© EARSeL and Warsaw University, Warsaw 2005. *Proceedings of 4th EARSeL Workshop on Imaging Spectroscopy. New quality in environmental studies*.
Zagajewski B., Sobczak M., Wrzesień M., (eds)

- IMPROVING CLASS SEPARABILITY - A COMPARATIVE STUDY OF TRANSFORMATION METHODS FOR THE HYPERSPECTRAL FEATURE SPACE

Mathias Bochow

GeoForschungsZentrum Potsdam, Section 1.4 Remote Sensing, Potsdam, Germany; mbochow@gfz-potsdam.de

ABSTRACT

Principle and practical aspects of three transformation methods for the hyperspectral feature space (MNF, DAFE, DBFE) are described. In two application cases (urban case: selected man-made materials, water and shadow; rural case: tree species) these methods are used for the discrimination of spectrally similar classes. Supervised classifications are conducted in the original feature space as well as in each of the transformed feature spaces. The achieved separabilities of the selected classes are compared and discussed with respect to the different transformation methods. The analysis is made on 2003 HyMap data (3 m GSD) for an area in and close to the city of Osnabrück in Lower Saxony.

INTRODUCTION

Hyperspectral sensors measure the surface reflectance of electromagnetic irradiation using a huge number of narrow adjacent bands. Due to the small spectral distance between two adjacent bands, they are often highly correlated, so that for a specific classification problem usually not all bands are needed. In fact, having to many bands can lead to worse classification results caused by two reasons: First, if there were some “good” bands that contain information for the separation of the given classes and some others that don’t and to all bands the same weight was given in classification, then the other bands would degrade the classification accuracy. To solve this problem, feature selection methods (e.g. Bhattacharyya distance, transformed divergence (i)) were developed to enable the selection of n bands which, for each number n, would result in the best classification accuracy. Figure 1 shows the achieved classification accuracy for the urban study case depending on the number of bands which were selected with Bhattacharyya distance. In this figure the second reason for reducing the number of bands before classification can be seen. For a limited number of training samples adding more bands will first increase the classification accuracy, raise to a maximum and then decrease again. This decrease is due to an inaccurate estimation of class statistics and is known as the Hughes phenomenon.

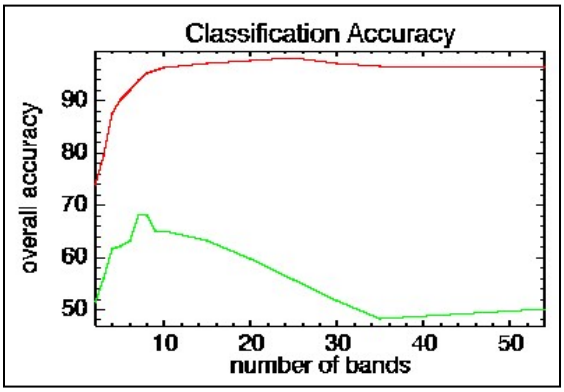


Figure 1: Classification accuracy for the classification of training (red line) and test samples (green line)

Moreover, in some cases the original feature space is suboptimal for separating spectrally similar classes. Consequently, transformation methods were developed that generate a smaller set of new bands by transforming the hyperspectral feature space with the aim of improving the separability of given classes. The methods have different principles, assets and drawbacks. Therefore, in this study three selected transformation methods – the Minimum Noise Fraction Transformation (MNF) (ii), the Discriminant Analysis Feature Extraction (DAFE) (i) and the Decision Boundary Feature Extraction (DBFE) (iii) – are compared and examined for their ability to aid discrimination between spectrally similar classes in two application cases.

DATA AND TEST SITES

The analysis was done for an urban and a rural test site in and close to the city of Osnabrück (Lower Saxony, Germany). The urban analysis was done on a hyperspectral dataset of DAIS 7915 (iv) with a ground resolution of 5 m, collected by the DLR in July 2002. The rural analysis was done on hyperspectral HyMap (v) data with a ground resolution of 3 m, collected by the DLR in July 2003. The DAIS data were processed by the DLR including rectification and radiometric calibration and were delivered as radiance data. Further pre-processing included atmospheric correction with FLAASH and the elimination of noisy bands. The HyMap data was delivered as radiometric calibrated radiance data. Rectification was done with PARGE (vi), followed also by an atmospheric correction with FLAASH and the elimination of noisy bands. For the determination of noisy bands their signal-to-noise ratio (SNR) was calculated utilizing the “homogeneous area method” (e.g. vii). This method implies that the image data contains an area which is homogeneous for the sensor. Thus, the variance in this area is due to the noise. For the SNR calculation the overall mean of each band was taken for the magnitude of the signal and the magnitude of noise was approximately calculated by the standard deviation of a water area assumed to be homogeneous. This was done on radiance data. Figure 2 shows the SNR quality of both datasets which show a significant difference in the SWIR.

