

## COMPARISOM OF WAVELET-BASED AND HHT-BASED FEATURE EXTRACTION METHODS FOR HYPERSPECTRAL IMAGE CLASSIFICATION

X.-M. Huang<sup>a</sup> and P.-H. Hsu<sup>a,\*</sup>

<sup>a</sup> Dept. of Civil Engineering, National Taiwan University, No.1, Sec. 4, Roosevelt Rd., Taipei City 10617, Taiwan –  
(r98521116, hsuph)@ntu.edu.tw

Commission VII, WG VII/3

**KEY WORDS:** Hyperspectral Image, Feature Extraction, Classification, Wavelet Transform, Hilbert-Huang Transform

### ABSTRACT:

Hyperspectral images, which contain rich and fine spectral information, can be used to identify surface objects and improve land use/cover classification accuracy. Due to the property of high dimensionality of hyperspectral data, traditional statistics-based classifiers cannot be directly used on such images with limited training samples. This problem is referred as “curse of dimensionality.” The commonly used method to solve this problem is dimensionality reduction, and feature extraction is used to reduce the dimensionality of hyperspectral images more frequently. There are two types of feature extraction methods. The first type is based on statistical property of data. The other type is based on time-frequency analysis. In this study, the time-frequency analysis methods are used to extract the features for hyperspectral image classification. Firstly, it has been proven that wavelet-based feature extraction provide an effective tool for spectral feature extraction. On the other hand, Hilbert-Huang transform (HHT), a relative new time-frequency analysis tool, has been widely used in nonlinear and nonstationary data analysis. In this study, wavelet transform and HHT are implemented on the hyperspectral data for physical spectral analysis. Therefore, we can get a small number of salient features, reduce the dimensionality of hyperspectral images and keep the accuracy of classification results. An AVIRIS data set is used to test the performance of the proposed HHT-based feature extraction methods; then, the results are compared with wavelet-based feature extraction. According to the experiment results, HHT-based feature extraction methods are effective tools and the results are similar with wavelet-based feature extraction methods.

### 1. INTRODUCTION

Imaging spectrometer, a technology which was developed in 1980's, can obtain hundreds of spectral bands simultaneously (Goetz *et al.*, 1985). The images acquired with spectrometers are called as hyperspectral images. These images not only reveal two-dimensional spatial information but also contain rich and fine spectral information. With these characteristics, they can be used to identify surface objects and improve land use/cover classification accuracies. In past three decades, hyperspectral images have been widely used in different fields such as mineral identification, vegetation mapping, and disaster investigation (Goetz *et al.*, 1985).

Because hyperspectral data have the property of high dimensionality, image processing methods which have been effectively applied to multispectral data in the past are not as proper as to hyperspectral data. For instance, it is ineffective when the traditional statistical classification methods are applied to hyperspectral images with limited training samples. In other words, the dimensionality increases with the number of bands, the number of training samples for classification should be increased as well (Hsu, 2007). This has been termed the “curse of dimensionality” by Bellman (1961). The commonly used method to solve “curse of dimensionality” is dimensionality reduction, which can be divided into two types: feature selection and feature extraction. For hyperspectral images, feature extraction is used to reduce the dimensionality more frequently (Hsu, 2003).

There are two types of feature extraction methods. The first type is based on the statistical property of data. For instance, principal components transform (PCT) is the most commonly used and simple method. Although it concerns the distribution of whole data, some useful features for hyperspectral data will be neglected easily. Discriminant analysis feature extraction (DAFE) is to maximize the between-class scatter and minimize the within-class scatter. Moreover, decision boundary feature extraction (DBFE), which was proposed by Lee and Landgrebe (1993), could find useful features by decision boundaries between different classes. Although DAFE and DBFE are effective and practical algorithms, there are some disadvantages. For example, the maximum number of feature in DAFE is the number of class minus one. Besides, in order to get reliable parameters in DAFE or to compute the decision boundaries in DBFE, it still needs adequate training samples (Fukunaga, 1990; Lee and Landgrebe, 1993).

