Nonlinear estimation of subpixel proportion via kernel least square regression

L. ZHANG*,†, B. WU†,‡, B. HUANG§ and P. LI†
†The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Peoples Republic of China
‡Spatial Information Research Center, Fuzhou University, Peoples Republic of China
§Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

(Received 16 April 2006; in final form 1 September 2006)

Spectral mixture analysis is an efficient approach to spectral decomposition of hyperspectral remotely sensed imagery, using land cover proportions which can be estimated from pixel values through model inversion. In this paper, a kernel least square regression algorithm has been developed for nonlinear approximation of subpixel proportions. This procedure includes two steps. The first step is to select the feature vectors by defining a global criterion to characterize the image data structure in the feature space and the second step is the projection of pixels onto the feature vectors and the application of classical linear regressive algorithm. Experiments using simulated data, synthetic data and Enhanced Thematic Mapper (ETM)+ data have been carried out, and the results demonstrate that the proposed method can improve proportion estimation. By using the simulated and synthetic data, over 85% of the total pixels in the image are found to lie between the 10% difference lines, and the root mean square error (RMSE) is less than 0.09. Using the real data, the proposed method can also perform satisfactorily with an average RMSE of about 0.12. This algorithm was also compared with other widely used kernel based algorithms, i.e. support vector regression and radial basis function neutral network and the results show that the proposed algorithm outperforms other algorithms about 5% in subpixel proportion estimation.

1. Introduction

Hyperspectral imaging spectrometers collect image data reflected from surface materials in many contiguous spectral bands. Owing to the complexity of land surface and the limited spatial resolution, a pixel in the remotely sensed hyperspectral imagery is typically a mixture of multiple electromagnetic radiances from assorted ground cover materials. Spectral unmixing is a quantitative analysis procedure used to recognize constituent ground cover materials (or endmembers) and obtain their mixing proportions (or abundances) from a mixed pixel. A general approach for spectral unmixing is to first build a mathematical model of the spectral mixture and then apply this model to implement spectral unmixing.

By modelling pixel signature in different ways, the unmixing methods can be generally grouped into two categories: linear mixture models (LMM); and non-linear
mixture models (NLMM). LMM are based on the assumption that each ground cover material only produces a single radiance and the mixed spectrum is a linear combination of ground cover radiance spectra. Hitherto, LMM have remained the dominant method for spectral unmixing analysis owing to their simplicity and generality (Settle and Campbell 1998, Hu et al. 1999, Settle 2002). LMM have been successfully applied to the abundances estimation problems in many areas, such as geology, forest and vegetation (Cross et al. 1991, Quarmby et al. 1992, Gong et al. 1994, Adams et al. 1996). However, LMM may not be appropriate for the cases where multiple scattering results in nonlinear mixture. Areas where multiple rock types are all visible on the region’s surface when viewing striated soils are typical examples of such cases. In these cases, the resultant mixture reflectance spectrum may best be described by assuming that the source radiation is multiply scattered by the randomly distributed endmembers before being collected by the imaging spectrometer. Typically, it has been found that NLMM simulates the physical phenomena more accurately (Mustard et al. 1998, Zhang and Li 1998); however, the models are usually complicated and application dependent (Borel and Gerstl 1986, Mustard and Pieters 1998). For instance, Mustard et al. (1998) showed that the NLMM produced a more accurate result after analysing the mixture of materials on the lunar surface using both the LMM and the NLMM. Mustard and Pieters (1998) also performed a quantitative analysis of mineral mixture spectra using a NLMM based on the bidirectional reflectance spectroscopy theory (Hapke 1981). In a nutshell, it could be stated that LMM has been widely employed for spectral unmixing analysis as it allows the application of mature mathematical methods, such as least squares estimation (LSE). NLMM is popular for its higher accuracy although there does not exist a simple and generic NLMM that can be utilized in various spectral unmixing applications. A natural problem then turns out to be whether we can utilize the nonlinear characteristics of spectral mixture to obtain the higher unmixing accuracy whilst simultaneously maintaining the simplicity as that of LMM.