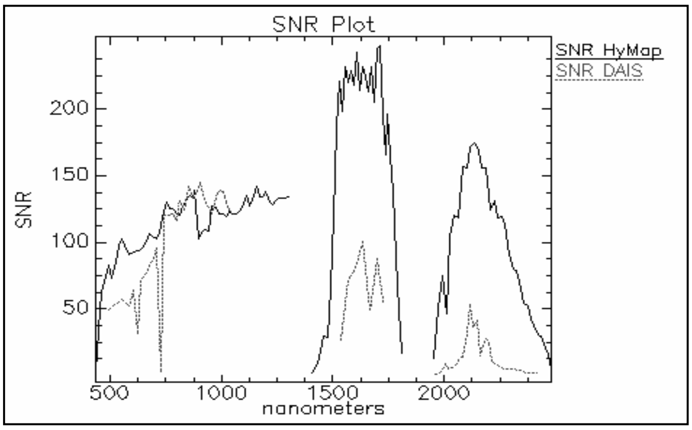


Figure 2: Comparison of the Signal-to-Noise-Ratios (SNR) of DAIS 7915 and HyMap

For both analyses – the urban and rural one – the focus of attention lies on the differentiation of selected spectrally similar classes. Training fields for the analyses were build from larger homogeneous regions – if available – otherwise from big single trees or single roofs, respectively, and were validated by high resolution image data and by ground truthing. The classes’ spectra (mean spectra of training fields) are shown in figures 3 and 4. The urban classes are all characterized by a low albedo and by a lack of strong absorption bands. The spectra of the tree species are characterized by a very similar shape.

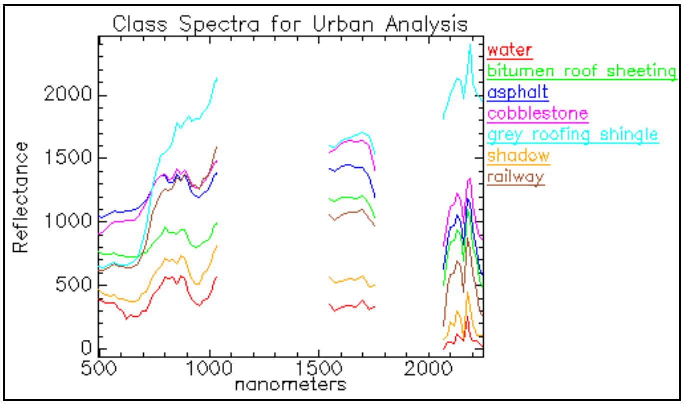


Figure 3: Urban analysis: Class spectra of selected man-made materials, water and shadow

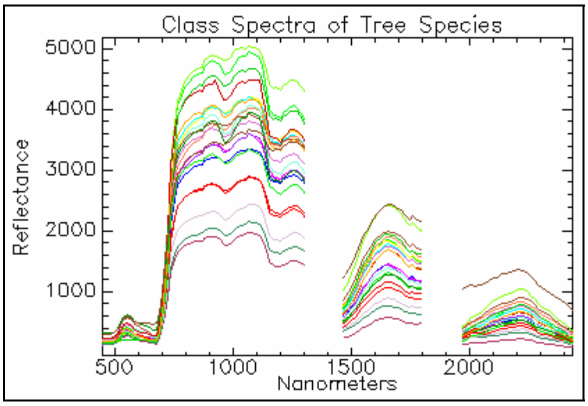


Figure 4: Rural analysis: Class spectra of 20 tree species.

