

PRIP-TR-93

November 19, 2004

Classification of Color Pigments in Hyperspectral Images

Christian Asinger

Abstract

Image spectroscopy in ultraviolet, visible and infrared regions provides a non-invasive method for analyzing paintings. It is used by restorers and art-historians to get valuable information about works of art without causing any damage to them. The purpose of this report is to analyze 15 different color pigments and three types of binder by means of near infrared hyperspectral images taken from 15 testpanels. After calibration three feature reduction methods, *principal component analysis*, *block-based principal component analysis* and *linear discriminant analysis*, are applied to the near-IR spectroscopic imaging data of the testpanels to reduce the originally 180 features to improve classification performance. Both, *principal component analysis* and *linear discriminant analysis*, are linear transformations (in contrast to e.g. Kernel PCA) that map the original data into a new vector space where certain constraints are met. On the basis of the outcomes of the feature reduction step *k-nearest neighbor* classification with *leave-one-out cross validation* is used to accomplish three different classification tasks. These three tasks contain the identification of the three binders, the identification of the color pigments for each binder separately and the identification of color pigments and binders together. The results of our work reveal that it is possible to distinguish the three types of binder as well as the color pigments for each binder separately with an accuracy of more than 90 percent. Classifying the 15 color pigments along with the three binders however results in a lower accuracy of at most 45 percent.

Contents

1	Introduction	2
2	Experiment	3
2.1	Testpanels, Data and Goals	5
2.2	NIR - Spectral Imaging System	7
3	Feature Reduction	7
3.1	Principal Component Analysis	8
3.2	Linear Discriminant Analysis	11
3.3	Block-Based PCA	12
3.4	Feature Reduction Results	14
3.4.1	PCA	14
3.4.2	Block-Based PCA	17
3.4.3	LDA	23
4	Classification	26
4.1	K-Nearest Neighbor Classification	26
4.2	Distance Measures	27
5	Results	27
6	Conclusion	31

1 Introduction

For restorers and art-historians it is of vital importance to uncover invisible information contained in artworks like the combination of materials used or the working technique followed by the artist [3]. For the restoration of a damaged part of a painting for instance it is necessary to determine the composition of color pigments in adjacent areas in order to guarantee that the new part does not differ from its surrounding. Another case where one needs information of a painting, which is a priori not visible to the human observer is if the underdrawing of a painting must be assigned to an artist [3].

The process of revealing hidden data in works of art requires an interdisciplinary and systematic approach, which includes invasive analytical techniques for the physical, chemical and structural characterization of the artwork. Analytical techniques are not applicable to in situ analysis and require extraction of samples, which is an invasive process and therefore unavoidably leads to permanent damage to the work of art [3]. However, the fact that every object of art is a unique piece emphasizes the necessity of working with non-invasive methods [10].

Popular non-destructive methods include the use of X-ray radiation (radiography), ultraviolet radiation (UV fluorescence) and near infrared radiation (IR reflectography). They permit the gathering of spectroscopic information without any sampling and therefore overcome the drawbacks of invasive analytical techniques. Reflectance spectroscopy provides the possibility of identifying pigments, monitoring the presence of alteration products and color changes on paintings. It also enables the uncovering of underdrawings due to the fact that pigment layers absorb much less infrared radiation compared to visible radiation [11].

James R. Mansfield et al. [9] use near infrared spectroscopic imaging data to analyze the remnants of a 16th century drawing damaged during a cleaning attempt. During the conservation attempt all the ink in the drawing was removed. Using hyperspectral images in the range from 650nm to 1050nm collected from the artwork and *linear discriminant analysis*, Mansfield et al. are able to identify regions where small amounts of the ink or its decomposition products are left.

Berns et al. [12] propose a method that uses visible spectrum imaging data to evaluate potential pigment combinations in paintings. The additional use of visible spectrum imaging data could be a complement to our work if we decided to identify the composition of mixed color pigments too. In the experiment this report is about, however, we only go for the identification of pure color pigments.