The other type of feature extraction methods is based on time-frequency analysis. For example, it has been proven that wavelet-based feature extraction provide an appropriate and effective tool for spectral feature extraction (Hsu, 2003). However, this method has some disadvantages; for instance, it has to select the wavelet basis function in advance, or it is not suitable for nonlinear data analysis. Hilbert-Huang transform (HHT) is a relatively new adaptive time-frequency analysis tool. It combines empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA), and has been used extensively in nonlinear and nonstationary data analysis. In this study, the wavelet transform and HHT are implemented on the hyperspectral data for physical spectral analysis. The spectral features are then extracted

---

\* Corresponding author.

based on the results of physically spectral analysis, so that we can get a small number of salient features, reduce the dimensionality of hyperspectral images and keep the accuracy of classification results. In our experiment, an AVIRIS data set is used to test the performance of the proposed HHT-based methods. Finally, the results are also compared with wavelet-based feature extraction methods.

## 2. HILBERT-HUANG TRANSFORM

Hilbert-Huang transform (HHT), first proposed by Huang *et al.* (1998), is a valid time-frequency analysis tool for nonlinear and nonstationary data. The HHT which consists of empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA) will be described briefly in this section.

### 2.1 Empirical mode decomposition

Empirical mode decomposition (EMD) can decompose time-series data into a series of intrinsic mode functions (IMFs) adaptively. These IMFs include different regions of frequency, and each IMF has two properties (Huang, 2005):

1. The number of extrema and the number of zero-crossing of an IMF must equal or differ at most by one.
2. All the local maxima and minima of an IMF are symmetric with respect to zero.

The EMD consists of the following steps:

1. First, identify all the local maxima and connect them by cubic spline function as the upper envelope for a signal,  $x(t)$ . Repeat the procedure for the local minima to generate the lower envelope.
2. Compute the mean  $m_1$  of the upper and lower envelopes, and let  $x(t)$  minus  $m_1$ . We will get first proto-IMF (PIMF) component,  $h_1$ :

$$x(t) - m_1 = h_1 \quad (1)$$

The procedure which obtain IMF components is called sifting process.

3. Proto-IMF,  $h_1$ , may not satisfy the definitions of IMF. Repeat the sifting process  $k$  times until the IMF meet the stoppage criteria.

$$h_{1(k-1)} - m_{1k} = h_{1k} \quad (2)$$

4. As soon as the IMF component satisfy the criteria, we will get first IMF,  $c_1$ , and separate  $c_1$  from  $x(t)$ .

$$x(t) - c_1 = r_1 \quad (3)$$

5. Since the residue,  $r_1$ , still contains information with long periods, it is treated as the data and repeat the sifting process. The result is

$$\begin{aligned} r_1 - c_2 &= r_2 \\ &\dots \\ r_{n-1} - c_n &= r_n \end{aligned} \quad (4)$$

Finally, by Summing up equation (3) and (4), we obtain

$$x(t) = \sum_{j=1}^n c_j + r_n \quad (5)$$

The EMD separates variations from the mean, and each IMF has its own physical meaning.

### 2.2 Hilbert Spectral Analysis

Having obtained the IMF components, we can apply the Hilbert transform to each IMF component and compute the instantaneous frequency. Then we can find the complex conjugate,  $\hat{c}_i(t)$ , of an IMF,  $c_i(t)$ , and have an analytic signal:

$$z_i(t) = c_i(t) + j\hat{c}_i(t) = a_i(t)e^{j\theta_i(t)} \quad (6)$$

$a_i(t)$  is the function of instantaneous amplitude, and  $\theta_i(t)$  is the function of phase angle. As a consequence, we can express the original data as the real part,  $RP$ , in the following form:

$$X(t) = RP \sum_{i=1}^n a_i(t)e^{j\theta_i(t)} = RP \sum_{i=1}^n a_i(t)e^{j\int \omega_i(t)dt} \quad (7)$$

Therefore, the Hilbert spectrum can be defined as:

$$H(\omega, t) = RP \sum_{i=1}^n a_i(t)e^{j\int \omega_i(t)dt} \quad (8)$$

We can also define the marginal Hilbert spectrum as:

$$h(\omega) = \int_0^T H(\omega, t)dt \quad (9)$$

In summary, the HHT, consisting of EMD and HSA, can decompose data adaptively and compute instantaneous frequency by differentiation rather than convolution. HHT is a superior data analysis tool for nonlinear and nonstationary data.