Table 1: List of tree species to discriminate in the rural analysis

Scientific name	**English name**	**German name**	**Used abbreviation**
Abies alba	European Silver Fir	Tanne	Ta
Acer platanoides	Norway Maple	Blut-Ahorn	BlAh
Aesculus hippocastanum	Horse Chestnut	Rosskastanie	Ka
Alnus glutinosa	Black Alder	Schwarz-Erle	Er
Betula pubescens	Downy Birch	Birke	Bi
Fagus sylvatica	European Beech	Rotbuche	Bu
Fagus sylvatica f. purpurea	Purple European Beech	Blut-Buche	BlBu
Fraxinus excelsior	Common Ash	Esche	Es
Larix decidua	European Larch	Lärche	Lä
Malus spec.	Apple tree	Apfelbaum	Apf
Picea abies	Norway Spruce	Fichte	Fi
Picea pungens "Glauca"	Colorado Blue Spruce	Blaufichte	BlauFi
Pinus sylvestris	Scotch Pine	Wald-Kiefer	Ki
Populus tremula	European Aspen	Zitter-Pappel	Pz
Populus x canadensis	Carolina Poplar	Hybridpappel	Ph
Prunus avium	Sweet Cherry	Vogel-Kirsche	Kv
Prunus spinosa	Blackthorn	Schlehe	Sl
Quercus petraea, Q. robur	Sessile Oak, Common	Trauben-, Stileiche	Ei
Salix alba	White Willow	Weide	We
Tilia cordata	Littleleaf Linden	Winter-Linde	Li

METHODS

The *Minimum Noise Fraction Transformation* (MNF; also called *Maximum Noise Fraction Transformation*) (ii) is commonly used to determine the inherent dimensionality of image data, to segregate noise in the data, and to reduce the computational requirements for subsequent processing (viii). The transformation consists of two cascaded *Principal Component Transformations* (PCA) and is based on the global covariance matrix and on an estimated noise covariance matrix. The first step transforms the noise covariance matrix into an identity matrix which decorrelates the noise in the data. The second PCA finds new components with decreasing variance of signal. The segregation of noise usually works better when a dark current file (which is like a photograph leaving the cap on the lens) is used for the calculation of the noise covariance matrix. Otherwise it can be derived from the image data itself using a shift-difference method(viii).

Compared to the MNF the *Discriminant Analysis Feature Extraction (*DAFE; also called *Canonical Discriminant Analysis* (CDA)) (i) is based on the within-class and between-class covariance matrix instead of the global covariance matrix. We can call it a class-dependent transformation resulting in different new features for different input classes. Therefore it is expected to be more appropriate for a specific classification problem like discriminating spectrally similar classes. The DAFE transformation finds new discriminant features by a linear combination of the original ones, so that the ratio

$$\frac{\sigma_B^2}{\sigma_w^2} = \frac{\textbf{between-class variance}}{\textbf{within-class variance}} \tag{1}$$

is maximized. Here σ_B^2 is the variance of the mean values of the classes and σ_w^2 is the average of the variances of all individual classes when projecting them onto the new feature axis. In a two dimensional, two class example the new feature axis calculated by the DAFE algorithm is shown in figure 5. If we now consider those two classes to be spectrally very similar, the whole point cloud of the dataset could look like in figure 6. This shows that the new components found by PCA or MNF whose calculation is based on the global covariance matrix are not sensible to a separation problem of spectrally similar classes.

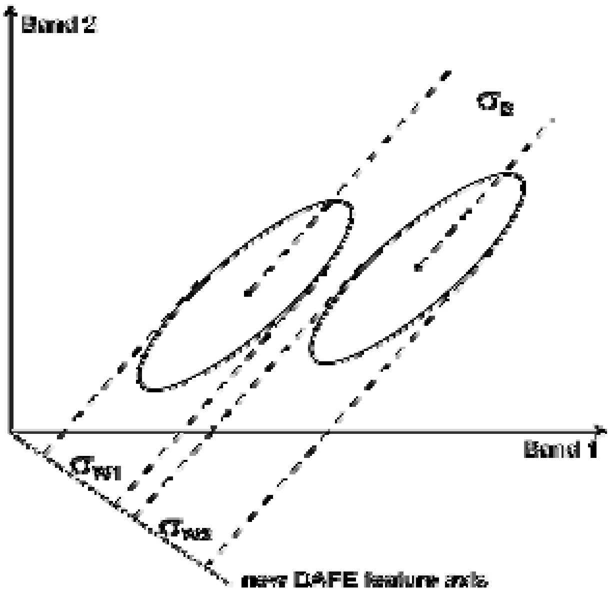


Figure 5: Illustration of the principle of DAFE. The new axis minimises σ^2_w while σ^2_B is maximised. (from (i); modified)

