

Hyperspectral Image Fusion Using Spectrally Weighted Kernels

Baofeng Guo, Steve Gunn, Bob Damper and James Nelson

Image, Speech and Intelligent Systems Group
School of Electronics and Computer Science
University of Southampton
{bg,srg,rid,jn}@ecs.soton.ac.uk

Abstract - *Target detection from hyperspectral imagery requires the fusion of information from hundreds of spectral bands. In this paper, we study such fusion in the context of hyperspectral image classification. Previous research applying kernel methods such as support vector machines (SVMs) to hyperspectral image classification has achieved performance competitive with the best available algorithms. However, few efforts have so far been made to extend SVMs to fit the specific requirements of this application, e.g., by building tailor-made kernels. To this effect, we propose a novel spectrally weighted kernel. Observation of real-life spectral signatures from the AVIRIS hyperspectral dataset shows that the useful information for classification is not equally distributed across bands. Hence, we propose the use of spectrally weighted kernels to assign weights to different bands according to the amount of useful information they contain. We have carried out experiments on the AVIRIS 92AV3C dataset to assess the performance of the proposed method. Results show potential for improvement in classification accuracy.*

1. Introduction

Traditional information fusion has focused on techniques to integrate data from multiple sensors where one-dimensional or two-dimensional sensor signals are typically used. However, evolution of technology has led to sensors which are currently capable of producing much higher-dimensional data in space and time. Hyperspectral sensors simultaneously measure hundreds of narrow and contiguous spectral bands from a wide electromagnetic spectrum (e.g., from visible light to mid-infrared region, 0.4 to 2.5 μm). The increase of spectral bands makes it possible to reduce overlap between classes [1, 2], and therefore greatly enhances the potential to discriminate subtle spectral difference among materials that multispectral imaging cannot achieve. In recent years, hyperspectral image classification has received significant attention in many civilian and military applications [1]. The purpose of hy-

perspectral image classification is to recognise materials or surfaces by analysing the spectral signatures extracted from hyperspectral sensors. They require the careful design of new algorithms that are able to fuse hundreds of spectral bands whilst minimising the effects from the curse of dimensionality.

One of the main problems encountered with hyperspectral image classification is the high dimensionality of hyperspectral data; for instance, the AVIRIS hyperspectral sensor [3] has 224 spectral bands. The performance of supervised classification methods is governed by the availability of training samples and the dimensionality of input space. Hughes [4] demonstrated the relationship between the recognition accuracy and model complexity for different training dataset sizes. Hughes' results show that the accuracies of recognition algorithms will not always increase with the number of signal dimensions used. Given a fixed amount of training data, classification accuracy, as a function of signal dimension, will reach a maximum and then decline. This is mainly because the limited amount of training data in high-dimensional spaces will degrade the estimation accuracy of a larger number of parameters such as the covariance matrix. In hyperspectral image classification, the amount of training data is always limited, and the proportion of training pixels to the number of bands can be relatively small. Although hyperspectral data have many spectral bands, care must be taken not to simply overfit the data in this higher dimensional space and degrade recognition accuracy. To limit the effect of Hughes phenomenon, or the 'curse of dimensionality', some form of prior should be placed on the class of classifiers. This can take many forms, such as smoothness, or a more direct prior based on a feature selection/extraction stage applied to reduce the dimensionality. Successful feature selection/extraction can reduce the redundancy in the data, but care must also be taken not to remove features that contain discriminatory information. Despite that feature selection or subspace extraction can be a time-consuming pre-processing step, furthermore, the limited number of training data may cause biased class

statistics, which can degrade feature selection or extraction. Hence, obtaining a robust method with a more integrated subspace approach is an important research topic in hyperspectral image classification.