In this paper, we present a nonlinear kernel-based least square regressive algorithm to unmix remote sensing images. This algorithm performs spectral linear mixing model regression in the feature space induced by a Mercer kernel (Vapnik 1998) and can be used to recursively construct the minimum mean squared-error regressor. The algorithm is formulated in terms of dot product in order to reduce computation complexity. It is implemented in two steps. The first step is to select the feature vectors by defining a global criterion to characterize the image data structure in the feature space and the second step is the projection of pixels on to the feature vectors and the application of the classical linear regressive algorithm.

To test the performance of the proposed algorithm, LSE, support vector regression (SVR) and radial basis function (RBF) neutral network were also implemented for comparison purposes. It has been demonstrated above that LSE is the widely used algorithm for subpixel proportion estimation, so it is an important comparison. SVR has been popular for regression problems in the past decade (Smola and Schölkopf 2004). It differs from conventional regressions in that it maps input data into a high dimensional reproducing kernel Hilbert space and uses an $\varepsilon$-insensitive loss function. As a result, SVR also has a sparse representation of solutions, which lends itself for comparisons. RBF was introduced into the neural network literature in the late 1980s. The RBF has been studied in multivariate approximation theory, particularly in the field of function interpolation, and recently received considerable interests in hyperspectral classification (Du and
Chang 1999, Kerri et al. 2001) and RBF has been reported to be a useful tool to perform proportion estimation of hyperspectral imagery. Therefore, it was also selected as comparison algorithm.

2. Kernel least square regression

Let \( r \) be an \( L \times 1 \) column image pixel vector in an multispectral or hyperspectral image where \( L \) is the number of spectral bands. Assume that \( M \) is an \( L \times p \) signature matrix, denoted by \( M = [m_1, m_2, ..., m_p] \), where \( m_i \) is a column vector represented by the \( i \)-th image endmember signature resident in the pixel vector \( r \), and \( p \) is the number of signatures of interest. Let \( a = (a_1, a_2, ..., a_p)^T \) be a \( p \times 1 \) abundance column vector associated with the fraction of the \( i \)-th signature in the pixel vector \( r \). A LMM assumes that the spectral signature of a pixel vector is linearly superimposed by spectral signatures of image endmembers \( m_1, m_2, ..., m_p \) present in the pixel vector \( r \) and can be described by:

\[
r = \sum_{i=1}^{p} a_i m_i + n \tag{1}
\]

where \( n \) is an \( L \times 1 \) column additive noise vector representing a measurement or model error. Equation (1) is a general LMM and can be solved by interpreting the noise as the error resulting from the goodness of fit in the least squares sense and then minimizing its least squares error. The model’s goodness-of-fit is assessed by the length of \( e \), using the sum-of-squared errors (SSE):

\[
\text{SSE}(\hat{\alpha}) = \|e\|^2 = r^T r - \hat{r}^T \hat{r} = r^T P_{\text{S}} r \tag{2}
\]

where \( P_{\text{S}} = I - M (M^T M)^{-1} M^T \).

LMM has been widely used for spectral unmixing analysis and can generate satisfactory results, i.e. meets the requirements in most applications (Gong et al. 1994, Adams et al. 1996); however, it is difficult for the method to capture a nonlinear relationship with a linear procedure and hence to reduce the estimated proportional accuracy in most cases. In order to overcome such a limitation, we extended the LMM and LSE to a nonlinear version to keep the simplicity of LMM and also support LSE algorithms. One way for a nonlinear extension is to lift the input space to a higher dimensional feature space by a nonlinear feature mapping and then to find a linear dimension reduction in the feature space. The basic principle behind kernel machines is that a Mercer kernel function, applied to pairs of input vectors, can be interpreted as an inner product in a high dimensional Hilbert space (the feature space) rather than the data points themselves, thus allowing inner products in feature space to be computed without making direct reference to feature vectors. The great success of the support vector machine is an example of the effective use of the kernel functions to capture nonlinear data structure (Schölkopf et al. 1999). The kernel principal component analysis (PCA) (Smola and Schölkopf 2000) and the kernel linear discriminate analysis (LDA) (Schölkopf et al. 1999, Smola and Schölkopf 2000) have also been introduced as nonlinear generalisations of the PCA and the LDA by kernel functions, respectively, and some of their interesting experimental results have been presented.