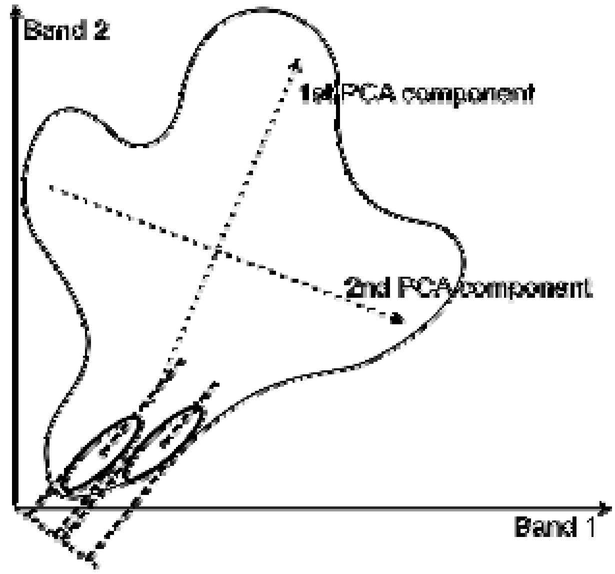


Figure 6: Possible position of the two classes within the entire point cloud. Axis orientation of the DAFE axis compared to the axis found by PCA or MNF

As DAFE the *Decision Boundary Feature Extraction* (DBFE) is also a class-dependent linear transformation which predicts the minimum number of features needed to achieve the same classification accuracy as could be obtained in the original space for a given classification problem and finds the needed feature vectors (iii). The new feature vectors are calculated based on the decision boundaries between the given classes. For a two class problem the decision boundary is the locus of points where the probability of a pixel vector to belong to class A or class B is equal (fig. 7). Now the new feature axis can be found to be normal to this decision boundary. Depending on dimensionality of the feature space a decision boundary can be a point, line, plane, hyper-plane, solid, hyper-solid, curved surface or curved hyper-surface.

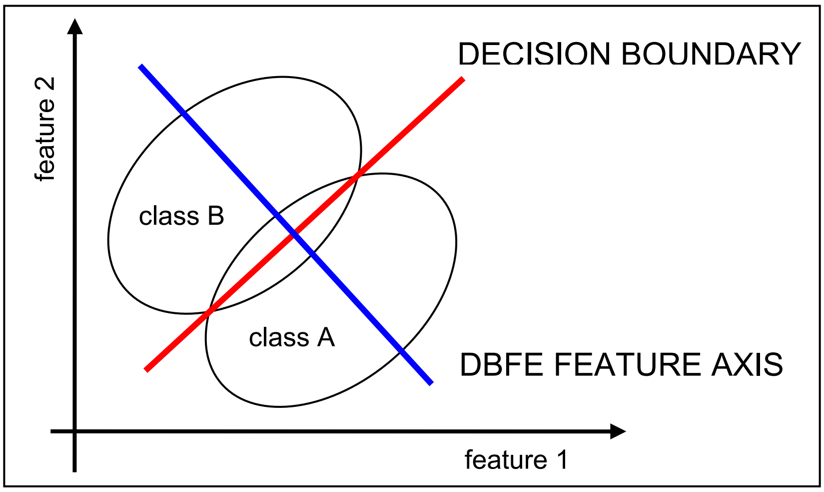


Figure 7: Decision Boundary (red line) for a two-class problem. (from (iii); modified)

A practical aspect to be mentioned about DAFE is that the number of features DAFE calculates for a classification problem with n classes is limited to n-1. Thus DAFE produces good results when less than n features are sufficient to discriminate classes. Otherwise DBFE, in general, gives better results (iii).

Although DBFE works for multi-class problems, too, it is optimized for fewer numbers of classes where ≥ n features are needed for separation. With a growing number of classes the quality of DBFE results may decrease. A limitation of the algorithm is that the number of pixels of each classes' training field has to be at least one more than the number of features of the input dataset.

Table 2: Degree of fitting of classifiers and transformation methods to the class statistics.

Methods fits to …	Nothing. No fit to the given classes.	Class means	Means and variances, averaged for all classes	Means and variances, individually for every class
SAM		X		
MinDist		X		
ML				X
MNF	X			
DAFE			X	
DBFE				X

