Effects of Hyperspectral Data Transformations on Urban Inter-class Separations using a Support Vector Machine

1M.A. Misman, 2H.Z.M. Shafri and 3Raja M. Kamil Raja Ahmad
1Institute of Advanced Technology,
2Department of Civil Engineering, Faculty of Engineering,
3Department of Electrical and Electronic Engineering, Faculty of Engineering,
Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Abstract: This study investigated the performance of different data types used in a hyperspectral data classification process. Data in the form of spectral reflectance, first derivative spectra and wavelet coefficients were used as inputs for the Support Vector Machine (SVM) algorithm used to classify five different classes. The first derivative spectra gave a lower classification accuracy (35.6%) than the spectral reflectance (82%) and the use of wavelet coefficients further improved the classification accuracy to 100%. Proper selection of the wavelet transformation method, the mother wavelet, the number of vanishing moments and the decomposition level could improve classification accuracy. In summary, wavelet coefficients could maximise discrimination capability as compared to the spectral reflectance and first derivative spectra.

Key words: Hyperspectral, first derivative, wavelet coefficients, support vector machine

INTRODUCTION

The availability of various remote sensing data provides an opportunity for users to fully utilise the data to achieve their goal with a maximum success rate. These opportunities arise because remote sensing technology has been used in many applications including mapping, the military, meteorology, agriculture and others. Different types of data have their own usefulness. Multispectral data with high spatial resolution is suitable for mapping and hyperspectral data is more suitable for subtle discrimination.

Hyperspectral data have been widely used because of their capability in discriminating between subtle variations among similar features and they could improve the user’s capability for gaining a greater understanding of various features. Although, it has been extensively used in agricultural applications (Rao et al., 2007; Gong et al., 2002; Mutanga et al., 2004; Chappelle et al., 1992; Hansen and Schjoerring, 2003), hyperspectral data have also been extensively used for urban and sub-urban applications (Heiden et al., 2007; Bassani et al., 2007).

Several methods such as clustering (Gomez-Chova et al., 2009; Oldeland et al., 2010) or indices (Oldeland et al., 2010; Huang et al., 2009; Kuckenberg et al., 2008; Devadas et al., 2008) are possible for discriminating between one feature and others. Classification (Laverington, 2010; Wang and Sousa, 2009; Lucas et al., 2008; Wilson et al., 2004; Castro-Esau et al., 2004) is also one of the methods used to discriminate between features. There are several factors that can contribute to the success of a classification process, including the selection of data type for use as input for the classification (Koger et al., 2003; Yang et al., 2009; Ouma et al., 2008), the selection of the optimal band (Phillips et al., 2009; Keshava, 2004; Murakami, 2004; Serpico and Moser, 2007) and the use of classifiers (Shafri and Ramle, 2009; Clark et al., 2005; Yang et al., 2009; Erbek et al., 2004; Ali et al., 2009; Nelson, 1981). These are the major factors that will impact the classification accuracy.

For hyperspectral data, the original data are normally in the form of the spectral reflectance. Exploitations or transformations of the original data could be performed to improve the classification results. Calculating the spectral derivative is one of the transformations that can be applied to the reflectance data and several studies have shown that derivative data could achieve better results than the reflectance data (Tsai and Philpot, 1998; Han, 2002). Wavelet transformations can also be used as a data transformation method and have been used for applications like feature extraction, image compression, feature detection and others (Hsu, 2007; Koger et al., 2003; Galford et al., 2008; Bruce et al., 2002; Loun et al.,

Corresponding Author: Helmi Zulhaidi Mohd Shafri, Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
Fig. 1: The study area and the AISA image

2007; Ping et al., 2009; Raju et al., 2008). However, the best data type has yet to be determined for use in classifying different urban-area features while using data from an Airborne Imaging Spectrometer for Application (AISA). This is particularly important in the Malaysian context because AISA is currently the only airborne hyperspectral sensor available through a commercial data provider. Thus, this study focused on determining the effects of data input selection on the classification of inter-class features by using AISA data.