### 2.3 Comparisons between Wavelet Transform and Hilbert-Huang Transform

Wavelet transform and Hilbert-Huang transform are both time-frequency analysis tools, so that they can analysis the variation of data in both time and frequency domain. Table 1 shows some differences between wavelet transform and HHT. Firstly, wavelet transform have complete theoretical base and have to define a basis function before using it; whereas, HHT with empirical theoretical base has an adaptive basis, which can analysis data adaptively. Second, wavelet transform computes frequency by convolution operation; while, the frequency is derived by differentiation rather than convolution in HHT. However, wavelet transform and HHT can present the results in time-frequency-energy space. Finally, wavelet transform is suitable for nonstationary data but is unsuitable for nonlinear data. On the contrary, HHT is suitable for both nonlinear and nonstationary data. Therefore, HHT is a superior tool for time-frequency analysis of nonlinear and nonstationary data (Huang, 2005).

	Wavelet Transform	HHT
Theoretical Base	Theory complete	Empirical
Basis	A prior	Adaptive
Frequency	Convolution	Differentiation
Presentation	Energy-time-frequency	Energy-time-frequency
Nonlinear Data	Unsuitable	Suitable
Nonstationary Data	Suitable	Suitable

Table 1. Comparisons between Wavelet and Hilbert-Huang Transform (Huang, 2005)

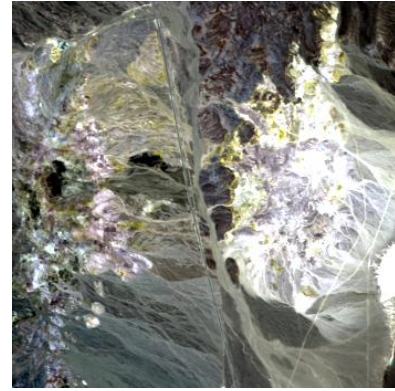
### 3. HYPERSPECTRAL IMAGE FEATURE EXTRACTION

#### 3.1 Datasets Description

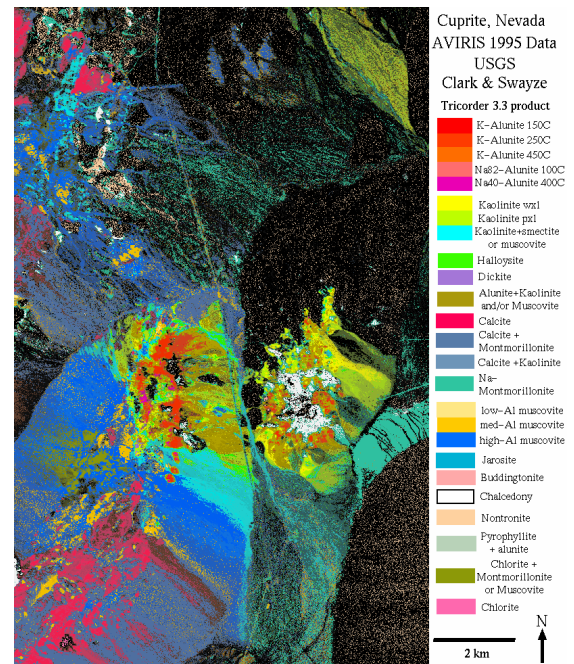
In this study, an AVIRIS data set is used to test the performance of using wavelet transform and HHT on hyperspectral image feature extraction and classification. The AVIRIS data set shown in Figure 1(a) is the well-known Cuprite data set, which is a mineral region at Nevada. The image size of the test field is 350×350. The number of bands is 224. Figure 1(b) also shows a mineral map produced in 1995 by USGS. In this study, we choose 6 classes from this map (Table 2) for feature extraction and classification. Table 2 also shows the number of training samples and check sample for image classification.

Class names	# of training samples	# of check samples
Alunite	100	50
Kaolinite	100	50
Muscovite	100	50
Calcite	100	50
Montmorillonite	100	50
Kaolinite+Semectite or Muscovite	100	50

Table 2. The 6 chosen classes



(a) False image



(b) Mineral map in Cuprite(USGS Spectroscopy Lab, 1998)