RESULTS

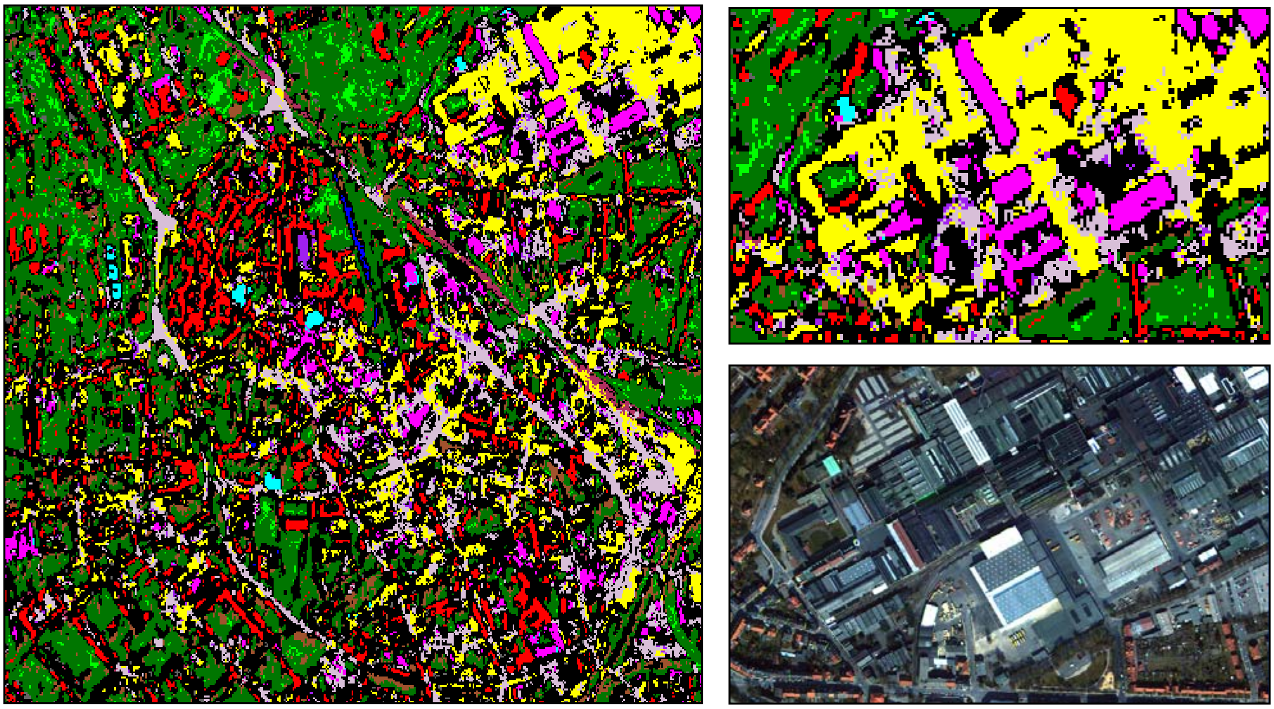
For both study cases the three transformation methods were applied to the image data. Classification was done on the transformed datasets and on the original image data using Maximum Likelihood (ML) and Minimum Distance (MinDist) classification and the Spectral Angle Mapper (SAM). For each transformation and classifier combination the overall accuracy for the classification of training and test fields is given in tables 5 and 6. Single input spectra for the SAM classification were calculated by averaging the classes' training field spectra. The number ground truth pixels for the rural site varies from less than 10 for a few classes to several hundreds. Enough validated pixels to divide them into independent training and test fields could be collected only for the classes European Beech, Common Oak, Norway Spruce and an apple tree class. Thus, the confusion matrices were calculated only with these four classes.

Table 5: Overall accuracies (training/test) for each transformation and classifier combination for the urban site.

	Orig. FS (54 bands)	MNF (17 bands)	DBFE (20 bands)	DAFE (6 bands)
ML	96.3 / 50.1	97.6 / 61.9	98.3 / 62.1	95.6 / 78.1
MinDist	55.9 / 52.5	54.3 / 54.7	54.2 / 51.6	87.7 / 77.4
SAM	68.3 / 58.9	52.7 / 48.3	56.3 / 54.5	84.3 / 75.0

In the urban application DAIS data was classified with the aim to distinguish the spectrally similar classes asphalt, bitumen roof sheeting, cobblestone, grey roofing shingle, railway, water and shadow. In the original feature space high confusion between the classes appears with a low overall accuracy for the test fields for all used classifiers. The big difference between the first and second value for the ML classification shows that ML suffers more from the Hughes phenomenon than the other classifiers because it fits more to the estimated class distributions. The MNF and DBFE transformations are not able to improve results significantly. In contrast, DAFE highly improves the accuracy for training and test fields while the dimensionality is reduced to 6 bands.

