a manner similar to the relevance selection of spectral and wavelet features described in the previous paragraph. We start by selecting a single wavelet coefficient that has the largest magnitude, take the inverse wavelet transform of the test sample set using this single wavelet coefficient, and then classify the reconstructed spectral signature using the MED classifier and tabulate the result. Next we add, to the first wavelet coefficient, the wavelet coefficient with the second largest magnitude. We perform the inverse wavelet transform of the test sample set using these two wavelet coefficients, classify the reconstructed spectral signatures using the MED classifier and tabulate the result. This process continues by adding wavelet coefficients one at a time in descending order of their magnitudes. For process **C** in Table 4, we report the number of wavelet coefficients that corresponds to the maximum classification accuracy using the MED classifier.

Our validation using an independent classifier shows that we can achieve more than 3% improvement in classification accuracy using the 76 relevance selected spectral features (process **B**) discovered with GRLVQI over the benchmark using all (194) spectral features (process **A**).

Selection of the largest magnitude wavelet coefficients as features (process **C**) may have an advantage over GRLVQI on the original spectral data (process **E**) if the application can tolerate a small degradation in classification accuracy. Classification accuracy is slightly worse (less than 2%), but achieved with less than half of the features!

Using GRLVQI selected wavelet coefficients produces the best results (process **D**). Classification accuracy is only slightly better than when using the GRLVQI selected spectral features (process **B**), but with less than 19% of

the coefficients. The story is similar with largest magnitude wavelet coefficient selection (process C). We boast an increase in classification accuracy of 2.7% with nearly 60% fewer coefficients. These results clearly demonstrate the potential of our relevance wavelet feature extraction and classification model.

Summary of Classification Results

	A	B	C	D	E	F
Classification Accuracy	91.9%	95.3%	93.1%	95.8%	97.3%	97.0%
Retained Features	194	76	37	15	36	81

Table 4: Classification accuracy and the corresponding number of significant relevance factors or wavelet coefficients for the six simulations listed in Table 3. The accuracies shown are computed as the average of three jack-knife runs. In each run, the overall accuracy is calculated as the average of the individual class accuracies on the test data.

In Figure 3 we show representative spectra and spectral relevance factors for the 23 classes we evaluate. Three primary regions are discovered as relevant for classification by GRLVQI. The first cluster of spectral relevance factors appears in the range of 0.45 μm to 0.68 μm . A second significant cluster appears between 0.7 μm and 1.1 μm , and a third cluster shows from 1.4 μm to 1.7 μm , although sparse around 1.5 μm . Two major regions, 1.1 μm to 1.4 μm and 1.7 μm to 2.5 μm , receive little attention from GRLVQI indicating that these regions do not provide significant information for the correct classification of the samples in our 23-class problem.