In detail, this report deals with the analysis of reflectance values of 15 different color pigments and three types of binder on the basis of near infrared hyperspectral imaging data. The goal of our work is to find out whether it is possible to distinguish between the different pigments and binders by means of their reflectance values in the range from 900nm to 1700nm. The imaging data were obtained from 15 testpanels which are structured like real paintings with a grounding layer, an underdrawing and a single pigment layer. Each pigment layer is a mixture of one color pigment and one of the three binders. Since the color pigments themselves are powders, the binders are necessary to make the pigments solvent for application. Because the three binders influence the transparency of the color pigments differently, in Figures 1, 2 and 3 the calibrated and filtered mean

reflectance values of the 15 color pigments are depicted separately for each binder. In Figure 1 pigments are mixed with binder *Scotch Glue* (Hautleim), in Figure 2 with binder *Linseed Oil* (Leinoel), and in Figure 3 the color pigments are mixed with binder *Egg Tempera* (Ei).

Our experiment consists of two main steps that lead us to the results of identification. The spectral distributions in Figures 1, 2 and 3 show that there is high correlation between neighboring bands in the original data what implies that not all of the 180 original bands (features) are needed for discrimination. Therefore the first step is a feature reduction step where three different approaches are tested. The first technique used is *principal component analysis* (PCA), because it is a classical technique in multispectral image analysis and because it maps the original bands into a new vector space where the new features are uncorrelated. The second approach, block-based PCA, was chosen because of the block-correlation nature of our data, as can be seen in Figure 7 in Section 3.3. The third method, *linear discriminant analysis* (LDA), is used as contrast to the two principal component transformations. Unlike PCA and block-based PCA, LDA maps the original data into a vector space where the separability between classes is maximized.

In the second main step the *k-nearest neighbor* (KNN) classification algorithm is applied both, to the original data set and to the outputs of the three feature reduction methods. KNN was chosen because it is a widespread algorithm for simple classification problems that is easy to implement and does not make use of class statistics like mean values or covariance matrices. In addition, an early research into our test data showed that unsupervised clustering algorithms like *K-means clustering* are unsuitable because within-class distances are too high compared to between-class distances, i.e. the classes do not form distinct clusters in feature space (e.g. Section 3.4.1, Figure 10). The results of our experiment show that it is basically feasible to distinguish the 15 color pigments and the three binders with the experimental setup and methods described in this report. From the three feature reduction algorithms LDA performed best in four out of five cases.

The next section explains the experimental setup and the goals of our experiment in detail. In Section 3 the three feature reduction methods are discussed and the results for each method are presented. Section 4 is concerned with the *k-nearest neighbor* classification technique and the distance measures used in it. Finally Section 5 provides the results of the classification step, which are discussed afterwards in Section 6.

2 Experiment

The first point in this section is concerned with the condition of the testpanels as well as the structure of the hyperspectral images, which are the basis of our work. Furthermore the purpose of the experiment which can be broken down into three tasks is outlined. The second point, this section deals with, is a listing of the technical specifications of the imaging system used for gathering the near-IR spectroscopic imaging data.