To obtain a completely classified image the classification result from DAFE/ML was combined with a SAM classification on the MNF transformed feature space which included all other surface materials / land covers of the image (e.g. several roof types, wood, meadows). The subset shown in figure 8 (upper right edge of the full image) shows an industrial area where most buildings are covered with bitumen roof sheeting (yellow). It shows the ability of the algorithm to distinguish between bitumen roof sheeting and the asphalted streets and parking areas around the buildings. Figure 9a shows the position of the class spectra vectors in the point cloud of a 3D-subspace of the original feature space of DAIS data and thus indicates their spectral similarity.



bitumen roof sheeting, asphalt, galvanized steel panel, red roof shingle, hardwood

Figure 8: Combined classification results from MNF- and DBFE-transformed datasets. The subset of the upper right edge shows an industrial area with mostly bitumen-roof-sheeting-covered buildings and some asphalted streets and parking areas around.

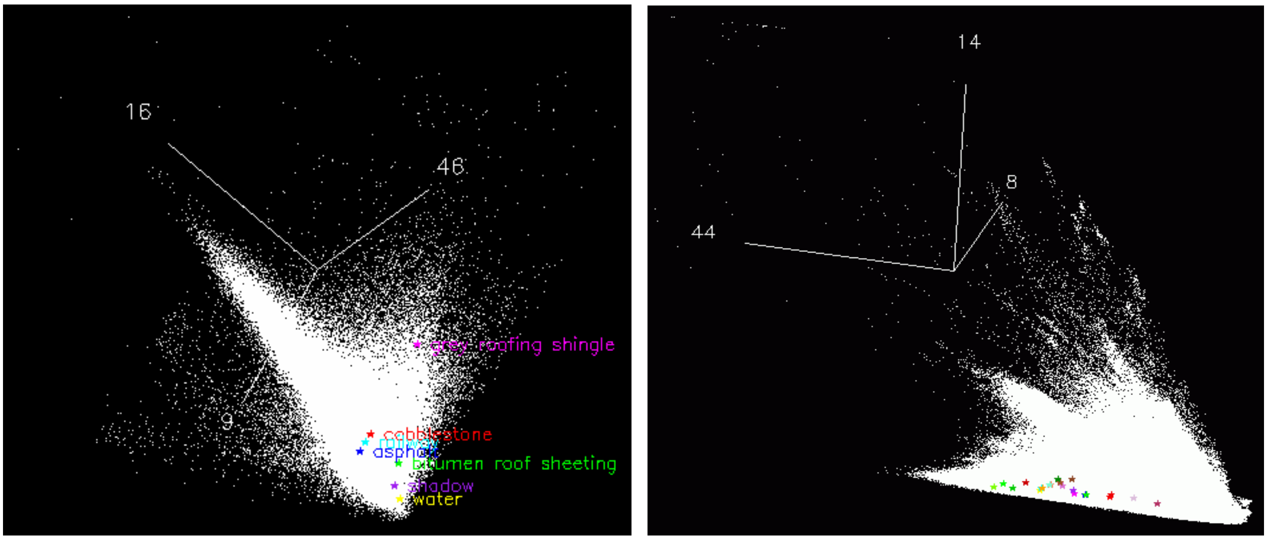


Figure 9a + 9b: Urban (a) and rural (b) class spectra vectors in the original feature spaces of DAIS and HyMap data.

In the rural analysis the aim was to distinguish 20 tree species (tab. 1) in the hyperspectral HyMap data. Figure 9b shows the position of the tree species' class spectra in the point cloud of a 3D-subspace of the original feature space (displayed bands: CIR). They spread along a bulge that points in the direction of the infrared axis but compared to the whole extend of the point cloud the area covered by the vectors of the class spectra is rather little keeping in mind that these vectors don't belong to one class but each to an individual class. The small differences between the vectors of the class spectra in the original feature space are hard to detect for common used classifiers. This can be seen in figure 10a which shows a plot of calculated spectral angles in the original feature space for three selected pixels. These pixels belong to three different tree species as indicated in the legend (Bu = European Beech; Ei = Oak; Ph = Carolina Poplar) and should have the lowest spectral angle at the corresponding class number (indicated by the arrows). This is not the case for Oak and Carolina Poplar, and although it is the case for European Beech, the difference between the lowest and the other values is very small. Consequently, the classification results for the original and also for the MNF-transformed dataset show a very high confusion between almost all classes.