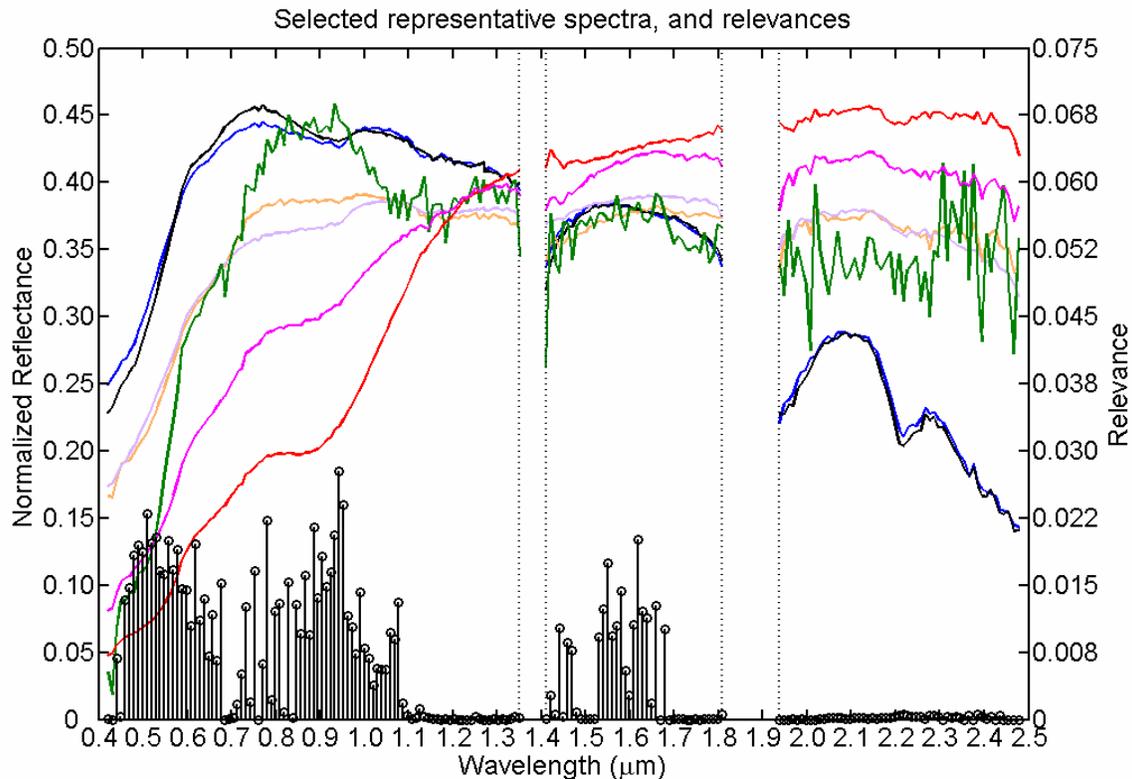


Figure 3: Average spectra of representative classes A (red), G (green), H (orange), L (magenta), O (purple), Q (black), and R (blue) and spectral relevance factors (black stem plot.) Relevance factors are the averages of three jack-knife runs. Classes were selected to show largest diversity. The dotted vertical lines indicate data fallout due to saturation of the water bands.

In Figure 4, we show a breakdown of the wavelet transformed spectra indicating the order in which the wavelet filters are applied (read left to right). Here *H* indicates that highpass filtering was performed on the data and *L* indicates lowpass filtering. As an example, *LH* means that first a lowpass filter was applied, then a highpass filter. The process is to extend the length of the original vector such that it is divisible the 2^L where *L* is the number of times the signal will be decomposed using the wavelet transform. We chose to perform a three level decomposition,

so our spectral dimension needs to be divisible by $2^3 = 8$. We extend the length of our 196-dimensional spectra to 200 dimensional vectors using zero padding. A highpass filter (H) is applied to the 200-dimensional vector and the result is downsampled by a factor of two yielding detailed signal information. Similarly, a lowpass filter (L) is applied to the 200-dimensional vector and the result is downsampled by a factor of two yielding coarse signal information. We iterate on the lowpass filtered region. This process is performed two more times to yield the three level wavelet decomposition.

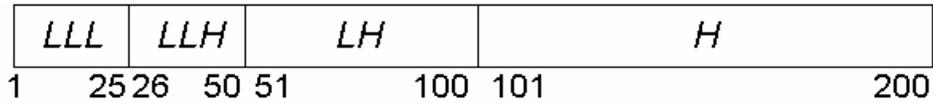


Figure 4: Illustration of the application of the highpass (H) and lowpass (L) filters (top) and the range of wavelet coefficients that apply to the specific filtering sequence (bottom) for a three level wavelet decomposition.

Figure 5 shows representative wavelet curves and the wavelet relevance factors GRLVQI assigns to the wavelet coefficients. We see that GRLVQI discovers that coarse signal information is most important for classifying our 23-class data set. Referencing Figure 4, we see that coarse signal information is contained in those wavelet coefficients which have been lowpass filtered only (coefficient indices 1-25). GRLVQI further discovers that some detailed coefficients are also necessary for classifying the 23-class data set. The detailed information comes from the LLH sub-band in Figure 4 (coefficients 26 through 50). We note that a group of large magnitude wavelet coefficients exists in the index range of 90-105, however, no large wavelet relevance factors are associated with this range. The lack of wavelet relevance factors in this region indicates that it is not important for classifying our data set and demonstrates that GRLVQI does not follow the largest magnitude wavelet selection method commonly associated with wavelet-based image processing.