Figure 1. An AVIRIS data set of Cuprite

#### 3.2 Wavelet-Based Feature Extraction

The orthogonal wavelet transform can decompose a signal into the low-frequency components that represent the optimal approximation, and the high-frequency components that represent detailed information of the original signal (Mallat, 1989). The decomposition coefficients in a wavelet orthogonal basis can be computed with a fast algorithm that cascades discrete convolutions with conjugate mirror filters (CMF)  $h$  and  $g$ , and subsamples the outputs. The decomposition equations are described as following:

$$\begin{aligned}
 a_{j+1}[p] &= \sum_{n=-\infty}^{\infty} h[n-2p]a_j[n] \\
 d_{j+1}[p] &= \sum_{n=-\infty}^{\infty} g[n-2p]a_j[n]
 \end{aligned} \quad (10)$$

$a_j$  is the approximation coefficients at scale  $2^j$ , and  $a_{j+1}$  and  $d_{j+1}$  are respectively the approximation and detail components at

scale  $2^{j+1}$  (Mallat, 1999). Because of sub-sampling, the length of the original signal is reduced after applying conjugate mirror filter to the original signal. With these properties, wavelet decomposition is implemented on dimensionality reduction of hyperspectral images (Hsu, 2003).

Linear and nonlinear wavelet-based feature extraction (WFE) methods are applied to hyperspectral images in this study. In linear WFE, the approximation coefficients  $a_j$  are regarded as features for classification. On the other hand, nonlinear WFE consider approximation coefficient  $a_j$  as well as detail coefficient  $d_j$ , and select  $M$  largest wavelet coefficients as important features for classification (Hsu, 2003). In the experiments, the wavelet function used in linear and nonlinear WFE is Daubechies 3 wavelet (Daubechies wavelet with 3 vanishing moments). The experiment procedure of wavelet-based feature extraction is illustrated in Figure 2.

### 3.3 HHT-Based Feature Extraction

According to the characteristics of time-frequency analysis of HHT, the HHT will be applied to spectral curves of each pixel in hyperspectral image. First of all, Hilbert-Huang transform is implemented on a spectral curve. The instantaneous frequency and amplitude of each component will be calculated. Then Hilbert spectrum is formed by using instantaneous frequency and amplitude. The residual information is also considered in this spectrum. After that, the  $M$  largest values in the Hilbert spectrum are selected as the important features of the spectral curve for classification. These features are sorted by the bands where the feature is located. If more than two features have same location of bands, sort the features according to their frequency. Finally, the extracted features are used as the inputs for classification. Maximum likelihood classifier is used in this study. The procedure of feature extraction using Hilbert-Huang transform is illustrated in Figure 2 as well.

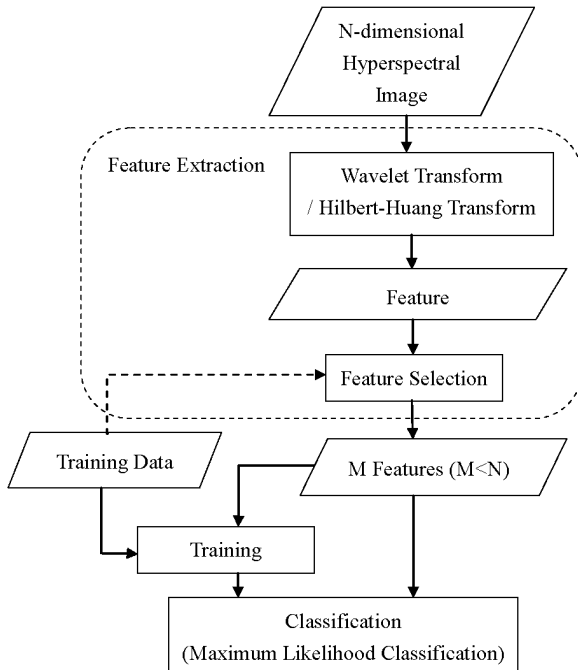


Figure 2. The flow chart of wavelet-based or HHT-based feature extraction

## 4. EXPERIMENTS

There are three experiments in this study. The first experiment is to compare the performance between unsupervised and supervised HHT-based feature extraction. The second experiment is to compare different WFE methods. Finally, the HHT-based feature extraction methods are compared with wavelet-based feature extraction methods mentioned in section 3.2 and 3.3.

### 4.1 Experiment I: Comparison of Classification Results between Unsupervised and Supervised HHT-Based Feature Extraction Methods

The purpose of the experiment is to test the performance of unsupervised HHT-based feature extraction method. In addition, the features of supervised HHT-based feature extraction are selected by computing Bhattacharyya distances from the features extracted by unsupervised HHT-based methods. The classification accuracies are calculated for various numbers of features by HHT-based methods.

