Adaptive Band Selection for Hyperspectral Image Fusion Using Mutual Information

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 - *Hyperspectral imagery consists of hundreds of spectra or bands whose intensity is measured at various wavelength. Fusing the multiple spectral bands can provide more potential to differentiate between natural and man-made objects, and significantly improve the capability of target detection and classification. Spectral band or wavelength selection is one of the fundamental problems in hyperspectral data fusion. It is one instance of the classical optimal subset selection problem, which is known to be computationally hard. In this paper, we propose a new information-based band selection method for hyperspectral image fusion, which uses an adaptive measurement of mutual information (MI). As derived from the concept of entropy, MI measures the statistical dependence between two random variables and therefore can be used to evaluate the relative utility of each band to classification. Experiments on the AVIRIS dataset show that the method effectively identifies redundant spectral bands. Removing 15% of the total bands increases accuracy by 1.76% relative to performance on all bands, whereas removing 45% of the bands gives only 1.34% loss of accuracy.*

1. Introduction

Hyperspectral sensors simultaneously measure hundreds of narrow and contiguous spectral bands with a fine spectral resolution, e.g., $0.01\mu m$. Hyperspectral data fusion requires the careful design of new algorithms that are able to fuse hundreds of spectral bands without suffering from the curse of dimensionality. In this paper, we study this issue in the context of hyperspectral bands selection.

The large number of spectral bands presents several significant challenges to hyperspectral image fusion. First, the increased number of spectral bands means the higher dimensionality of hyperspectral data. For instance, the AVIRIS hyperspectral sensor [1] has 224 spectral bands ranging from $0.4\mu m$ to $2.5\mu m$, and the original signature is 224 dimensional. It is known that the dimensionality of input space strongly affects the performance of many supervised classification methods [2]. Although there may be hundreds of bands available for analysis, not all bands in the whole spectrum contain the discriminatory information for classification. In some parts of the spectrum, certain material might have sole spectral reflectance feature; on the other parts of the spectrum, the uniqueness of its spectral reflectance might be not very significant. Furthermore, the high dimensionality inevitably results in a larger volume of data. Whether using conventional detection algorithms or modern methods, the requirements for storage space, computational load and communication bandwidth are factors particularly in real-time applications that have stringent constraints.

To limit the negative effects incurred by higher dimensionality, huge data volume and redundant information, it is advantageous to remove parts of the spectral bands which convey no discriminatory information. In the past, many band selection techniques have been proposed [3, 4]. These methods can be roughly categorised into three groups, i.e., search-based methods, transform-based methods and information-based methods. When the number of spectral bands is the order of one hundred, as in hyperspectral data, the search-based methods are unfeasible due to the problem of 'combinatorial explosion'. To improve the search efficiency, several optimisation algorithms have been applied, such as hill climbing, genetic algorithm, and greedy approaches. However, considering the immense number of possible band combinations, the computational cost is still prohibitively high and such approaches suffer local minima problems. Transform-based methods apply matrix transforms like Principal Component Analysis (PCA) or eigenvector analysis to project data onto a space of lower dimension [5, 6]. Such a transformation usually causes a loss of the original physical meaning of the spectral data, which would make it difficult to interpret intuitively, for example to further exploit the resultant bands in data-level fu-

0-7803-9286-8/05/$20.00 © 2005 IEEE

sion. The other problem is that this technique usually requires the capture of the full data cube before applying the transform, which is a disadvantageous factor for real-time processing. The information-based methods directly measure the information content of each individual band. If the measured information content is related (e.g., proportional) to the level of discriminatory capability, the band selection can be carried out by choosing those bands with the higher information content. Common information metrics include the entropy, the contrast and correlation [4, 7]. Compared with the transform-based methods, one advantage of the information-based methods is that the selected subset of hyperspectral data still retains its original appearance. Other important band selection techniques include a trade-off scheme between the spectral resolution (i.e., wideband or narrowband) and spatial resolution [8], maximisation of Spectral Angle Mapper [9], high-order moments [10], wavelet analysis [11].