MATERIALS AND METHODS

The data used in this study were acquired in November 2009 over an area in Kuala Lumpur, Malaysia (Fig. 1a). With a pixel size of approximately 1×1 m, the AISA image covers a spectral region of 400-1000 nm with a spectral resolution of approximately 5 nm for 128 bands. The AISA image used in this study is shown in Fig. 1b.

The three major steps involved in this study were image processing, data transformation and classification. The general methodology of this study is shown in Fig. 2.

Image processing: The AISA image required processing before feature extraction could be made. The image was converted to reflectance because it was collected in radiance mode to minimise the effects of the atmosphere. The conversion method used was a log-residual method in ENVI software. After the conversion to reflectance was made, a Minimum Noise Fraction (MNF) transformation was performed. This step minimised the noise in the image. Twenty seven first-MNF bands were used to obtain a de-noised image that had less noise than the original image obtained using the Inverse MNF method. Next, the Pixel Purity Index (PPI) was performed to find the pixels of highest purity in the image that represent certain features.

Feature extraction: The feature extraction was performed based on the PPI result. Five features were selected for further processing: vegetation, water, road, concrete and rooftop. One hundred pixels were extracted from the de-noised image for each feature. Each pixel represented the signal for that particular feature. Those signals were used for further processing. Figure 3 shows the average spectral reflectance for the selected features.

Data transformation: The methodology for data transformation is shown in Fig. 4. Three major steps were involved in this process and included spectral derivative conversion followed by wavelet conversion of the spectral reflectance and derivative data using a Continuous Wavelet Transform (CWT) and a Discrete
Wavelet Transform (DWT). Selection of the mother wavelet was also performed along with selection of the number of vanishing moments and the level of decompositions.

The original spectral reflectance data can be transformed into other types of dimensional data by applying mathematical operations. Using transformed data may provide better information and understanding than using the original data. For instance, the spectral derivative enhances the spectral differences in certain parts of the spectrum, removes multiplicative factors and reduces the effect of the soil background (Tsai and Philpot, 1998; Gong et al., 2001). All of the spectral reflectance samples used in this study were transformed into first derivative by using Eq. 1 (Dawson and Curran, 1998).

\[
DR_{\lambda i}(t) = \frac{(R_{\lambda i+1}) - R_{\lambda i})}{\Delta \lambda}
\]

where, FDR is the first-derivative reflectance at wavelength \( \lambda \), \( R_{\lambda i} \) is the reflectance at wavelength \( j \), \( R_{\lambda i+1} \) is the reflectance at wavelength \( j+1 \) and \( \Delta \lambda \) is the difference in wavelength between \( j \) and \( j+1 \).

Wavelet transformations were then applied to the spectral reflectance and the first-derivative data. A wavelet is a mathematical function used to divide a given function (or signal) into different scale components and a wavelet transform is the representation of a function by wavelets. Wavelets have advantages over traditional Fourier transform for representing functions that have discontinuities and sharp peaks. Wavelets also have advantages for deconstructing and reconstructing a signal. Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT) are two types of wavelet transformations.

There are several types of mother wavelets and each mother wavelet has its own characteristics. The mother wavelets are different in terms of their orthogonality,

Fig. 3: Average spectral reflectance for vegetation, water, road, concrete and rooftop features

Fig. 4: Flowchart for the methodology of data transformation
support, regularity and symmetry. As a result, wavelet coefficients vary according to the selection of mother wavelet. This study uses only seven mother wavelets (Fig. 5a-g): Haar, Daubechies, Symlet, Coiflet, Biorthogonal, Reverse Biorthogonal and Discrete Meyer.