Figure 3 shows the classification accuracies with different HHT-based feature extraction methods. We can find that both unsupervised and supervised HHT-based methods have good classification accuracies. The classification results both conform to Hughes phenomenon that accuracy increases at first and then accuracy decline when the number of features increases with constant number of training samples. Compared with unsupervised HHT-based method, supervised HHT-based method can improve classification accuracy apparently. Supervised HHT-based feature extraction can achieve better classification accuracy (90%) with six and seven features, whereas unsupervised HHT-based method has lower accuracy (81.33%) with four features. Therefore, supervised HHT-based method can have better results by calculating the separability of different classes with training samples than unsupervised HHT-based method.

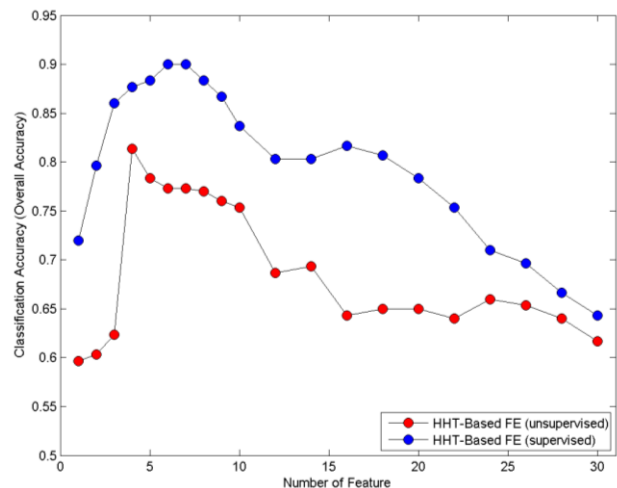


Figure 3. Comparison of classification accuracies between HHT-based methods

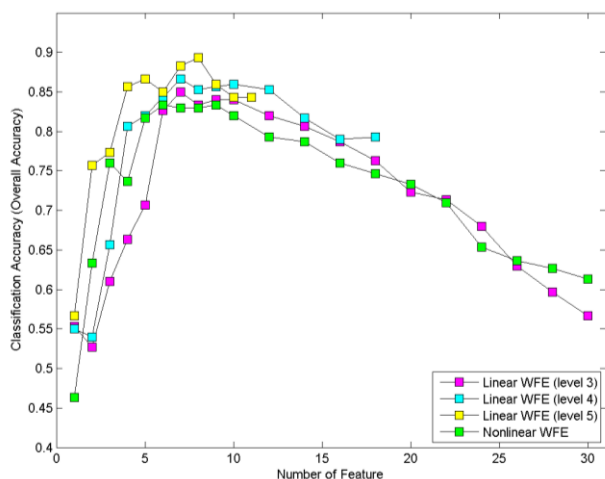


Figure 4. Comparison of classification accuracies between WFE methods

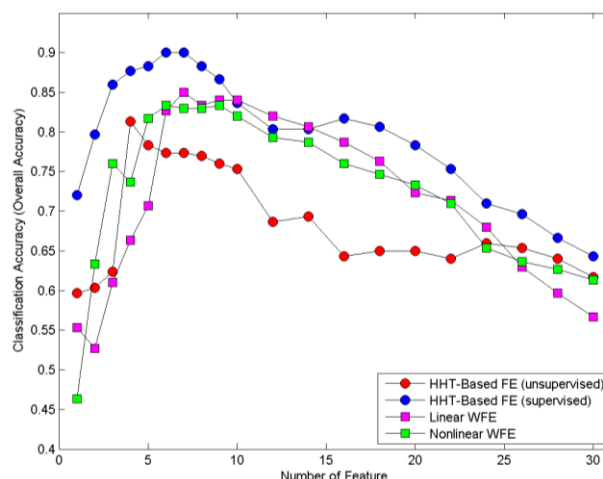


Figure 5. Comparison of classification accuracies between wavelet-based and HHT-based feature extraction methods

#### 4.2 Experiment II: Comparison of Classification Results between Wavelet-based Feature Extraction Methods

The purpose of the experiment II is to compare the performance between wavelet-based feature extraction (WFE) methods. In linear WFE, we use Daubechies 3 wavelet as basis function and decompose signal from level 3 to level 5. The numbers of features in different linear WFE methods are thirty, eighteen and eleven respectively.