In this paper, we propose a new information-based band selection method for hyperspectral image classification, which uses an adaptive measurement of mutual information (MI). As derived from the concept of entropy, MI measures the statistical dependence between two random variables and therefore can be used to evaluate the relative utility of each band to classification. Although entropy [4, 7] and mutual information [5, 12, 13] have potential for band selection, their usefulness was not fully exploited. One of the significant weaknesses regarding the MI-based method is that it heavily relies on the availability of a reference map [1]. The reference map is usually obtained by a ground survey or human labelling. Considering its limited availability, it becomes prohibitive to apply this technique in practice. In this paper, we consider an adaptive strategy to estimate the mutual information using a priori knowledge of the scene. By comparing the spectral-signatures of interest in a signature library, it is not difficult to assume a scope of key spectra, in which the most useful information is located given a specific task. The estimated reference map is then obtained by calculating the mean of the spectral images in the key spectra. Based on the estimated reference map, we calculate the MI of each spectral band to estimate the information content it contains common to the classification objective. The spectral bands are then selected on the basis of their associated MI values. Since this proposed method no longer relies on the given ground reference, the application scope and adaptability is extended.

Experiments are carried out to evaluate the effectiveness of the proposed method using a public dataset AVIRIS 92AV3C. The results show that for the task of four-class vegetation classification, the proposed method can remove 45% of the redundant bands at the expense of a mere 1.34% accuracy loss. If 15% of the total bands are removed, the accuracy increases to about 1.76%.

The remainder of this paper is organised as follows. In Section 2, we discuss the band selection method based on information-theoretic analysis, including the entropy and mutual information analysis. In Section 3, we propose the adaptive MI-based band selection scheme. The experiments on accuracy analysis are carried out to test the performance of the proposed method, which are reported in Section 4. Finally, we end this paper with some conclusions.

2. Hyperspectral band selection through information-theoretic analysis

A. Entropy-based band selection

According to Shannon's information theory, entropy measures information content in terms of uncertainty. Let A be a random variable taking values in the set a with probability distributions $p_A(a)$. The entropy is defined by

$$H(a) = -\sum_a p_A(a) \log p_A(a) \tag{1}$$

Some methods [4, 7] directly use the entropy as a criterion for band selection. In these methods, the entropy is calculated to estimate the level of information contained in each individual band or wavelength interval. The spectral bands are then ranked in a certain order according to their entropy values. Band selection is performed by choosing the bands with the higher entropy values.

The entropy is effective in measuring the information content represented by uncertainty. However, this information measurement does not always, as often assumed, coincide with the requirement of image classification. In fact, the method only is effective when the uncertainty calculated by the entropy encodes the most relevant information for target classification. By examining the definition of entropy, it can be seen that such an assumption does not always hold. From Equation 1, it is shown that the entropy is a function of a single variable A. In other words, $H(a)$ is calculated on a single signal, no reference to an objective. This means that the amount of information measured by the entropy lacks a point of reference or benchmark. Hence, there is no guarantee that the amount of entropy always matches the information content useful for target classification.

B. Priority index represented by mutual information

To improve the entropy-based methods, a logical approach is to extend the information measurement to two variables: one for the measured spectral image itself; the

1 In this paper, the reference map denotes a marked image regarding the scene, i.e., the ground truth map, in which each pixel is correctly labelled to a class.

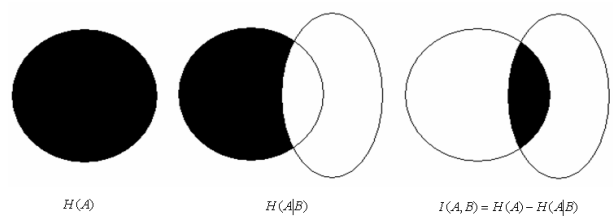

Figure 1. Illustration of Mutual Information.

other for the target image that is directly related to the classification objective. Mutual information (MI) provides an ideal framework to measure the similarity between two random variables, and was introduced for band selection in [12, 5].