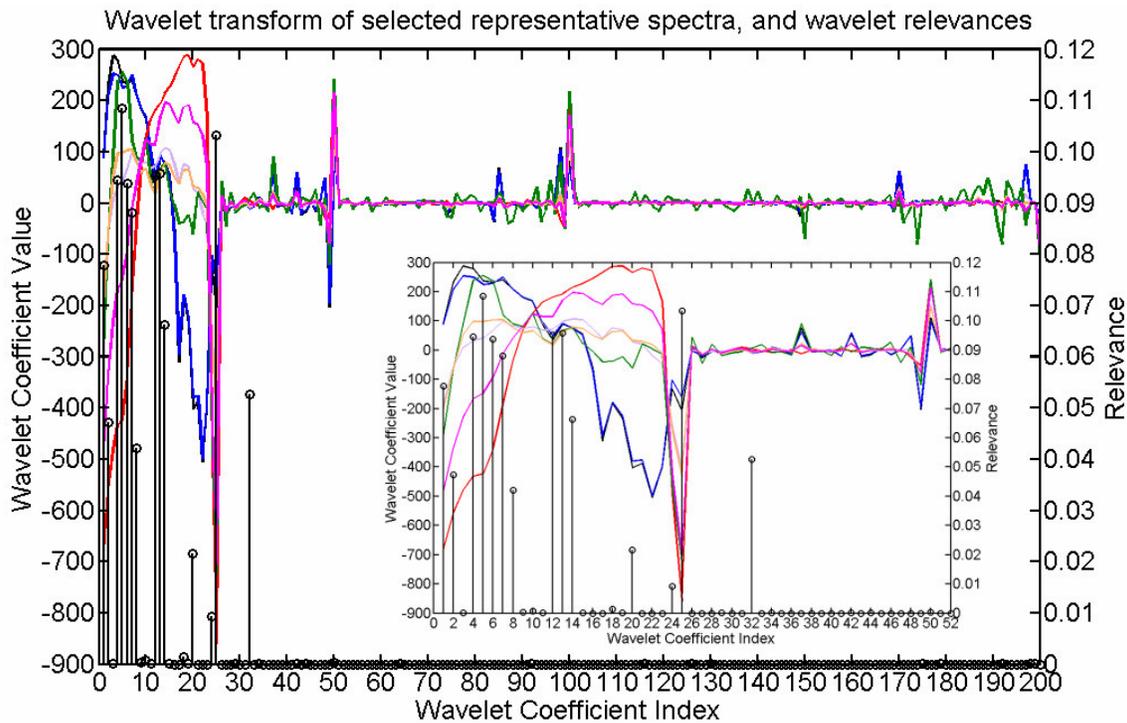


Figure 5: Wavelet transform of class means for classes A (red), G (green), H (orange), L (magenta), O (purple), Q (black), and R (blue) with wavelet relevance factors obtained by GRLVQI (black stem plot) in the wavelet domain. Relevance factors are the averages from three jack-knife runs. Classes were selected to show largest diversity.

In order to compare spectral relevance factors and wavelet relevance factors, we apply the inverse wavelet transform to the wavelet relevance factors. We rescale the inverted wavelet relevance factors using the L_1 -norm. Three observations from Figure 6 are worth noting. First, the highly correlated nature of the inverted wavelet

relevance factors indicate how well the wavelet transform decorrelates the data. Second, GRLVQI in the wavelet domain places greater emphasis in the 1.4 μm to 1.5 μm and the 2.2 μm to 2.3 μm regions than GRLVQI does when processing on the reflectance features. Third, there are some negative values in the inverted wavelet relevance factors. It is unclear at this time how one should interpret the negative relevances.

One might expect GRLVQI to discover that the wavelet indices corresponding to the largest magnitudes or the largest variances over the entire training set are the most important for classification. To dispel the idea that GRLVQI selects coefficients based on largest magnitude or largest variance, we compare them both with the largest wavelet relevance factors in Figure 7. We choose to compare 15 wavelet coefficients because this corresponds to the best classification achieved by the MED classifier (process **D**) when wavelet coefficients are selected using the largest wavelet relevance factors. In Figure 7 (top), the mean of all wavelet signatures is shown as the black curve. The 15 largest magnitude wavelet coefficients are indicated by the red stem plot. The wavelet coefficients corresponding to the 15 largest wavelet relevance factors are indicated by the blue stem plot. The wavelet coefficients that are among both the largest magnitudes and the largest wavelet relevance factors are indicated by the black stem plot. Similarly, the standard deviation of all wavelet signatures is shown as the black curve in Figure 7 (bottom). The 15 largest standard deviations of the wavelet coefficients are indicated by the red stem plot. The wavelet coefficients corresponding to the 15 largest wavelet relevance factors are indicated by the blue stem plot, and the wavelet coefficients included in both the set of largest standard deviations and the set of largest wavelet relevance factors are indicated by the black stem plot. Clearly, although there is overlap in both plots, the largest relevances assigned by GRLVQI to wavelet coefficients are not the same as the wavelets with largest magnitudes or largest variances.