In order to establish a linear problem in the feature space that corresponds to the nonlinear problem in the input space, how to map input vectors \( r \in \mathbb{R}^L \) into the feature space and how to handle its possible high dimensionality need to be
specified. If an L-dimensional input pixel vector \( r \in R^L \) is mapped into a high dimensional space \( \mathcal{R} \) by a nonlinear mapping function \( \phi \).

\[
\phi : R^L \to \mathcal{R}, r \to \phi(r)
\]  

This high dimensional space \( \mathcal{R} \) is often called feature space. Thus equation (1) is typically written in the following form in the feature space:

\[
\phi(r_i) = \sum_{i=1}^{p} \alpha_{\phi(i)} \phi(m_i) + \phi(n_i)
\]  

where \( \phi(m_1), \phi(m_2), ..., \phi(m_p) \) vectors are endmember spectra in the features, \( \alpha_{\phi(i)} \) is a coefficient vector that accounts for the abundances of each endmember spectrum in the feature space. Since the feature space \( \mathcal{R} \) is usually very high, to employ the high dimension data, an inner product in a feature space is calculated by a positive definite kernel function \( k \), which satisfies:

\[
\langle \phi(x), \phi(y) \rangle = k(x, y)
\]  

where \( \langle ..., \rangle \) denotes the inner product.

Consider \( M_p = [m_1, m_2, ..., m_p] \) and define the kernel matrix \( K \) of dot products as:

\[
K = (k_{ij}), 1 \leq i \leq p, 1 \leq j \leq p
\]  

where \( k_{ij} = \phi^T(m_i)\phi(m_j) \). If the inner product \( \langle ..., \rangle \) form of equation (2) was replaced by \( k(r, y) \), it becomes:

\[
\text{SSE} \hat{a} = ||e||^2 = r^T r - \hat{r}^T \hat{r} = k(r, r) - k^T(r, M_p) K^{-1} k(r, M_p)
\]  

where \( k(r, M_p) = [k(r, m_1), k(r, m_2), ..., k(r, m_p)] \).

As the dimension of \( \mathcal{R} \) is higher than that of the original space, the selective \( \phi(m_1), \phi(m_2), ..., \phi(m_p) \) vectors cannot be used to form the basis for characterizing the image data structure in the feature space. Hence, using the kernel trick to obtain the value of \( \alpha_{\phi(i)} \) in equation (4) does not mean ‘the achievement of’ abundances. In order to obtain the meaningful proportions from each pixel by equation (4), we proposed a feature vector selection and linear regression algorithm. This algorithm includes two steps. The first step is to select the feature vectors by defining a global criterion to characterize the image data structure in the feature space and the second step is the projection on to the feature vectors and then apply the classical linear regressive algorithm. Figure 1 shows the architecture of the proposed algorithm using the feature vectors selection and linear regression.