MI is a basic concept in information theory to measure the statistical dependence between two random variables [14, 15]. 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:

$$I(A,B) = \sum_{a,b} p_{AB}(a,b) \log \frac{p_{AB}(a,b)}{p_A(a) \cdot p_B(b)} \qquad (2)$$

From Equation 2, it is not difficult to find that MI is related to entropy by the following equations:

$$\begin{aligned} I(A,B) &= H(A) + H(B) - H(A,B) \\ &= H(A) - H(A|B) \\ &= H(B) - H(B|A) \end{aligned} \qquad (3)$$

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 Shannon entropy definition, the joint and conditional entropies can be written as:

$$H(A,B) = -\sum_{a,b} p_{AB}(a,b) \log p_{AB}(a,b) \qquad (4)$$

$$H(A|B) = -\sum_{a,b} p_{AB}(a,b) \log p_{A|B}(a|b) \qquad (5)$$

If we model the spectral images and the corresponding reference map 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 can be seen as a benchmark of the ideal classification result, the MI between a spectral image and its reference map can be used to measure the relative utility of this spectral band to the classification objective (see Figure 1).

Based on the MI criterion stated above, we calculated the mutual information of each spectral image (i.e., each spectral band) in the dataset AVIRIS 92AV3C and the corresponding reference map. This is illustrated in Figure 2 (the solid curve), where the AVIRIS spectral bands are shown along the abscissa from 1 to 220, and their MI values benchmarked with the reference map are plotted along the ordinate axis.

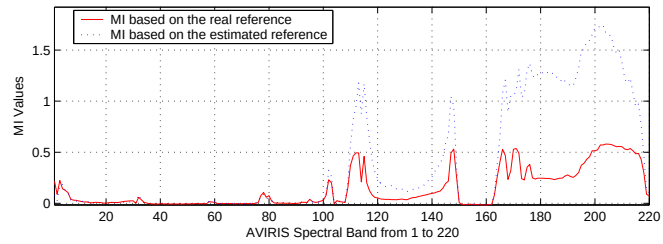

Figure 2. Mutual information of AVIRIS bands 1-220.

The MI curve in Figure 2 shows that the spectral bands 164 to 220 and 110 to 150 have higher MI values than other bands. This reinforces our visual evaluation regarding the utility of each spectral band to classification. Comparing the MI curve in Figure 2 to the selected examples of the AVIRIS hyperspectral images in Figure 3 and Figure 4, it can be seen that the spectral bands more similar to the reference map (see Figure 5) have higher values appeared in the MI curve. For example, the spectral bands from 191 to 200 in Figure 3 appear to be more like the reference map (i.e., contain more pro-classification information). Their MI values are much higher than those of the spectral bands from 103 to 109 in Figure 4. In particular, the MI curve clearly displays the region of water absorption, i.e., the lowest regions in the MI curve (see 104-108, 150-163 in Figure 2). From this comparison, we see that the MI of a spectral band with respect to the reference map is consistent with our visual observations regarding the relative importance of each spectral band to classification. The value of MI can therefore be used to encode the relative utility of a spectral band. For the aim of band selection, we call these MI values the priority index. Thus, the band selection criterion can be roughly presented in the following rule: the higher the priority index is, the higher the probability to retain this spectral band.

One of the significant weaknesses regarding the MI-based method is that it relies heavily on the availability of

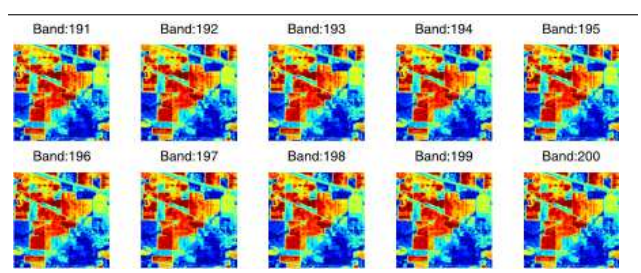

Figure 3. Examples of AVIRIS spectral images, band 191 - 200.

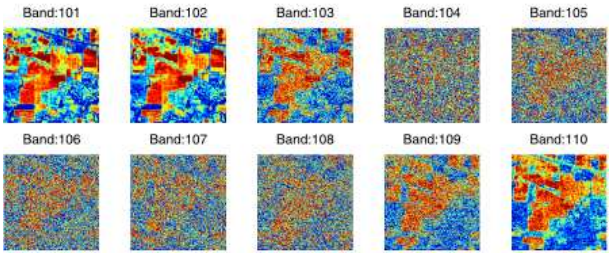

Figure 4. Example of AVIRIS spectral images, band 101 - 110.