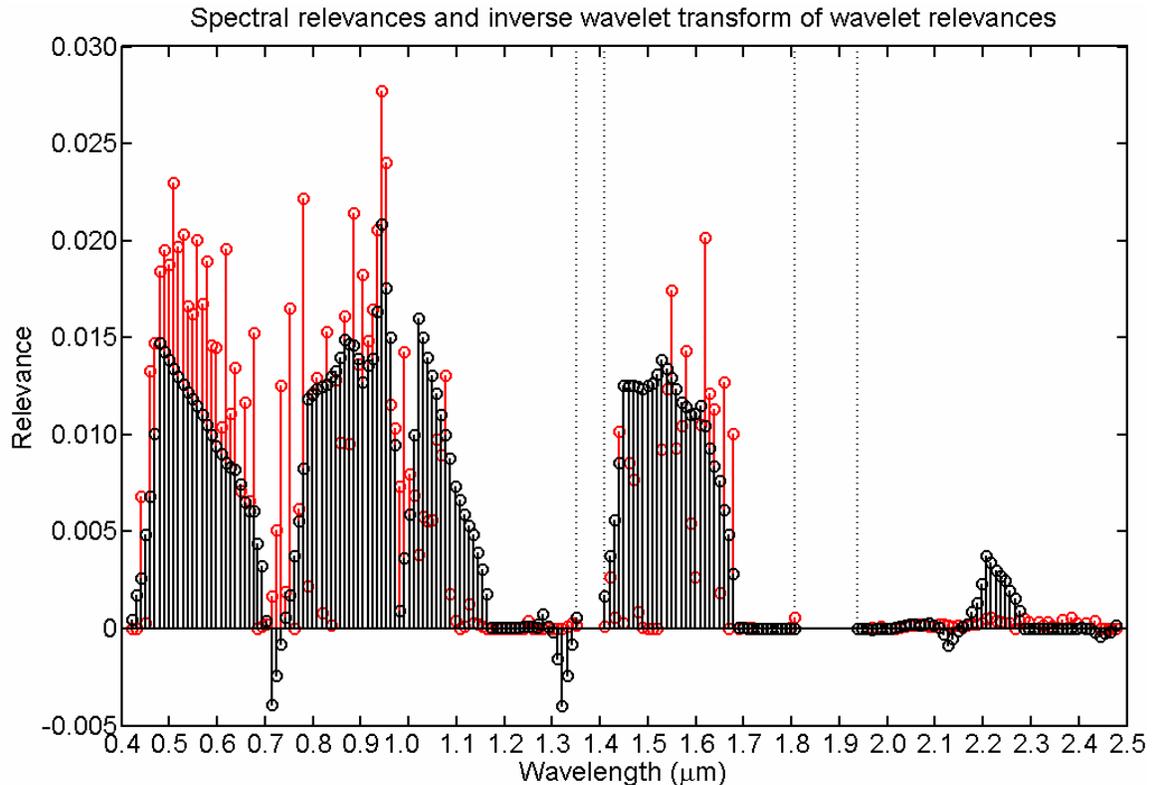


Figure 6: Relevances obtained by GRLVQI in the spectral domain (red) compared to the inverse wavelet transform of the relevances obtained by GRLVQI in the wavelet domain (black). The dotted vertical lines indicate data fallout due to saturation of the water bands.

A Parseval's theorem holds for orthonormal wavelets (Burrus, 1998), therefore we can compare the energy of the relevant spectral features, relevant wavelet coefficients, and largest magnitude wavelet coefficients (Table 4). Total energy is calculated as average pixel energy of the test sample set. For relevance weighted energy, we use the following:

$$E_{\lambda} = \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^n \lambda_i (x_i^m)^2$$

where i is the dimension index, m the training sample index, M the total number of training samples, and λ_i the relevance factor associated with dimension i . Energy calculations for non-weighted coefficients is accomplished by setting $\lambda_i = 1$ for all i .