![Figure 1. Architecture of the proposed algorithm.](image-url)
An unsupervised method has been developed for iterative selecting feature vectors in the \( \mathbb{R} \) space so as to capture the image data structure. Our method is as follows. Initially, we select a feature vector denoted by \( m_1 \). If there is knowledge about a material signature available as feature vector \textit{a priori}, then it can become the first feature vector; otherwise we choose the pixel vector as the first feature vector with maximum length, i.e. the brightest pixel in the scene. One reason for this selection is based on the fact that the brightest pixel may correspond to a pixel containing a material with the largest radiance spectrum in the image scene on the condition of LSE measuring the Euclidean distance. A feature spectrum in pixel space may also be a feature spectrum in kernel space in that the mapping function is monotone. In this case, the initial pixel that has the maximal distance from it tends to be a pair of the brightest and darkest pixels. Then the initial target signature is applied with an orthogonal subspace projector specified by equation (7) with \( M_S=m_1 \) to all image pixel vectors. A spectral feature signature is then found, which is denoted by \( m_2 \) with the maximum projection in the orthogonal complement space, denoted by \( \langle m_1 \rangle^\perp \) that is orthogonal to the space, linearly spanned by \( \langle m_1 \rangle \). A second spectral feature signature \( m_3 \) can be found by applying an orthogonal subspace projector \( P_{[m_1, m_2]}^\perp \) with \( M_S=[m_1, m_2] \) to the original image and a spectral signature that has the maximum projection is selected as \( m_3 \). The above procedure is repeated until all the spectral feature signatures are found or a stopping rule is met. A similar algorithm proposed by Ren and Chang (2000) has been developed, but their algorithm performs on the pixel vector; thus, their algorithm is only suitable for linear cases. Our algorithm works on kernel feature vector; thus, it is appropriate to nonlinear case.

Let \( \varepsilon \) be the prescribed error threshold, the corresponding algorithm is described as follows:

1. Initial condition: select the pixel vector with maximum length denoted by \( m_1 \). Set \( i=1 \) and \( M_1=\{m_1\} \).
2. Apply \( P_{m_1}^\perp \) via (7) to all image pixel vectors in the image.
3. Find another signature, denoted by \( m_2 \), which has the maximum orthogonal projection \( m_2=\arg \max \{||e||^2\} \), and set \( i=2 \) and \( M_2=\{m_1, m_2\} \).
4. If \( m_1^TP_{[m_1, m_2]}^\perp m_1<\varepsilon \), stop iterations; otherwise \( i=i+1 \).
5. Find the \( i \)-th target generated at the \( i \)-th stage by \( m_i=\arg \max \{||e||^2\} \), where \( M_{i-1}=[m_1, m_2, ..., m_{i-1}] \) is the target signature set generated at the \( i \)-th stage.
6. The above procedure repeats until the error threshold is met or a predefined number of feature vectors are obtained.

Once the feature vectors \( M_S \) are selected, they define a subspace \( M_S \) in \( \mathbb{R} \), \( (S \Rightarrow \mathbb{P}) \). For any pixel vector \( r_i \) in the scene, a suitable coefficient vector, \( w=(w_1, w_2, ..., w_s)^T \), can be found to satisfy the following approximate linear dependence condition:

\[
\left\| \sum_{i=1}^{s} w_i \phi(m_i) - \phi(r_i) \right\| \leq \varepsilon
\]  

Note that the smaller \( \varepsilon \) was, the more feature vectors were selected, which means that \( \varepsilon \) is an important parameter determining the level of sparsity. Sparsity is related to generalization ability and is considered a desirable property in learning algorithms (Schölkopf and Smola 2002, Vincent and Bengio 2002), as well as in...
signal processing (Duda and Hart 1973). The ability of a kernel machine to correctly generalize from its learned experience to new data can be shown to improve as the number of its free variables decreases, which means that sparsification may be used as a regularization instrument. The feature vector selection can also be viewed as the definition of the hidden layer of a multi-layer neural network (Johnson et al. 1992). The number of hidden neurons on which the data are projected corresponds to the number of feature vectors selected in $\mathcal{R}$. However, the feature vectors selection (FVS) has the advantage of providing a number of hidden neurons for a given kernel. The kernel trick makes it possible to proceed to select data in the feature space $\mathcal{R}$. For a neural network, the feature space remains hidden and is never explored.