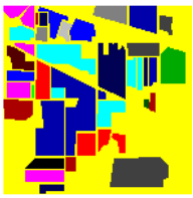

Figure 5. The AVIRIS 92AV3C reference map .

a given reference map. From Equation 2 and Equation 3, it is seen that the calculation of MI must involve a reference variable. Considering the limitation of its availability, it becomes prohibitive to apply this technique in practice. It is possible to construct an approximation to the reference map from the spectral signature library. This is a simple way to extend this method to scenarios where the reference map is not available. However, in many cases, this approximation may be too coarse. To improve the adaptability of this method, we propose a modified MI-based band selection scheme described in Section 3.

3. Adaptive MI-based band selection scheme

Motivated by the reasons discussed above, we consider a revised measurement of the mutual information. In this case, instead of calculating the MI based on the reference map, R, we compute the MI based on an *estimated* reference map, $\widehat{R}$, which is assumed to be easy to obtain. If $\widehat{R}$ is sufficiently similar to the ideal classification result, the MI between a band image and the reference map R can be approximated using the MI benchmarked by the estimated one, $\widehat{R}$.

In our proposed method, the estimated reference map is designed using available *a priori* knowledge. From the spectral signature library, we can analyse the relevant spectral signatures that we are interested in with respect to a specific task. Comparing these signatures, it is not difficult to procure an approximate range of spectra in which the signatures of interest differentiate each other significantly. The region of the spectrum that we shall call key spectra and denote by S, is assumed to contain the most typical discriminatory-information. For example, in the vegetation dominant scene of AVIRIS 92AV3C, the task is to classify vegetation for agriculture survey. By analysing the typical signatures of several classes of vegetation, we make an initial conclusion that the most important spectrum for this task roughly falls into the last 50-70 bands (e.g., from bands 170 to 220, see Figure 3). Based on this prior, we use the average of the bands in the key spectral region S to estimate a reference map.

Let $I_j \in S, 1 \leq j \leq M$ be the set of spectral images that have been assumed as key bands, then the estimated reference map $\widehat{R}$ can be obtained by calculating the algorithm mean of the spectral images in the key bands set, i.e.,

$$\widehat{R} = \frac{1}{M} \sum_{j=1}^{M} I_j \tag{6}$$

Considering that many spectral bands might also be useful and can be seen as potential candidates of key spectra, an alternative scheme to estimate the reference could be an iterative process to search a more accurate estimate. Moreover, some algorithms derived from the analysis of differential entropy are helpful to identify an optimal estimate of the reference, which will not depend on any subjective judgement any more. It is should be noted that with a possible better reference estimate, both of the above schemes involve more bands combinations, and will increase the computational complexity.

To verify the accuracy of this estimation method, we compare the MI results based on the estimated reference map and the real one, i.e., the $MI(I_i, \widehat{R})$ and $MI(I_i, R)$. The dotted curve in Figure 2 represents the mutual information calculated using the estimated reference map on AVIRIS 92AV3C dataset. Compared with the MI result based on the real reference map (the solid curve in Figure 2), it can be seen that the overall shapes of two curves are very similar. The two MI curves have the identical higher value regions and lower value regions. Moreover, the detailed tendencies of rising and falling in these two curves are almost the same. This means that, for the aim of relative comparison, the priority index calculated using the estimated reference map is consistent with that using the real one. Thus, the estimated reference map can be considered as an alternative representation of the ideal classification objective in the calculation of MI.

It is should be noted that the MI value is a relative value calculated based on two images involved. By observing the two MI curves presented in Figure 2, it is seen that although the two curves are quite similar in shape, their absolute values are significantly different since they are based on two different references. Hence, in the practical applications, the threshold for band selection should be chosen based on the correct reference configuration. For example, if a ground truth is available, the threshold should be determined from the solid-curve in Figure 2. On the other hand, if the refer-

ence is estimated, the threshold should be determined from the corresponding dotted-curve in Figure 2.