Kernel methods, such as support vector machines (SVMs) [5, 6], have already shown superior performance in many machine learning applications. In contrast to most connectionist techniques, SVM learning involves solving a convex optimisation problem with a unique global solution. SVMs create optimal classifiers in the sense of a maximum margin. Unlike many parametric methods, SVMs do not rely on estimating the density functions underlying the training dataset, and as such is less affected by the high dimensionality of input space [7]. The SVM determines a separating boundary that is defined by a subset of weighted training samples, called support vectors. Both the support vectors and the multiplying coefficients are obtained as the solution to a quadratic program. Hence, as long as the training samples are representative enough, the search for the separating boundary will not be significantly affected by the statistical estimation and the number of training samples.

With the successful application of SVMs to a wide variety of different pattern recognition areas, such as 3D object recognition, image classification, character recognition, etc, SVMs have recently attracted increasing attention in remote-sensed multi/hyper-spectral communities [7, 8, 9, 10, 11, 12, 13, 14]. Previous literature applying SVMs to hyperspectral data classification [10, 11, 12] has shown competitive performance with the best available classification algorithms. However, the full potential of SVMs, such as developing custom kernels to integrate prior domain knowledge has not been fully explored; little attention has been drawn to the characteristic that parts of the spectrum will provide a much richer descriptor for classification than other parts of the spectrum.

In this paper, we propose the use of spectrally weighted kernels (SW kernels) to take better advantage of SVM techniques for hyperspectral image classification. We firstly highlight the non-uniform information distribution of the sensor information for classification among hyperspectral bands based on examples of two real-life spectral signatures and the AVIRIS hyperspectral dataset. Then we propose a simple modification to the kernel function to take into account the difference of the relative utility of each spectral band. An outstanding problem regarding the SW kernels is determining the spectral weight values. To address this we propose a scheme to estimate the spectrally weighted coefficients by computing the mutual information between spectral bands and the training targets derived from the ground-truth reference map, so as to encode the utility difference of various spectral bands.

The remainder of this paper is organised as follows. In Section 2, we discuss the non-uniform information distribution of hyperspectral data. In Section 3, we briefly review the basic SVM method and then propose the spectrally weighted kernels for hyperspectral data classification. We discuss how to use the mutual information to estimate the spectrally weighted coefficients in Section 4. Several experiments are carried out to test the performance of the proposed method, which are presented in Section 5. Finally, we end this paper with some conclusions and a proposal for future work.

2. Unbalanced information distribution

The majority of previous research [10, 11, 12] has directly applied standard SVMs to hyperspectral image classification, without introducing a custom kernel. Typically, generic kernels such as linear, polynomial and Gaussian are applied, without special consideration to the nature of hyperspectral data. This work aims to incorporate the characteristics of hyperspectral data, such as non-uniform information distribution among the spectral bands, spatial characteristics, etc.

The fundamental physical principle underlying hyper/multi-spectral recognition is that the spectral reflectance of the component materials can characterise their molecular composition and shape. The spectral reflectance curves for different materials will be dissimilar. The classification is carried out by analysing the electromagnetic reflectance as a function of the wavelength, i.e., the spectral signature. Hyperspectral sensors sample signals in a wide spectrum and it is to be expected that different parts of spectrum will have differing representative capabilities to distinguish the objects of interest. In some parts of the spectrum materials may have a much more unique spectral reflectance signature than other parts of the spectrum. The Figure 1¹ illustrates two real-life examples of spectral reflectance from two classes of vegetation. There is an overlap between the two classes due to the natural spectral variability of reflectance. If we do not consider the second or higher order statistics, these two classes are more separable in the region [1.6-1.8 μ m] than other heavily overlapped regions like [0.5-1.2 μ m], [1.3-1.5 μ m], [1.9-2.0 μ m] or [2.4-2.6 μ m]. Within the classification framework, we can say that the region [1.6-1.8 μ m] is relatively more descriptive. Therefore, it would be expected that the learning task should place greater emphasis on the sensor information in the region [1.6-1.8 μ m]. To illustrate this further, Figure 2 and Figure 3 show a subset of a hyperspec-

¹ This figure is taken from reference [2].