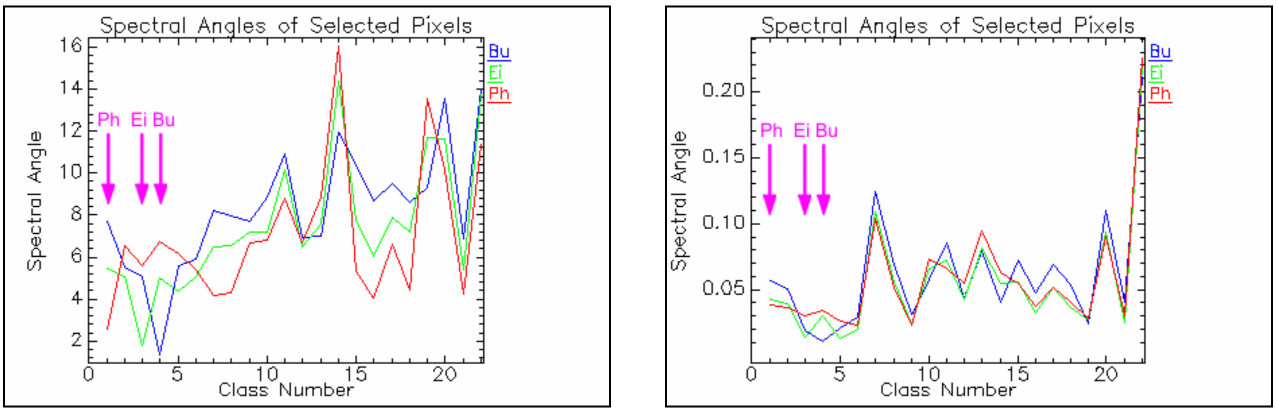


Figure 10a + b: Plot of calculated spectral angles for three selected pixels in the original (a) and DAFE-transformed (b) feature space. Arrows indicate the class numbers of the three tree species where the lowest spectral angle should appear for correct classification.

On the contrary, a DAFE transformation is highly conductive to separation of the tree species as shown in figure 10b. Here, the spectral angles after a DAFE transformation are plotted for *the same* pixels shown in figure 10a. Now all pixels would be classified correctly and the distance between the right and the wrong classes has increased drastically. This improvement can be seen also in the classification result where most of the classes appear to be separable (fig. 11) and in the increased overall accuracy for the test fields in table 6. Here also MNF and DBFE show improved accuracies for the test fields. As mentioned above the accuracies for all 20 tree species could not be calculated because of low numbers of validated ground truth pixels. But it can be seen in the classified images that the MNF and DBFE transformation does not improve separability when all tree species are involved. For the DBFE this can be explained with the method's problem with high numbers of input classes.

Table 6: Overall accuracies (training/test) for each transformation and classifier combination for the rural site. Transformations and calculated values are only based on the classes European Beech, Common Oak, Norway Spruce and an apple tree class.

	Orig. FS (116 bands)	MNF (21 bands)	DBFE (6 bands)	DAFE (3 bands)
ML	100 / 84.0	100 / 97.3	100 / 93.7	100 / 98.1
MinDist	83.5 / 89.0	83.4 / 89.2	99.6 / 96.9	100 / 98.3
SAM	91.9 / 71.9	82.7 / 89.9	99.5 / 96.7	100 / 96.3

Figure 11 shows the classification result for the combination DAFE/MinDist. To optically improve the image it was filtered after classification with a moderate majority filter which substitutes the center pixel of a three by three kernel with the most frequent grey value (= class value in this case) of the kernel if and only if the grey value of the kernel center does not appear in it's N8 neighborhood.