In contrast to the scheme using the given reference map, the proposed method only requires limited and easily available *a priori* knowledge. Considering that the estimated reference map achieves almost the same MI curve that is based on the real reference, it is possible to substitute the real one in the calculation of priority index. Then, the MI-based band selection can be extended to the applications in which the reference map is difficult to obtain. This scheme also produces an *adaptive* estimation of the reference map. From Equation 6, it can be seen that the estimated reference map only depends on the data itself and would automatically be updated with the new data. In this way, not only is the priority index estimated, but it is also adaptive to the image content as the investigated scene changes.

Comparing with the previous research [4, 5, 7, 12], the significant difference of the proposed method can be summarised as two points: (1) we use MI to measure the utility of each band image based on a target-related reference; and (2) the MI is calculated with respect to an adaptive reference map rather than a given reference map. In terms of the first difference, the proposed method directly estimates the utility of each band to classification rather than the correlation between two band images. This avoids the problem where two spectral bands are highly correlated but are also highly valuable to classification. For the second difference, the proposed method does not rely on the availability of a given reference map, and is therefore suitable for more applications.

4. Experimental results

To assess the proposed band selection method, we performed classification experiments on the hyperspectral dataset, AVIRIS 92AV3C [2].

Kernel methods, such as support vector machines (SVMs) [16, 17], have already shown themselves to provide superior performance in many machine learning applications. Previous literature applying SVMs to hyperspectral data classification [18, 19, 20] have presented competitive performance with the best available classification algorithms. Hence, we choose SVMs as the classifiers in our hyperspectral classification experiments. As a supervised classification method, the SVMs need a training procedure to construct the classifier from examples before applying to those unseen instances. We randomly choose 20% of examples from every class as the training set. The remaining 80% of pixels form the test set.

The main objective of band selection is to remove the redundant spectral bands whilst not degrading the classification accuracy. The experiments are then implemented to assess the change of classification accuracy when certain numbers of spectral bands are removed each time using the proposed band selection method. This analysis is useful to examine the effectiveness of the proposed method on reducing redundancy in the hyperspectral data. In this experiment, the curve of classification accuracy is defined as the change of accuracy with various numbers of bands removed. This curve can indicate the effectiveness of the band selection method. For example, the slope of the curve characterises the capability of band selection to retain the useful information. An ascending or flat curve (i.e., a positive or small value slope) suggests that the accuracy was improved or changed slightly. The rapid descending curve (i.e., a larger negative slope) means that the accuracy degraded significantly. Our main concern is to cut off the redundant bands as much as possible whilst retaining accuracy. Hence, a 'good' band selection method should produce a curve that has a wide range of ascending or flat regions. Meanwhile the band selection should be avoided in the rapid descending regions.

The 92AV3C dataset consists of 220 [3] contiguous spectral bands with wavelengths range of $0.40\mu m$ to $2.52\mu m$. This means that every pixel in this dataset is measured over 220 spectral bands. In this experiment, we gradually cut off a certain number of bands with the lower MI values. Specifically, the number of bands removed is incremented by 5% or 11 bands each time, until 80% bands, i.e., 176 bands from total 220 bands, are cut off. The classification accuracy is tested using the remaining data. Based on these results, we plot an accuracy curve against each percentage of bands cut from 0% to 80%; this is illustrated in Figure 6. In Figure 6, the x-axis represents the percentage of removed bands from 0% to 80%, and the y-axis the classification accuracy tested using the remaining subset of spectral bands. It can be seen that the accuracy curve experiences three phases:

- $s1$: At the beginning of the curve, a slow rise is observed from band cut percentage 0% to 15% (See Figure 6)
- $s2$: In the region where spectral bands are cut off between 15% to 45%, the classification accuracy decreases slightly in an approximately linear manner
- $s3$: In the region where spectral bands are cut off between 45% to 80%, the classification accuracy drops rapidly

2 Dataset AVIRIS 92AV3C can be downloaded from a website at Purdue University (ftp://ftp.ecn.purdue.edu/ biehl/ MultiSpec/).