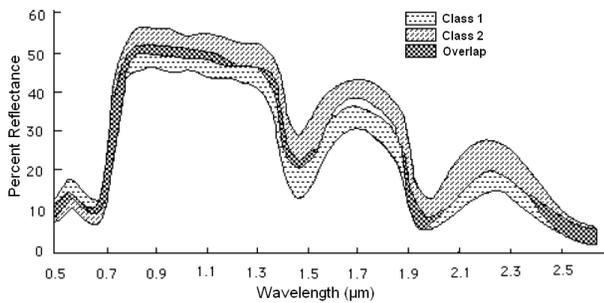


Figure 1. Spectral responses for two classes of vegetation^[2].

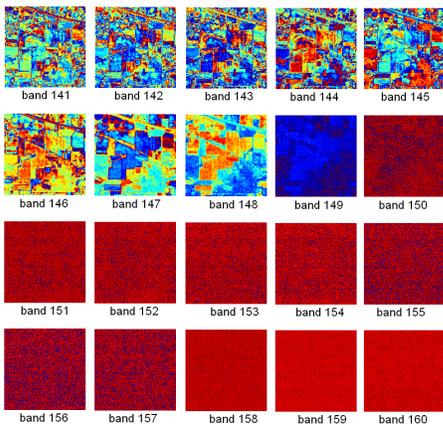


Figure 2. False-colour images of AVIRIS, from spectral band 141 to 160.

tral image² from the AVIRIS dataset. This dataset consists of a 220 dimensional hyperspectral image (4 of the 224 bands from the sensor contain zero response and have been removed) with a spatial resolution of 125×125 pixels.

Comparing these hyperspectral bands with the ground-truth reference map (Figure 4), we can see that the inherent scale varies across the bands. For example, in bands 181 to 200, we can roughly distinguish the outline of the reference map, whereas in bands 150 and 160, this becomes impossible. However, it is not possible to conclude that the lower bands do not contain useful discriminatory information, since they may contribute towards higher order information.

Figure 1 reveals that the non-uniform distribution of discriminatory information can exist in hyperspectral data. In the set of spectral bands, some bands may contain more useful information for classification than other bands. This introduces the potential to emphasise effective spectral bands or windows, with spectral management algo-

² All hyperspectral images illustrated in this paper are transformed using a suitable false colour palette.

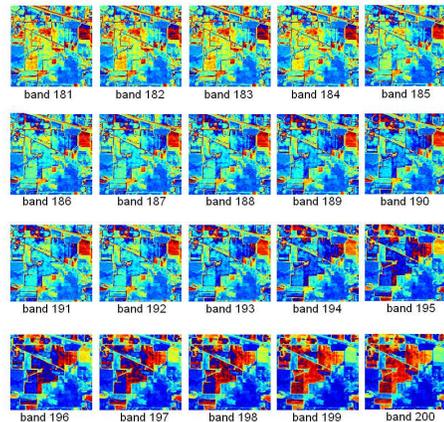


Figure 3. False-colour image of AVIRIS, from spectral band 181 to 200.

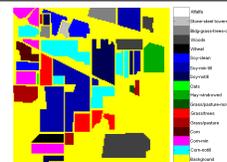


Figure 4. AVIRIS reference map.

rithms. In the SVM-based classification framework, a straightforward approach is to modify the kernel by assigning different weights to different bands automatically, or adaptively, according to the amount of useful classification information contained within the band. To this end, we propose the use of spectrally weighted kernels to better exploit this specific characteristic in hyperspectral data.

3. Spectrally weighted kernels

To present the proposed spectrally weighted kernel, let us first introduce the relevant SVM formulae that will be used in our discussion. A detailed introduction of SVMs theory can be found in [5, 15, 16].