Both DWT and CWT were used in this study. Each spectral reflectance and first-derivative spectrum was transformed or decomposed into wavelet coefficients with nine levels of decomposition. For DWT, each decomposition process produced two types of coefficients, the approximation coefficient (cA) and the detail coefficient (cD). Their values are illustrated in Fig. 6. Only the detail coefficient from each level was used as an input for classification. DWT and CWT were tested to determine the best method of wavelet transformation for use in distinguishing features.

Both the spectral reflectance and the first derivative data were decomposed using the selected wavelet families to study the effects of mother wavelet selection on discrimination capability and to investigate which of the mother wavelets performed best. Haar, Daubechies (db1-db20), Symlet (sym1-sym20), Coiflet (coif1-coif5), Biorthogonal (bior1.1-bior6.8), Reverse Biorthogonal (rbio1.1-rbio6.8) and Discrete Meyer (dmey) are the seven mother wavelets that were tested in this study. They were decomposed, level-by-level, up to level nine. Since the transformation was applied to both the spectral reflectance and the first-derivative spectra, the analysis generally produced two groups of wavelet coefficient data. One group contained the wavelet coefficients derived from the spectral reflectance dataset and the other group was derived from the first-derivative dataset. The classification of datasets containing the spectral reflectance, the first-derivative and the wavelet coefficients were performed after the wavelet transformation process was completed.

**Support vector machine classification:** Classification was one of the methods used for information extraction. Various supervised and unsupervised classification algorithms may be used to assign data to one possible class. The choice of classifier (i.e., decision rule) depends
Fig. 7: Illustration of the SVM process in two-dimensional space. Blue dots represent data from group 1 while red dots represent data from group 2.

Fig. 8: First-derivative spectra of the spectral reflectance of each class.

on the nature of input data and the desired output (Jensen, 2005). The Support Vector Machine (SVM) method was selected because it is considered the most suitable classifier for handling limited samples (Chi et al., 2008). SVM is a supervised learning method used for classification and regression. It constructs a separating hyperplane between two sets of data in n-dimensional space. The hyperplane will maximise the margin between the two data sets. A good separation is achieved by the hyperplane that has the largest distance separating the neighbouring data points of both classes. Larger margins will lower the generalisation error of the classifiers.

The spectral reflectance, the first derivative of spectral reflectance, the wavelet coefficients derived from the spectral reflectance and the wavelet coefficients derived from the first-derivative spectra were used as inputs for SVM classification. Fifty samples from each group were used as a training dataset. To eliminate any potential bias, only the remaining separate samples were used as testing dataset. The performance of the spectral reflectance, first-derivative and all-wavelet coefficients data were evaluated based on the classification accuracy (Fig. 7).

RESULTS AND DISCUSSION

First derivative transformation: The values of the spectral reflectance data depend on many factors including sun illumination. Use of the first-derivative data has an advantage over use of the reflectance data because it minimises the effects of sun illumination. Several studies have established the usefulness of the first-derivative data for separating classes more effectively than the reflectance data (Tsai and Philpot, 1998; Han, 2002). Figure 8 shows the first-derivative spectra of the average spectral reflectance for each class.

Wavelet coefficients: Different wavelet coefficient values were obtained from the decomposition of different mother wavelets. There were also differences in wavelet coefficient values obtained from the spectral reflectance and first-derivative spectra. Although, the spectral reflectance and the first derivative spectra were decomposed using the same mother wavelet, their wavelet coefficient values resulting from decomposition were different. It has been suggested for some time that the use of spectral derivatives can reduce the illumination and other effects, thus, the use of spectral derivatives as input data may offer further benefits when applying wavelet analysis (Blackburn and Ferwerda, 2008). In addition, the number of vanishing moments used for the decomposition also affected the resulting wavelet coefficient values. Figure 9 shows an example of the wavelet coefficient values obtained from spectral reflectance data and first-derivative spectra obtained from a vegetation sample. Generally, the coefficient values of each mother wavelet varied because of different characteristics the mother wavelets. The coefficient values derived from first-derivative spectra were also smaller than the coefficient values derived from the spectral reflectance. More visualization of wavelet transformation of spectral reflectance and first-derivative using DWT and CWT methods by Blackburn and Ferwada (2008).