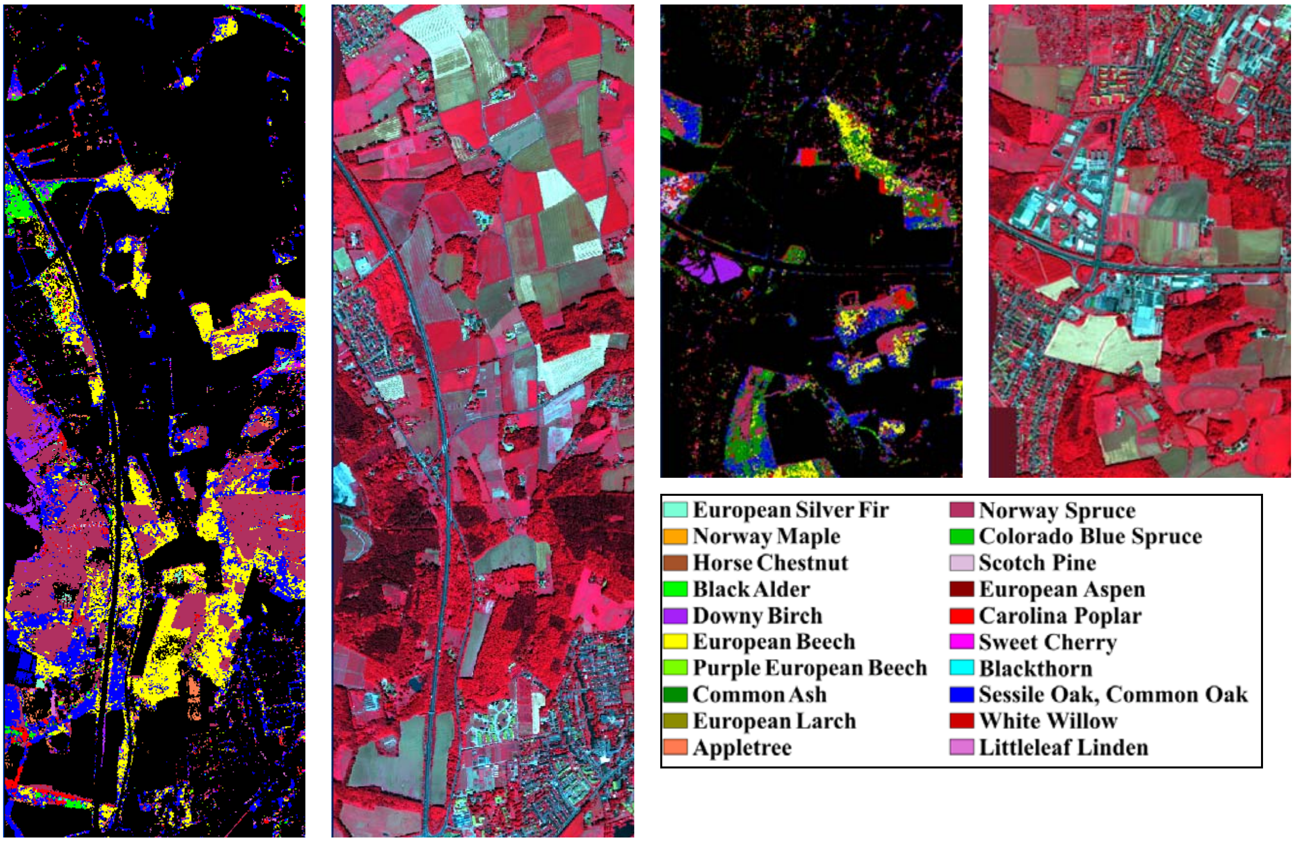


Figure 11: Classification result for the tree species classification. Left is in the north, right in the south of Osnabrück.

CONCLUSIONS

The MNF improves classification results when all or the main classes covering the image scene are included. But because it is based on the global covariance matrix there was no improvement for the discrimination of spectrally similar classes. Here DAFE and DBFE showed a great improvement while DAFE should be applied if the number of classes is high and the number of features needed to distinguish is less than the class number. On the other hand DBFE should be chosen if the number of classes is low and the number of features needed to distinguish is greater than the class number.

The ongoing development in sensor technology will produce sensors which will be able to record more and more slightly differences between observed objects. This research has shown that the methods to extract them from the data are already existing.

ACKNOWLEDGEMENTS

I like to thank Ansgar Greiwe for the rectification of the HyMap data with PARGE. Thanks go also to Ronald Janowsky for helping with mapping of training fields and to Karl Segl for some programming.

References

i Richards, J. A., Jia, X., 1999. Remote Sensing Digital Images Analysis – An Introduction. 3rd rev. and enlarged ed., Springer, Berlin, pp. 249–251.

ii Green, A., Berman, M., Switzer, P. & M. Craig, 1988. A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal. In: IEEE Transactions on Geoscience and Remote Sensing 26(1): 65–74.

iii Lee, C., Landgrebe, D. A., 1993: Feature Extraction Based on Decision Boundaries. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15(4): 388–400.

iv DLR: http://www.dlr.de/caf/mf-ev/Infrastruktur/Hardware/DAIS/dais_en.html

v HyVista Corp.: http://www.hyvista.com

vi Schläpfer, D., Schaepman, M. E., and Itten K. I., 1998. PARGE: Parametric Geocoding Based on GCP-Calibrated Auxiliary Data. In: Proc. of SPIE Int. Symp. on Opt. Sc., Eng. and Instr., San Diego (CA), pp 334-344.

vii van der Meer, F. D., de Jong, S. M., 2001. Imaging Spectrometry - Basic Principles and Prospective Applications. (Remote Sensing and Digital Image Processing, Vol. 4), Kluwer Academic Publishers, Dordrecht, p. 36.

viii Reseach System Inc., 2003. ENVI User's Guide. September 2003 Ed.