Let $\mathbf{x}_i = (x_1, x_2, \dots, x_N)$ be a N -dimensional hyperspectral data vector, $y_i \in \{+1, -1\}$ be the classification targets, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)$, be the Lagrange multipliers, M the number of examples and b a threshold. The SVM classifier can be represented as:

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^M y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right)$$

where $K(\mathbf{x}, \mathbf{x}') = \Phi(\mathbf{x})^T \Phi(\mathbf{x}')$ is an appropriate kernel function which has a corresponding inner product expansion, Φ . The commonly-used functions are polynomials and Gaussian radial basis functions (RBFs):

$$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^d \quad (1)$$

$$K(\mathbf{x}, \mathbf{x}') = \exp\{-\gamma\|\mathbf{x} - \mathbf{x}'\|^2\} \quad (2)$$

For a hyperspectral data vector $\mathbf{x}_i = (x_1, x_2, \dots, x_N)$, the component x_i corresponds to the reflectance value in a specific hyperspectral band i , $1 \leq i \leq N$. Generic kernels, e.g. Equation 1 and Equation 2, regard each component x_i with equal emphasis in their projection into feature space. However, this paper has motivated that it is advantageous to moderate the spectral information according to the richness of the descriptor. For example if the component x_i is a reflectance value in the spectrum where two classes can be clearly separated such as the region [1.6-1.8 μm] in Figure 1, providing this feature with a larger effect in feature space could improve classification, and similarly reducing it when it confers little description.

Under the SVM framework, the simplest way of reflecting the above consideration is via a modification of the kernel function to assign different weights to different bands. Hence, we assign a series of weights, $\mathbf{s} = (s_1, s_2, \dots, s_N)$, to scale each feature x_i in the hyperspectral data vector before mapping into the feature space. To simplify notation we introduce a diagonal matrix, $\mathbf{S} = \text{diag}(\mathbf{s})$. Given this weighting the spectrally weighted kernels for a polynomial and an RBF can be written as:

$$K_{sw}(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{S}^T \mathbf{S} \mathbf{x}' + \mathbf{1})^d \quad (3)$$

$$K_{sw}(\mathbf{x}, \mathbf{x}') = \exp\{-\gamma\|\mathbf{S}(\mathbf{x} - \mathbf{x}')\|^2\} \quad (4)$$

The necessary and sufficient condition for deciding whether a function is a kernel is given by the Mercer's theorem. It is easy to prove that the spectrally weighted kernels still satisfy Mercer's condition, since they correspond to a simple scaling of the input space. Substituting the spectrally weighted kernels into Equation 1, gives the corresponding spectrally weighted SVM.

4. Estimation of the spectrally weighted coefficients

In the SW kernels, the spectrally weighted coefficients s_i , determine the relative influence of each band to the kernel values. A suitable paradigm is required to estimate these coefficients. Possible methods include learning or estimating these from the data, or through prior knowledge. Here, we propose a straightforward method which uses the Mutual Information (MI) to estimate the spectrally weighted coefficients.

4.1. Mutual information

MI is a basic concept from information theory to measure the statistical dependence between two random variables [17], and has been applied in many areas including image registration [18, 19] and information fusion [20]. Given

two random variables, A and B , with marginal probability distributions, $p_A(a)$ and $p_B(b)$, and joint probability distribution $p_{AB}(a, b)$, MI is defined as:

$$I(A, B) = \sum_{a,b} p_{AB}(a, b) \log \frac{p_{AB}(a, b)}{p_A(a)p_B(b)} \quad (5)$$

MI is related to entropy by the following equations:

$$\begin{aligned} I(A, B) &= H(A) + H(B) - H(A, B) \\ &= H(A) - H(A|B) \\ &= H(B) - H(B|A) \end{aligned} \quad (6)$$

where $H(A)$ and $H(B)$ are the entropy of A and B , $H(A, B)$ their joint entropy, and $H(A|B)$ and $H(B|A)$ the conditional entropies of A given B and of B given A , respectively. Using the definition of Shannon entropy, these entropies can be written as:

$$H(A) = - \sum_a p_A(a) \log p_A(a) \quad (7)$$

$$H(A, B) = - \sum_{a,b} p_{AB}(a, b) \log p_{AB}(a, b) \quad (8)$$

$$H(A|B) = - \sum_{a,b} p_{AB}(a, b) \log p_{A|B}(a|b) \quad (9)$$

MI can measure the statistical dependence or information redundancy between two random variables. If we model hyperspectral images and the corresponding reference map³ as random variables, MI can be used to estimate the dependency between a spectral image and the reference map. This is helpful to investigate how much information a spectral image contains about the reference map. Since the reference map is the objective of training, the MI between a spectral band and its reference map can be used as an approximation to the utility of this spectral band to classification. However, it should be noted that this method only considers the effect of each individual band to classification independently and as such will not detect higher order interactions for more complex situations.