3 The original band number of the AVIRIS sensor is 224. Four of them are all zeros and are not used. See http://dynamo.ecn.purdue.edu/ biehl/MultiSpec/aviris/_documentation.html

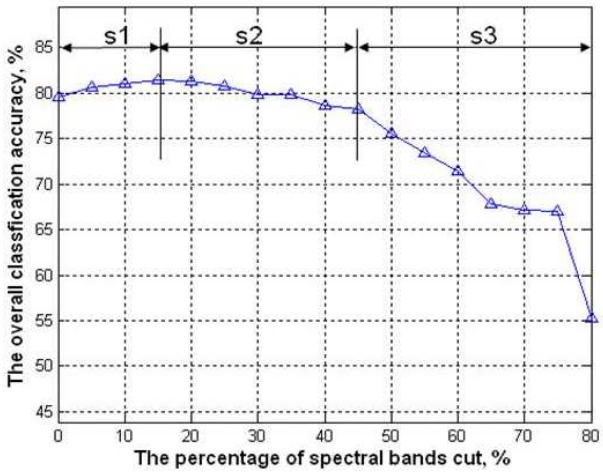

Figure 6. Classification accuracy against various bands cutting percentage.

Number of bands cut off	Number of bands remained	Percentage of bands selected	Change of accuracy
33	187	85%	1.76%
99	121	55%	-1.34%
176	44	20%	-24.38%

Table 1. Change of classification accuracy after the band selection

Figure 6 shows that in a wide range of the x-axis, i.e., from band cutting percentage 0% to 45%, the accuracy curve is ascending or flat. This intuitive observation testified the effectiveness of the proposed band selection method. To quantitatively measure the performance of the band selection, the change of classification accuracy is calculated for each phase. The accuracy difference is measured between the beginning of the curve where no bands are cut off and the end point of each phase. A positive value of accuracy change indicates that the accuracy is improved although some bands are removed. Similarly, a negative value means that the removal of these bands degrades the accuracy. The results are tabulated in Table 1.

From Table 1, it is seen that if we choose 85% top ranked bands (i.e., to cut off the first 15% low MI value bands), the classification accuracy is increased by 1.76%. This phenomenon means that the classification accuracy is actually improved in the beginning of the band selection, even though some bands are removed. It reveals that hyperspectral data contain some bands that are not only redundant or useless, but also detrimental to classification. We call these bands (i.e., the 33 92AV3C bands removed in the rising region $s1$) the negative bands. Cutting off these bands by no means degrades the accuracy but improve it.

In the phase $s2$, if we choose 55% top ranked bands (i.e., to cut off 45% low MI values bands), the accuracy only degrades about 1.34%. This means that in 92AV3C, 30% or 66 bands are less informative or redundant to classification since removing them just causes a slight degradation of classification accuracy ($< 1.5\%$). So we name the bands cut off in the flat region $s2$ as the redundant bands.

If we further cut off the bands from the percentage 45% to 80%, the accuracy curve is seen to drop rapidly from about 78% to 55%. Removing this part of bands, i.e., the bands whose MI values are in the top 55% portion, causes a significant accuracy loss. This means that these bands contain important information to classification, and should be retained in priority. As a result, the bands in the dropping region $s3$ are denoted as the positive bands.

It is noted that in previous research [18, 21], the bands affected atmospheric problems are normally identified by the prior knowledge of experts. For example, [18] indicates that the bands from 104 to 108, 150 to 163 and 220 are needed to discard. From a practical point of view, the method of expert judgement is subjective and impractical in highly changeful scenarios. In this sense, the proposed method provides a better way to discard these useless bands. The proposed method correctly and automatically identified all of these bands. The experimental results illustrated in Figure 2 clearly show that the bands with the lowest MI values coincide exactly with the bands identified by the experts in [18]. Moreover, the proposed method uses an objective metric, mutual information, to measure the level of usefulness of each band. This completely avoids the weakness of subjective judgement. Finally, this method does not adopt a given reference map but an adaptive one, by which the selected 'bad' bands will be updated with different scenes and weather. Hence, for the task of discarding 'bad' bands, the proposed method provides a more reliable and flexible alternative.

Since the bands are selected by the corresponding MI values, the band numbers/ranges for different phases might also be identified approximately by observing the MI curves in Figure 2. For example, the bands with lowest MI values in the region from 150 to 163 will appear in phase $s1$, and the following lower MI bands will be fallen in phase $s2$. Figure 7 illustrates five lowest MI bands in phase $s1$ and five highest MI bands in phase $s3$.