Moreover, $M_S$ being a basis implies that the matrix $M_S^T M_S$ has the full rank and its inverse exists. Hence an orthogonal basis can be defined:

$$\Omega = M_S (M_S^T M_S)^{-1}$$

We then apply the dot product projection to obtain the parameter space of original pixels:

$$\pi_\phi(r) = (M_S^T M_S)^{-1} M_S^T \phi(r) K^{-1} K(M_S, r)$$

Since rank($\pi_\phi$) = $S$ and the parameter space is linear, this suggests that a linear blind source separation on $\pi_\phi(r)$, based on simultaneous digitalization techniques, can be performed to obtain $S$ linear directions of separated nonlinear components in the input space (Harmeling et al. 2003). However, the solutions with the highest eigenvalues may not correspond to the source, but to some function of these sources (Dominique and Alistair 2003). In our experiment, owing to the complexity of remote sensed data, it did not generate acceptable results by performing linear blind source separation on parameter space. Hence we resorted to regression on $w_i$ to obtain the proportional value.

$$\hat{a}_i = w_i A + b$$

Given a set of training data $(w_i, \hat{a}_i)$ where $w_i$ is projection (10) onto $\pi_\phi$ of the sample $r_i$, the goal of the algorithm is to estimate the vector $\hat{a}_i$, the output of the function. The mean square error is minimized using the classic LSE technique. It has been shown that equation (11) leads to the posterior probability estimation and gives the best approximation of the Bayes decision function (Duda and Hart 1973).

### 3. Experiments

To evaluate the performance of the proposed algorithm, three datasets were used. One is a simulated dataset based on Hapke’s bidirectional approximating reflectance model (Hapke 1981), which is used to test nonlinear approximating ability of the proposed algorithm. The second is a synthetic airborne visible infrared imaging spectrometer (AVIRIS) image for comparing the performance of the proposed kernel least square (KLS) algorithm with two other algorithms. The third dataset is real image, which was derived from IKONOS and Landsat-7 ETM+ images. Three
typical performance measures (Li et al. 2002, Liu and Wu 2005) were used: root mean square errors (RMSE); bivariate distribution functions (BDF) between the real and estimated subpixel proportions; and error bound. The BDF can help visualize the accuracies of prediction by mixture models, while the RMSE is used for evaluating total accuracies.

3.1 Simulated dataset

Hyperspectral reflectance spectra were selected from the original pushbroom hyperspectral imager image with 80 bands, which was acquired on 9 September 1999 over Xiaqiao Town, Jiangsu Province, China. The wavelength of this image ranges from 0.42 μm to 0.85 μm. The main ground materials in the area are road, water and vegetation. Their spectra are distinguished from each other; hence, selected as endmembers for simulating data. For each of the three ground-cover materials 15 spectra have been chosen to keep the spectral variety. For each class, these spectra were separated into two groups, among which five spectra were selected randomly for system training and the other ten spectra were utilized for system testing. As such, the testing data do not include any information from the training data, which ensures a fair and reliable testing. Figure 2 shows the 45 hyperspectral signals that are utilized for the simulated dataset.

Given three endmember spectra, \( r_1, r_2 \) and \( r_3 \), and their abundances, \( a_1, a_2 \) and \( a_3 \), the nonlinear mixed pixel spectra are synthesized based on Hapke’s bidirectional approximating reflectance model.

\[
R(i, e) = \frac{wH(\mu)H(\mu_0)}{4(\mu + \mu_0)}
\]

(12)

where \( R(i, e) \) = bidirectional reflectance, \( w \) = single scattering albedo, \( \mu = \cos i \), \( \mu_0 = \sin i \), \( i \) = angle of emergence

\[
H(\mu) = \frac{1 + 2\mu}{1 + 2\mu\sqrt{1 - w}}
\]

is multi-scattering function.

In all of the following experiments, the approximation to \( H(\mu) \) should not greatly affect the abundance estimation. According to Johnson et al. (1992), the albedo of a mixture is a linear combination of the single-scattering albedos of its endmember constituents. Thus, if we can convert all of the pixels in an image to albedo, the popular linear unmixing methods for estimating abundance should provide more accurate results (Kerri et al. 2001). A mixed pixel spectrum, \( \tilde{R} \), can be synthesized:

Figure 2. The pushbroom hyperspectral imager reflectance signals of road, water and vegetation utilized for the simulated nonlinear mixture spectra.
1. Transfer $r_1$, $r_2$ and $r_3$ to the single scattering albedo $w_i$ using equation (12).