4.2. Estimation of the spectrally weighted coefficients using MI and reference map

Considering random variables A and B as two images, marginal distributions of A and B can be estimated by counting the number of times each value occurs in the images and dividing by the total number of pixels [19] (i.e., the normalised histogram). Thus estimation for marginal distribution $p_A(a)$, $p_B(b)$ and the joint probability distribution

³ The reference map denotes a marked image of the scene, i.e., the ground truth map, in which each pixel is correctly labelled to a class

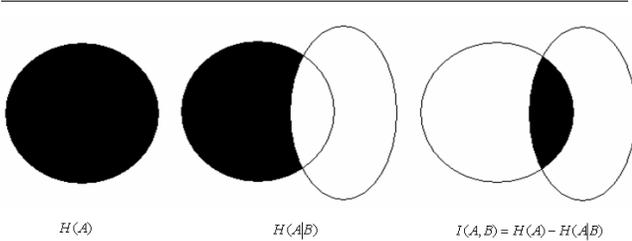


Figure 5. Illustration of Mutual Information.

$p_{AB}(a, b)$ can be obtained by normalisation of the marginal and joint histograms of both images.

In supervised hyperspectral image classification, e.g., SVM-based methods, a reference map is usually available for the aim of training or learning. Let A be the reference map and B a spectral image band, the entropy $H(A)$ is known to be a measure of the amount of uncertainty about A , while $H(A|B)$ is the amount of uncertainty left in A when knowing B . From Equation 6, $I(A, B)$ is the reduction in the uncertainty of reference map A by the knowledge of hyperspectral image B . Hence, MI can be interpreted as the amount of information that hyperspectral image B contains about reference map A (see Figure 5). In other words, MI has the ability to reflect the amount of information that a hyperspectral band B contains about the reference map A . Since the reference map contains ground truth information, this can be used to estimate the individual band contribution in isolation to the classification task. Thus, $I(A, B)$ can be used to estimate the spectrally weighted coefficients. In our SW kernels, we use MI to estimate the weighted coefficients in Equation 3 and Equation 4. The higher the MI, the higher the weight that is assigned to the spectral band.

5. Experiments

To test the proposed method, we implemented the spectrally weighted scheme as shown in section 3 and 4, and ran the algorithm on AVIRIS 92AV3C hyperspectral dataset ⁴.

We assessed the performance of the proposed method by comparing to a standard SVM with no spectral weighting of the kernel as adopted in [7, 10, 9]. The performance measurement adopted assessed the classification accuracy of the proposed method on a hold-out set.

We chose four types of ground plants as the classification objects including corn-notill (class A), soybeans-notill (class B), soybeans-min (class C), and grass/trees (class D). In the four classes, we randomly chose 20% of examples from every class as the training set. The remaining 80% of pixels form the testing set. The SVM hyperparameters, C ,

⁴ AVIRIS dataset can be downloaded from a public website at Purdue University: <ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/>.

	Overall classification accuracy
Method using polynomial kernel	80.11%
Method using SW polynomial kernel	80.65%

Table 1. Comparison of overall classification accuracy, polynomial kernel, $d = 3$

	Overall classification accuracy
Method using RBF kernel	76.33%
Method using SW RBF kernel	77.78%

Table 2. Comparison of overall classification accuracy, RBF kernel, $\gamma = 1$

and any additional kernel parameters, were optimised using part of the training set as a validation set.