Figure 4 shows the classification accuracies between WFE methods. First of all, the results of linear and nonlinear WFE methods are similar, and the best accuracies of linear WFE (level 3), linear WFE (level 4), linear WFE (level 5) and nonlinear WFE are 85%, 86.67%, 89.33 and 83.33%. In addition, the classification accuracies increase slightly in linear WFE, when the level of decomposition increases.

#### 4.3 Experiment III: Comparison of Classification Results between Wavelet-Based and HHT-Based Feature Extraction Methods

In experiment III, the purpose is to compare the performance between wavelet-based and HHT-based methods. The classification accuracies with different methods are showed in Figure 5. First of all, the classification accuracies of WFE methods and HHT-based methods are all conformed to Hughes phenomenon that classification accuracy increases at first and then declines as the number of feature grows.

Compared with the results of different methods, linear and nonlinear WFE have similar classification results which have been mentioned in section 4.2. Also, the results of unsupervised HHT-based method are similar to WFE methods but the accuracies decrease obviously when the number of feature is more than ten. Finally, supervised HHT-based feature extraction can achieve better classification results than any other methods.

#### 5. CONCLUSION

In this study, two feature extraction methods using Hilbert-Huang transform were proposed to extract useful features for hyperspectral image classification. The results of HHT-based methods are compared with wavelet-based feature extraction methods.

According the experiments, the results of unsupervised HHT-based methods are similar to the result of WFE which is implemented in this study, but the accuracies of unsupervised HHT-based method are unstable when the feature increases. Subsequently, when computing the separability of different classes with training samples, supervised HHT-based method can have better result than unsupervised HHT-based method and can reach 90% classification accuracy with six or seven features. Furthermore, it also has superior classification accuracies than linear and nonlinear WFE. By extracting features from Hilbert spectrum, we can not only reduce the dimensionality of hyperspectral image but also get a small number of salient features for classification. Therefore, Hilbert-Huang is an appropriate and effective tool for hyperspectral image analysis.

In the future, the effectiveness of HHT-based methods still could be improved. In addition, the objects in the experiments are mainly the minerals. It is another object to investigate that HHT-based feature extraction methods proposed in this study are suitable and have similar/better results than WFE methods for other kind of material objects such as metropolitan area of vegetation area.

#### REFERENCES

- AVIRIS Airborne Visible/Infrared Imaging Spectrometer, 2011. AVIRIS Data - Ordering Free AVIRIS Standard Data Products. [http://aviris.jpl.nasa.gov/data/free\\_data.html](http://aviris.jpl.nasa.gov/data/free_data.html) (14 Jul. 2011).
- Bellman, R., 1961. *Adaptive Control Processes: A Guided Tour*, Princeton University Press.
- Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*, Second edition, San Diego: Academic Press, Inc.

- Goetz, A. F. H., G. Vane, J. E. Solomon, and B. N. Rock, 1985. Imaging Spectrometry for Earth Remote Sensing, *Science*, 228:1147-1153.
- Hsu, P.-H., 2003. *Spectral Feature Extraction of Hyperspectral Images using Wavelet Transform*, Ph.D. Thesis, National Cheng Kung University, Tainan, Taiwan, R.O.C..
- Hsu, P.-H., 2007. Feature extraction of hyperspectral images using wavelet and matching pursuit, *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(2):78-92.
- Huang, N. E., Z. Shen, S.R. Long, M. C. Wu, H. H. Shin, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998. The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis, *Proc R. Soc. London*, A454: 903-995.
- Huang, N. E., 2005. Introduction to Hilbert-Huang Transform and Some Recent Developments. *The Hilbert-Huang Transform in Engineering*, Taylor & Francis Group, LLC, Boca Raton, pp.1-23.
- Lee, C. and D. A. Landgrebe, 1993. Analyzing High-Dimensional Multispectral Data, *IEEE Transactions on Geoscience and Remote Sensing*, 31: 792-800.
- Mallat, S., 1989. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7): 674-693.
- Mallat, S. 1999. *A Wavelet Tour on Signal Processing*, Academic Press, New York.
- USGS Spectroscopy Lab, 1998. Imaging Spectroscopy Material Maps: Cuprite Introduction.  
<http://speclab.cr.usgs.gov/map.intro.html> (14 Jul. 2011).