2. Generate three stochastic values within $[0,1]$ as $a_1$, $a_2$ and $a_3$ satisfying $a_1 + a_2 + a_3 = 1$; hence, the mixture albedo turns to be $\rho = \sum_{i=1}^{k} a_i w_i$.

3. Inverse the mixture albedo $\rho$ to the mixture reflectance $R$, and add additional Guassia noise $N(0,1)$.

4. A mixed pixel spectrum, $\bar{R}$, can be synthesized as $\bar{R} = R\left(1 + \frac{N(0,1)}{SNR}\right)$, where $SNR$ is the signal to noise ratio.

With the simulated data, the kernel function was chosen as the Polynomial kernel with parameter $d=2$. It can be inferred from table 1 that with the threshold $\varepsilon$ decreasing, the number of selected feature vectors increases. Thus, $\varepsilon$ is an important parameter to determine the level of sparsity.

A total of 2500 simulated data points are generated with $SNR=20:1$. Among them 500 samples were selected randomly for training and the rest for testing. Since $SNR$ equals to 20:1, the threshold $\varepsilon$ is defined as 0.05, to exclude the effect of noise.

It can be found from figure 3(a) that the RMSE of the three materials is 0.016, 0.062,
and 0.074, respectively. Points along a 1 : 1 line on the BDF graph indicate that a prediction matches exactly with the real proportion. The smaller the difference between the predicted and the real proportion, the closer the point to the diagonal 1 : 1 line. Figure 3(a) illustrates that almost all points within the 10% error bound reach the accuracy of 98.8%, 93.6% and 89.2% respectively. The same data were used for traditional linear regression. It can be found from figure 3(b) that the linear unmixing results is worse than the former with the RMSE of the three materials as 0.1051, 0.2940, and 0.2589, respectively. Points along a 1 : 1 line on the BDF graph indicate that a prediction is tremendously partial to the real proportion. The main reason for the partiality is that the simulated data have strong nonlinear properties. This experiment shows that classical linear regression is not suitable for nonlinear data, but our algorithm can perform rather well.

3.2 Synthetic dataset

An AVIRIS dataset (scene 1 of the 1997 flight line) of Moffett Field, California was also selected to test our algorithm. The dataset is available online (http://aviris.jpl.nasa.gov) in radiance and reflectance units. This dataset (512 × 614 pixels and 224 spectral bands per scene) contains 12 bit digital numbers without any atmospheric corrections for absorption. To remove atmospheric, water absorption bands and low Signal-to-Noise Ratio (SNR) bands, we selected 90 bands, which wavelength ranges from 0.468 μm to 1.284 μm to compose an image for our experiment. Figure 4 is a sub-scene (350 × 350) extracted from the original image.

Synthetic imagery has the advantage of avoiding co-registration and radiometric correction errors between the lower and higher resolution images. The synthetic images (70 × 70) that we generated were degraded at a 5 : 1 scale using an average filter. Degradation of a hard classification yields fractional images for each class during the classification. These fractional images do not contain any uncertainty as it originates from degradation instead of a classification process. Consequently, the

![Figure 4](image_url)
subpixel proportion estimation error only reflects the performance of the proposed algorithm and validation is facilitated as the original hard classification can be used as reference material.

The kernel function for the experiment was also chosen as the Polynomial kernel, 500 (about 10%) pixels were randomly selected from the degraded image as training samples. Since the multinomial rank of kernel function decides learning ability, it is important to select the parameter. Table 2 indicates that the number of feature vectors becomes larger with the increase of the multinomial rank $d$, whereas the RMSE is not the case. When $d$ equals to 3, the algorithm generates the best results with the RMSE of water, soil and vegetation being 0.0291, 0.0793 and 0.0805, respectively.