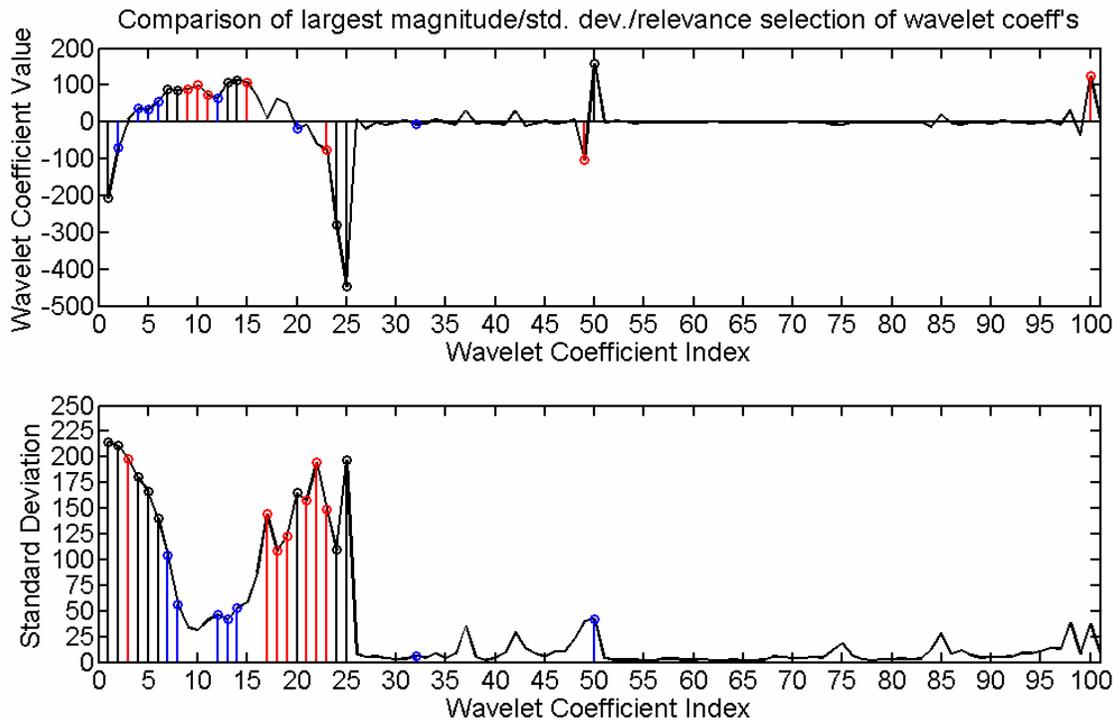


Figure 7: Top) Comparison of the 15 largest magnitude wavelet coefficients (red stem plot) with the 15 wavelet coefficients corresponding to the 15 largest wavelet relevance factors (blue stem plot.) The black curve is the mean wavelet representation for all training samples. The black stem plot shows the indices for which the corresponding wavelet belongs to both red and blue sets. **Bottom)** Comparison of the 15 largest standard deviations of the wavelet coefficients (red stem plot) with the 15 wavelet coefficients corresponding to the 15 largest wavelet relevance factors (blue stem plot.) The black curve is the standard deviation of the wavelet representation of all training samples. The black stem plot shows the indices for which the corresponding wavelet belongs to both red and blue sets.

We discuss the energy of the retained features to show that there is not a consistent relationship between retained energy and achieved classification accuracy. Table 4 shows that keeping all available spectral channels has the most energy yet has the weakest MED classification performance (process **A**). Keeping the spectral features associated with the 76 largest spectral relevance factors (process **B**) has the smallest energy yet is outperformed by keeping the 15 wavelet coefficients associated with the 15 largest wavelet relevance factors (process **D**). Keeping the 37 largest magnitude wavelet coefficients (process **C**) has the second largest energy yet is outperformed by the 76 SF and 15 WF cases.

Average Retained Energy per Pixel for MED Classification

	w/o relevance	w/o relevance	Accuracy
76 SF	$4.99*10^3$	$7.99*10^3$	95.3%
15 WF	$5.88*10^3$	$4.62*10^4$	95.8%
37 WC	$9.45*10^3$	NA	93.1%
194 SF	$2.49*10^7$	NA	91.9%

Table 4: The retained energy for the 76 spectral features (76 SF) corresponding to the 76 largest spectral relevance factors, the 15 wavelet features (15 WF) corresponding to the 15 largest wavelet relevance factors, the 37 largest magnitude wavelet coefficients (37 WC) retained using more typical wavelet processing, and for all spectral features (194 SF).

SUMMARY & DISCUSSION

The wavelet transform is a powerful tool used in data analysis for a broad range of applications. We introduced a new method of wavelet coefficient selection that couples the power of the recently introduced neural paradigm called relevance learning with wavelet analysis. Our new approach results in exceptional classification performance with a significantly reduced feature set.