Through the above analysis of Figure 6 and Table 1, it is shown that the proposed method is capable of identifying the noise bands (the phase $s1$), the redundant bands (the phase $s2$), and the useful bands (the phase $s3$). Therefore via the proposed method, the band selection can be successfully implemented by eliminating the noisy and redundant bands whilst preserving the useful bands.

Apparently, any band selection methods can archive the saving on storage space and transmission bandwidth. The extent of saving directly relies on how many spectral bands

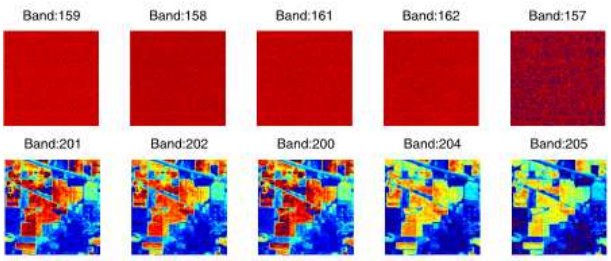

Figure 7. Spectral images extracted from phase $s1$ (the first row) and $s3$ (the second row).

are cut off. The more bands removed, the much saving on storage space and communication bandwidth. The key problem is if the performance of classification accuracy can remain unchanged or within a degradation tolerance. The above results confirmed that the proposed band selection retained a decent accuracy whilst removing the redundant bands.

5. Conclusion

This paper has presented a new band selection method based on mutual information analysis for hyperspectral image classification. While most information-based band selection approaches only use the entropy or the mutual information with a given reference map, the proposed method revises them by devising an adaptive measurement of mutual information. For applications where the ground reference is not available, the scheme can prove useful. As shown by the AVIRIS dataset, the proposed method could effectively reduce the redundant bands with a minor classification accuracy loss (45% bands removal with 1.34% accuracy loss). Whilst a decent classification accuracy is preserved, the processing time, storage space and communication bandwidth are significantly saved using the subset data.

Acknowledgements

This research was supported by the Data Information Fusion Defence Technology Centre, United Kingdom, under 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] AVIRIS. Airborne visible/infrared imaging spectrometer. http://aviris.jpl.nasa.gov/.

[2] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transaction on Information Theory*, 14:55–63, 1968.

[3] G. Petrie, P. Heasler, and T. Warner. Optimal band selection strategies for hyperspectral data sets. In *IEEE International Geoscience and Remote Sensing Symposium Proceedings*, pages 1582–1584, July 1998.

[4] P. Groves and P. Bajcsy. Methodology for hyperspectral band and classification model selection. In *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pages 120–128, 2003.

[5] C-I. Chang, D. Qian, T-L. Sun, and M. Althouse. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2631–2641, 1999.

[6] M. Velez-Reyes and L. Jimenez. Subset selection analysis for the reduction of hyperspectral imagery. In *IEEE International Geoscience and Remote Sensing Symposium Proceedings*, volume 3, pages 1577–1581, 1998.

[7] P. Bajcsy and P. Groves. Methodology for hyperspectral band selection. *Photogrammetric Engineering and Remote Sensing Journal*, 70(7):793–802, July 2004.

[8] J. Price. Spectral band selection for visible-near infrared remote sensing: spectral-spatial resolution tradeoffs. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5):1277–1285, 1997.

[9] N. Keshava. Best bands selection for detection in hyperspectral processing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3149–3152, 2001.

[10] D. Qian. Band selection and its impact on target detection and classification in hyperspectral image analysis. In *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pages 374–377, 2003.

[11] S. Kaewpijit, J.L. Moigne, and T. El-Ghazawi. Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4):863–871, 2003.

[12] C. Conesea and F. Masellia. Selection of optimum bands from tm scenes through mutual information analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 48(3):2–11, 1993.

[13] 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.

[14] N. Blachman. The amount of information that y gives about x. *IEEE Transactions on Information Theory*, 14(1):27–31, 1968.

[15] 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.

[16] C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2(2):121–167, 1998.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.