To test the performance of the proposed algorithm, SVR and RBF were also implemented with the data for comparison purposes. Figure 5 illustrates that all three kernel-based algorithms performed very well. The RMSE of all materials with KLS, SVR and RBF are less than 0.09. Among them, the fractions of water and soil have the smallest RMSE by using KLS algorithm; however, RBF achieves the smallest RMSE in the vegetation fraction. The average RMSE of KLS, SVR and RBF is 0.0629, 0.0652 and 0.0657, respectively. This experiment shows that the proposed algorithm outperforms SVR and RBF in terms of RMSE. It can also be found from figure 5 that the nonlinear method greatly outperforms linear mixture spectral analysis (LMSA) even if the procedure of producing proportion in synthetic imagery is linear as the variance of homogeneous materials may result in a nonlinear mixture among materials (Song 2004). The classical linear regression is not suitable for this case, which leads to comparative lower accuracy.

### 3.3 ETM dataset

A real ETM+ dataset was also conducted in order to test the performance of proposed algorithm. We obtained the Landsat/ETM+ standard product over the Shenzhen special zone of China acquired on 2 January 2000 from the Beijing Remote Sensing Ground Station. This product was geo-referenced with the spatial resolution of 30 m and the solar zenith angle of 34.5. The IKONOS data over the same region were acquired on 20 December 2001. Although one year difference exists between the two images, both of them were acquired in the same season. Hence, their reflectance spectra are assumed to be similar. In such a case, the spectral signature may be obtained through inverse spectral mixture analysis to derive endmember signatures or training pixels given the fractional covers that can be obtained from the IKONOS image. As a result, we first performed geometric

---

Table 2. The relationship between multinomial rank and feature vector, and root means square error (RMSE) correspondingly ($\varepsilon=0.001$).

<table>
<thead>
<tr>
<th>Rank</th>
<th>Feature vectors</th>
<th>Water</th>
<th>Soil</th>
<th>Vegetation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td></td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>0.0772</td>
<td>0.1359</td>
<td>0.1346</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>0.0305</td>
<td>0.0830</td>
<td>0.0839</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>0.0291</td>
<td>0.0793</td>
<td>0.0805</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>0.0296</td>
<td>0.0835</td>
<td>0.0881</td>
</tr>
<tr>
<td>5</td>
<td>67</td>
<td>0.0632</td>
<td>0.0968</td>
<td>0.0980</td>
</tr>
<tr>
<td>6</td>
<td>90</td>
<td>0.0808</td>
<td>0.1303</td>
<td>0.1323</td>
</tr>
</tbody>
</table>
correction on it, and so each corrected pixel of the output image has the same universal transverse mercator (UTM) coordinates as the ETM+ data. The ETM+ spatial resolution was re-sampled to 32 m, eight times the IKONOS data (4 m). Three steps were included to calculate the fractional land cover for each ETM+
pixel. First, the IKONOS image was classified into four classes (city, vegetation, soil and water) using a maximum likelihood classifier assisted by interpretation as the shade reflection intertwines with water. Since the IKONOS image has a higher spatial resolution, the classified accuracies deem to be higher with manual interpretation. However, we cannot provide quantitative evaluation due to a lack of field investigations. Recent studies have found that the typical scale of urban reflectance is between 10–20 m (Small 2003). Therefore, most 4 × 4 m IKONOS pixels are spectrally homogeneous and can be reasonably assumed to be pure materials. Second, the classified IKONOS image was registered with the ETM+ images using the ground control points with the total RMSE error less than 0.3. Third, the classification map was overlaid to the ETM+ image to estimate the fraction of land cover classes for each pixel at the ETM+ data scale.