Classification results: The performance of all data types that were tested in this study was assessed by their classification accuracies. Generally, higher classification accuracy indicates better feature discrimination.

The classification results of the spectral reflectance and the first-derivative data are shown in Table 1. The classification accuracy for the spectral reflectance was...
Table 1: Classification accuracy of reflectance and first-derivative spectra

<table>
<thead>
<tr>
<th>Data</th>
<th>Classification accuracy (%)</th>
<th>Kappa coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflectance</td>
<td>82.0</td>
<td>0.78469</td>
</tr>
<tr>
<td>First derivative</td>
<td>35.6</td>
<td>0.35663</td>
</tr>
</tbody>
</table>

higher than the classification accuracy of the first derivative. The classification accuracy achieved by using spectral reflectance was 82.0% and the classification accuracy of first derivative was only 35.6%. The classification accuracy drops significantly when using derivative-transformed data. Several other studies also obtained similar results in which the classification accuracy of derivative is lower than the classification accuracy of spectral reflectance (Jones et al., 2010; Li and He, 2008; Zhang et al., 2006). For this study, the derivative spectra could not improve the discrimination capability and had worse discrimination capability than the spectral reflectance. The low classification accuracy achieved by using the derivative data may be caused by noise in the spectra because the spectral derivative is very sensitive even to small variations in the data. This factor was believed to have contributed to the low overall classification accuracy achieved when using derivative spectra.

The classification accuracy for wavelet coefficients can vary with the selection of transformation method, mother wavelet, number of vanishing moments and decomposition level. Some of the mother wavelets gave higher classification accuracies than the original dataset, while some gave lower classification accuracies. This difference shows that not all mother wavelets were suitable for feature discrimination. Koger et al. (2003), Zhang et al. (2005) and Bruce et al. (2002) got similar result pattern when classifying using different mother wavelet coefficients data sets. Koger et al. (2003) studied on detecting pitted morning glory in soybean using wavelet analysis. The classification accuracies were varied for most of the mother wavelets. Only several mother wavelets gave same accuracy. This indicated that classification accuracy is depending on the selection of mother wavelet. Selecting the number of vanishing moments was also as critical as the selection of the mother wavelet. This result is shown in Fig. 10.

Figure 11 shows the maximum accuracy of every level while using wavelet coefficients derived from reflectance and spectral-derivative data with DWT transformation. In general, the maximum accuracy increased with the number of decompositions used. The accuracy for one-level decomposition was lower than the accuracy at higher-level decompositions. This result may be due to the noise occurring at every level. Lower decomposition
Fig. 10: Classification accuracies achieved with the Daubechies wavelet while using different numbers of vanishing moments

Fig. 11: Maximum classification accuracies achieved from each level of decomposition while using the DWT wavelet coefficient. This data was based on the maximum accuracy for each level, independent of the mother wavelet, the number of vanishing moments and the level of decomposition.

Levels may contain more noise than the higher decomposition levels. The coefficient values at lower decomposition levels were small and smaller coefficient values were often assumed to be noise in the data. Besides that, the number of coefficients at a lower decomposition level was greater than the number of coefficients at higher decomposition levels. Wavelet coefficients at lower scale are sensitive to narrow or local spectral features because they are derived from high-pass filters, which is similar to the derivative spectra (Zhang et al., 2006). These factors could affect classification accuracy with regard to the selection of decomposition level.

Figure 11 also suggests that the classification accuracy while using wavelet coefficients derived from reflectance was higher than the accuracy of wavelet coefficients derived from derivative spectra. An exception was the three-level decomposition. This result shows that wavelet coefficients derived from spectral reflectance are more reliable for discriminating between features than wavelet coefficients derived from derivative spectra. Maximum classification accuracy could be achieved by using wavelet coefficients derived from the reflectance when using a five-level decomposition. Wavelet coefficients derived from derivative spectra only gives maximum accuracy when using wavelet coefficients with a seven-level decomposition.