SVMs were originally conceived for binary classification problems. There are two main approaches to combining several binary SVM classifiers to form a multi-class machine:

1. ‘one against one’: This method constructs $\binom{k}{2} = k(k-1)/2$ classifiers, where k is the number of classes. Each classifier is trained pairwise on data from two classes. A majority voting strategy is applied to decide which class the test data is in.
2. ‘one against the others’: The method constructs k SVM classifiers to separate class k from the remaining classes. Then all k classifiers are applied to the test data, and the label of the classifier with the strongest prediction is selected. We adopt the ‘one against one’ and majority voting as our multi-class scheme. As a result, the four training sets are used to train six binary SVM classifiers, namely, $C_{A|B}$, $C_{A|C}$, $C_{A|D}$, $C_{B|C}$, $C_{B|D}$ and $C_{C|D}$.

The experimental results are shown in Table 1 to Table 4. Table 1 and Table 2 are comparisons of the overall classification accuracy based on polynomial and RBF kernel. From Table 1 and Table 2, it can be seen that the proposed method based on the spectrally weighted method outperforms the un-weighted method in terms of overall classification accuracy in both the polynomial and RBF kernel. As analysed later, the extent of the improvement may depend on several factors, like the scene content, the estimation accuracy of weighted coefficients, etc.

To further illustrate the improvement over the un-weighted method of the proposed method, we also present those classification accuracies based on each individual classifier and enable a detailed comparisons between

	Method using polynomial kernel	Method using SW polynomial kernel
$C_{A B}$	90.20%	90.92%
$C_{A C}$	86.24%	87.18%
$C_{A D}$	99.86%	99.86%
$C_{B C}$	84.08%	83.56%
$C_{B D}$	99.66%	99.66%
$C_{C D}$	98.68%	98.78%

Table 3. Comparison of individual classifiers' accuracy, polynomial kernel, $d = 3$

	Method using RBF kernel	Method using SW RBF kernel
$C_{A B}$	89.27%	89.63%
$C_{A C}$	83.08%	85.22%
$C_{A D}$	99.78%	99.86%
$C_{B C}$	81.72%	80.64%
$C_{B D}$	99.66%	99.66%
$C_{C D}$	98.59%	98.64%

Table 4. Comparison of individual classifiers' accuracy, RBF kernel, $\gamma = 1$

the un-weighted scheme and the proposed scheme. The advantage of so doing is to see exactly which classifier(s) has been improved, since the overall classification only indicates an average result on the whole and often conceals some useful information about the improvement distribution among each individual classifier and to what extent this improvement is. Additional benefit can also be obtained to find which individual classifier performed poorly while the overall improvement is verified. This can provide useful information to assess and then improve the multi-class scheme, e.g., majority voting.

Table 3 and Table 4 show the classification results of each individual classifiers, i.e., $C_{A|B}$, $C_{A|C}$, $C_{A|D}$, $C_{B|C}$, $C_{B|D}$ and $C_{C|D}$, based on polynomial and RBF kernel, respectively. In this experiment, only data from the related classes attend the test of the corresponding classifier. For example, only data from classes B and C are used to test the classifier $C_{B|C}$, and other data from classes A or D will not attend this test. From Table 3 and Table 4, it can be seen that the spectrally weighted scheme outperformed or drew in five of all six classifiers both in polynomial and RBF kernel.

It can be seen from Table 1 to Table 4 the classification accuracy is improved by adopting the proposed method. However, the extent of changes is relatively not very significant. To explain this, three possible reasons are analysed and listed as follows:

1. The kernel function has not yet been particularly adjusted to the specific classification task, for example to be optimised in the SVMs learning procedure. In

other words, the weighted scheme proposed in this section, i.e., directly imposing the weighted coefficients to the input vector in the polynomial or RBF kernel functions has not fully tuned to use the information-variation among different spectral bands in SVMs. In this sense, the proposed kernel function may not necessarily produce a better feature space and higher classification accuracy. Further improvement of this method is the next step of our research.

2. We verified that the spectrally weighted coefficients calculated from the mutual information matches our visual evaluation of the amount of discriminatory information contained in each spectral band. The more amount of discriminatory information, the higher the MI value is. However, the relatively utility evaluated on the basis of 'human visual observation' dose not always guarantee that it is proportional to the real amount of discriminatory information. Our future research will look into if a more suitable scheme to calculate the spectrally weighted coefficients exists.
3. Finally, the proposed utility evaluation method estimates the importance of each spectral band in isolation, and therefore may fail to find the band combinations which are important together but poorly performed individually.