Figure 6. ETM+ and IKONOS reflectance image over the SHENZHEN special zones of China acquired on 21 December 2001 (circle points are ground control points).
Without loss of generality, a sub-area (150 × 150) on the ETM image was selected for our experiment. Figure 6 shows the reflectance sub-image of ETM (a) and IKONOS (b) for the selected area. For the purpose of training 200 ETM pixels from each class, with abundance larger than 0.5, were randomly selected from the image. The sizes of the training samples were deemed sufficient for training the algorithm in the study areas (800 pixels, about 4%). The testing dataset includes all the pixels of the entire image. For this dataset, the RBF kernel was chosen as the kernel function. Cross validation method is used to estimate the parameter $\sigma$ and $\varepsilon$, which ranges from $[0.01, 0.05, 0.1, 0.5, 1.0]$ and $[1e-7, 1e-6, 1e-5, 1e-4, 1e-3]$ respectively. Following this procedure, the kernel parameter optimizes to $\sigma=0.1$, and the threshold $\varepsilon=1e-5$. The RMSE of the four materials is 0.143, 0.115, 0.074 and 0.146, respectively. As can be observed from figure 7, most of the points lie between the 10% difference lines. These error distribution patterns are also clearly shown in figure 8 in a different format. For each material, the graph indicates how many predictions fall within a given percentage of field measurement, as the city class is not an endmember in the scene, but a synthesis of road, building and other concrete constructions. The accuracies of city and water are lower than vegetable and soil, yet the reflectance spectra of the city are large. The predicting accuracies using the real data are lower than the simulated and synthetic data probably because: (i) there may exist geo-referenced and geometric corrected errors; (ii) the classification error of IKONOS may reduce unmixing accuracies; (iii) there is radiometric correction and reflectance retrieval between IKONOS and ETM+; (iv) the difference among homogeneous spectra is large.

Comparison was also conducted with this dataset using the same kernel function and parameter $\sigma$. Table 3 shows the results of all three algorithms. All the material fractions by KLS algorithm have the smallest RMSE except water, in which RBF achieves the smallest RMSE. The average RMSE of KLS, SVR and RBF is 0.119,

![Figure 7. Bivariate distribution function of the real test data, from left to right is barren land (in the city), vegetable, soil and water, the root means square error (RMSE) of each material is listed on the upper left corner of each chart.](image)
0.124 and 0.129, respectively. It also shows that KLS uses 16 feature vectors, which is less than those of SVR (19 support vectors) and RBF (36 hidden nodes). Thus, the experiments show that the proposed algorithm outperforms RBF and SVR again.

4. Conclusions

This paper has presented a kernel-based least square mixture model making use of the nonlinear characteristics of spectral mixture to obtain the higher unmixing accuracies. With the simulated data, the proposed method accomplishes perfect results for all three materials with RMSE less than 0.08, while the traditional linear regression is not suitable for estimation of this nonlinear dataset with the maximal RMSE more than 0.25. With the synthetic imagery of AVIRIS, all of KLS, SVR and RBF algorithms perform very well. With the real data of ETM+ and IKONOS, the proposed method also achieves satisfactory accuracies with the RMSE of the four materials being 0.143, 0.115, 0.074 and 0.146, respectively. Comparative experiments show that the proposed algorithm outperforms RBF, SVR and LMSA.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 40471088, 40523005, the 973 Program of the People’s Republic of China under Grant No. 2006CB701302, and the Foundation of State Key

Table 3. Comparison of root means square error (RMSE) by the three kernel-based algorithms.

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature vectors</th>
<th>City</th>
<th>Vegetable</th>
<th>Soil</th>
<th>Water</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSMA</td>
<td>–</td>
<td>0.209</td>
<td>0.111</td>
<td>0.108</td>
<td>0.198</td>
<td>0.156</td>
</tr>
<tr>
<td>RBF</td>
<td>36</td>
<td>0.156</td>
<td>0.116</td>
<td>0.079</td>
<td>0.152</td>
<td>0.126</td>
</tr>
<tr>
<td>SVR</td>
<td>19</td>
<td>0.148</td>
<td>0.111</td>
<td>0.086</td>
<td>0.151</td>
<td>0.124</td>
</tr>
<tr>
<td>KLS</td>
<td>16</td>
<td>0.143</td>
<td>0.115</td>
<td>0.074</td>
<td>0.146</td>
<td>0.119</td>
</tr>
</tbody>
</table>

LSMA=; RBF=radial basis function; SVR=support vector regression; KLS=kernel least square.
References