Following we note several issues that remain open questions as a result of this study, and should be examined in subsequent work.

We do not exclude the possibility that better wavelet representation (e.g., different orthogonal wavelets, odd or even length bi-orthogonal wavelets, or adaptive wavelets) using more typical largest magnitude wavelet coefficient selection could yield as good as, or better classification than the spectral features discovered by GRLVQI. This has not been investigated.

We addressed wavelet feature extraction for classification in this paper. One may view feature extraction as a form of compression. We realize, however, that there may be a desire to code the remaining transform coefficients. Our wavelet feature extraction method potentially allows one to very efficiently code the coefficients selected by GRLVQI for the purposes of classification. Our model inherently saves bit resources by significantly reducing the number of coefficients that one needs to code. The problem of coding 200 potential wavelet coefficients and indices is reduced to coding a single 200-dimensional vector and 15 retained features and indices. We surmise that bit-plan coding would offer an efficient method of coding the retained coefficients with little overhead.

Our feature selection algorithm using GRLVQI in the wavelet domain is very successful. However, we feel it necessary to consider some aspects related to remotely sensed hyperspectral data if we are to leverage the full power of the wavelet transform. We are currently researching ways to address two specific issues. The first issue is how to treat the deletion of the spectral channels corresponding to data fallout (e.g., due to saturation of the water bands). The discontinuities created by gluing separate regions together manifest in the wavelet coefficients of the transformed data and can produce false features. Such false features are a concern because they can potentially cause GRLVQI to waste relevance resources learning features that are an artifact of the transform. The second issue is that the wavelet transform assumes uniformly distributed band centers. However, spacing of the band centers in remotely sensed hyperspectral images is typically non-uniform. We are currently investigating the impact of these issues on the wavelet transform and the effect on the classification accuracy and feature reduction potential of GRLVQI in the wavelet feature space.

ACKNOWLEDGEMENTS

Helpful private communications with Dr. Richard G. Baraniuk, Dr. Barbara Hammer, and Dr. Thomas Villmann are greatly appreciated.

MM is supported by the Air Force Institute of Technology, Wright-Patterson AFB, OH. Special thanks to Mr. Raymond Haren for his continued support.

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

REFERENCES

- Burrus C.S., R.A. Gopinath, H. Guo (1998). *Introduction to Wavelets and Wavelet Transforms: A Primer*, Upper Saddle River, NJ: Prentice Hall.
- DeSieno, D. (1995). Adding a conscience to competitive learning, in *Proceedings of the IEEE International Conference on Neural Networks I*, New York, NY, July, 117–124.
- Hammer, B., and T. Villmann (2002). Generalized relevance learning vector quantization, *Neural Networks*, 15:1059–1068.
- Green, R. O. (1996). Summaries of the 6th Annual JPL Airborne Geoscience Workshop, 1. *AVIRIS Workshop*, Pasadena, CA, March 4–6.
- Kohonen, T. (2001). *Self-Organizing Maps*, 3rd ed. Springer-Verlag Berlin Heidelberg.
- Mendenhall, M., and E. Merényi (2006). Relevance-based Feature Extraction for Hyperspectral Images, *IEEE Transactions on Neural Networks*, Submitted.
- Merényi, E., R. Singer, and J. Miller (1996). Mapping of spectral variations on the surface of mars from high spectral resolution telescopic images, *ICARUS*, 124:280–295.
- Merényi, E. (2000). Precision mining of high-dimensional patterns with self-organizing maps: Interpretation of hyperspectral images, in *QuoVadis Computational Intelligence: New Trends and Approaches in Computational Intelligence. Studies in Fuzziness and Soft Computing.*, P. Sincak and J. Vascak, Eds., 54. Physica-Verlag. Available: <http://www.ece.rice.edu/~erzsebet/publications.html>
- Moon, T., and E. Merényi (1995). Classification of hyperspectral images using wavelet transforms and neural networks, in *Proceedings of the SPIE: Wavelet Applications in Signal and Image Processing III*, Andrew F. Laine; Michael A. Unser; Mladen V. Wickerhauser; Eds., 2569:725–735.
- Oehler, K. L., and R. M. Gray (1995). Combining image compression and classification using vector quantization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):461–473.
- Sato, A., and K. Yamada (1996). 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, 423–429.
- Zhang, X., N. H. Younan, and C. G. O'Hara (2005). Wavelet domain statistical hyperspectral soil texture classification, *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):615–618.