6. Conclusions

In this paper, we have proposed an extension to the SVM-based classification method using spectrally weighted kernels for hyperspectral image integration. This research is motivated by the observation that the useful information for classification is not equally distributed among each spectral band. We have shown that by a simple spectral customisation of the kernel using a mutual information criterion it is possible to increase the classification performance. Further research is aimed at a more advanced customisation of the kernel, based on learning an improved kernel from the data. As such this procedure should be able to exploit higher order signature information that is contained in the spectrum rather than the simple independent procedure adopted in this paper. To this end the approach has potential to offer significant performance increases over techniques operating in generic kernel spaces. Other possible research is also identified such as revising the estimation scheme of the spectral weighted coefficients to better reflect the utility of each spectral band; and investigation of other approaches such as relevance vector machines.

Acknowledgements

This research was supported by the Data Information Fusion Defence Technology Centre, United Kingdom, un-

der DTC Project 8.2. The authors would like to thank Erik Blasch and other reviewers for comments and suggestions relating to this paper.

References

- [1] D. Landgrebe. Hyperspectral image data analysis. *IEEE Signal Processing Magazine*, 19(1):17–28, 2002.
- [2] D. Landgrebe. On information extraction principles for hyperspectral data: A white paper. Technical Report, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, U.S., 1997.
- [3] AVIRIS website. NASA/JPL: airborne visible/infrared imaging spectrometer. <http://aviris.jpl.nasa.gov/>.
- [4] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transaction on Information Theory*, 14:55–63, 1968.
- [5] C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2(2):121–167, 1998.
- [6] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, Pittsburgh, Pennsylvania, United States, 1992.
- [7] J. Gualtieri and R. Cromp. Support vector machines for hyperspectral remote sensing classification. In *Proceeding of the 27th AIPR Workshop: Advances in Computer Assisted Recognition*, pages 121–132, Washington DC, 1998.
- [8] M. Brown, H.G. Lewis, and S.R. Gunn. Linear spectral mixture models and support vector machines for remote sensing. *IEEE Trans. on Geoscience and Remote Sensing*, 38(5):2346–2360, 1999.
- [9] C. Huang, L.S. Davis, and J.R. Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749, February 2002.
- [10] C.A. Shah, P. Watanachaturaporn, M.K. Arora, and P.K. Varshney. Some recent results on hyperspectral image classification. In *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, NASA Goddard Spaceflight centre, Greenbelt, MD, October 2003.
- [11] M. Lennon, G. Mercier, and L. Hubert-Moy. Classification of hyperspectral images with nonlinear filtering and support vector machines. In *Proceeding of IGARSS 2002*, oronto, Canada, June 2002.
- [12] F. Roli and G. Fumera. Support vector machines for remote-sensing image classification. In Sebastiano B. Serpico, editor, *Proc. of SPIE Image and Signal Processing for Remote Sensing VI*, volume 4170, 2001.
- [13] G. Camps-Valls, L. Gomez-Chova, J. Calpe-Maravilla, J.D. Martin-Guerrero, E. Soria-Olivas, L. Alonso-Chorda, and J. Moreno. Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1530–1542, July 2004.
- [14] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, Aug 2004.
- [15] V. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, September 1999.
- [16] S.R. Gunn. Support vector machines for classification and regression. Technical Report, School of Electronics and Computer Science, University of Southampton, Southampton, U.K., 1998.
- [17] N. Blachman. The amount of information that y gives about x. *IEEE Transactions on Information Theory*, 14(1):27–31, 1968.
- [18] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information medical imaging. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [19] J. Pluim, J. Maintz, and M. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.
- [20] E. P. Blasch and M. Bryant. Information assessment of sar data for atr. In *Proceedings of IEEE National Aerospace and Electronics Conference*, pages 414–419, Dayton, OH, July 1998.