The same pattern of results was also achieved when using wavelet coefficients that had been transformed using CWT transformation (Fig. 12). Lower classification accuracy was achieved when using lower levels of decomposition and wavelet coefficients derived from reflectance gave better classification accuracies than wavelet coefficients derived from derivative spectra. The
Table 2: Classification accuracies of the reflectance, the first-derivative spectra, the best wavelet coefficients of reflectance and the derivative

<table>
<thead>
<tr>
<th>Data</th>
<th>Classification accuracy (%)</th>
<th>Kappa coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflectance</td>
<td>82</td>
<td>0.78469</td>
</tr>
<tr>
<td>First derivative</td>
<td>35.6</td>
<td>0.30665</td>
</tr>
<tr>
<td>Symlet12 Level 5 DWT (Reflectance)</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Haar Level 8 DWT (Derivative)</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Rbio3.7 Level 8 CWT (Reflectance)</td>
<td>92</td>
<td>0.60106</td>
</tr>
<tr>
<td>Rbio3.1 Level 9 CWT(Derivative)</td>
<td>67.6</td>
<td>0.62535</td>
</tr>
</tbody>
</table>

only difference between DWT and CWT was that the CWT transformation could not produce wavelet coefficients that provided a maximum classification accuracy of 100% (as with DWT). This result showed that the DWT transformation was much better than the CWT transformation for discriminating features.

The classification accuracy for wavelet coefficients derived from first-derivative spectra did not produce better accuracies than wavelet coefficients derived from the spectral reflectance. This may have been due to noise contamination in the original data, thus resulting in poor classification results. However, there was an improvement in classification accuracy for wavelet coefficients of derivative spectra compared to the classification accuracy of the derivative spectra. This result proved that wavelet transformation could be used as one of the transformation methods for improving the output result and achieving a maximum success rate.

The overall accuracy of the data types analysed in this study is summarised in Table 2. The spectral reflectance gave better classification results than the derivative spectra. Furthermore, the wavelet coefficients of the spectral reflectance and the derivative spectra may maximise the classification accuracy. The results using Symlet 12 and reflectance data can be considered better as it requires only five-level decomposition compared to the use of Haar wavelet with derivative spectra that requires eight-level decomposition. Decomposition to a higher level would demand more processing time. Thus, Symlet 12 with a five-level decomposition transformed from reflectance data by using the DWT transformation was the best wavelet coefficient for discriminating between features.

Although this study has proven that Symlet 12 is the best wavelet coefficients in discriminating between features, this wavelet will not guarantee to give the best results when applied to other data sets. Previous studies by Bruce et al. (2002), Koger et al. (2003) and Zhang et al. (2005) and have shown differences in terms of the performance of the best mother wavelet achieved from their studies.

CONCLUSIONS

In summary, the selection of data types was essential to achieve the desired output with minimum error. The selection of wavelet transformation method, mother wavelet, number of vanishing moments and level of decomposition also played an important role in achieving better results. This study showed that the first-derivative spectra do not necessarily provide better classification accuracies than the spectral reflectance data. Some of the mother wavelets derived from spectral reflectance or from derivative spectra resulted in lower classification accuracies than the original dataset and some of the mother wavelets produced improvements in the classification results. This study showed that the potential of using wavelet-transformed data for discriminating features with a maximum success rate. Better ways of dealing with noise could be investigated to utilise the spectral derivative fully in future studies using this data.

ACKNOWLEDGMENTS

The authors would like to thank the Ministry of Higher Education Malaysia for the research grant and UPM for providing a graduate scholarship to aid in the completion of this research. Aeroscan Precision (M) Sdn. Bhd. is also acknowledged for providing the test data.

REFERENCES