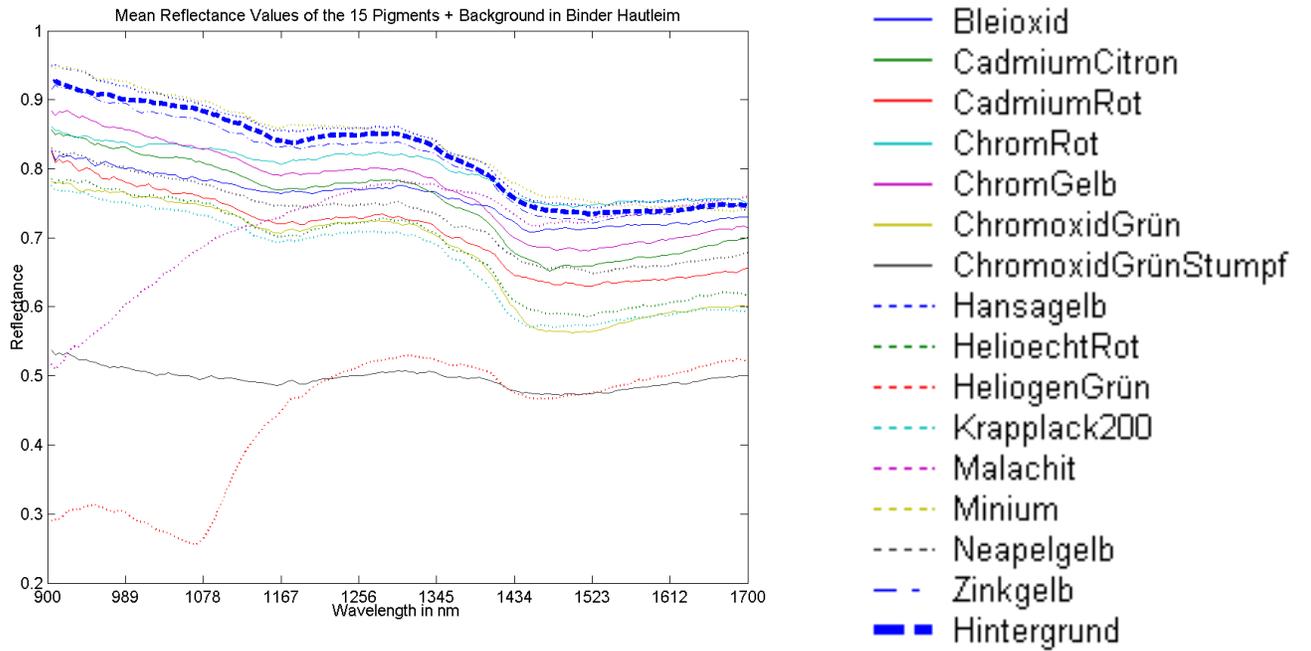


Figure 1: Mean Value of the Pigments Spectral Distribution for Binder Scotch Glue

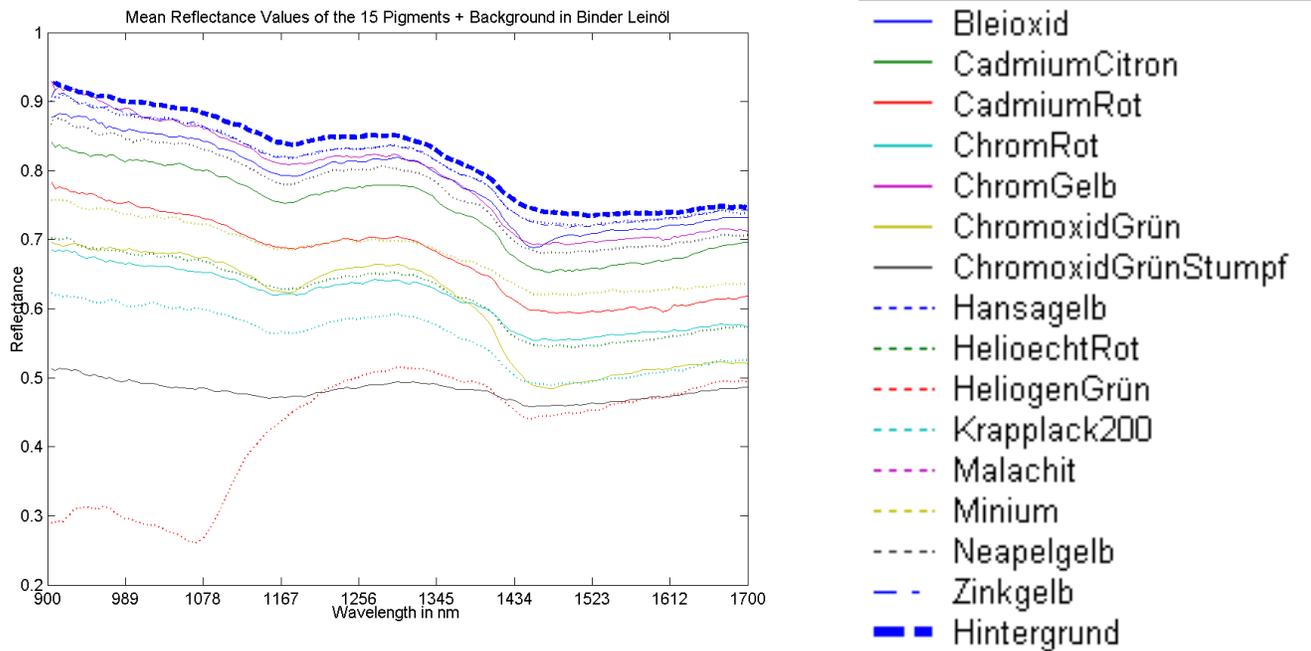


Figure 2: Mean Value of the Pigments Spectral Distribution for Binder Linseed Oil

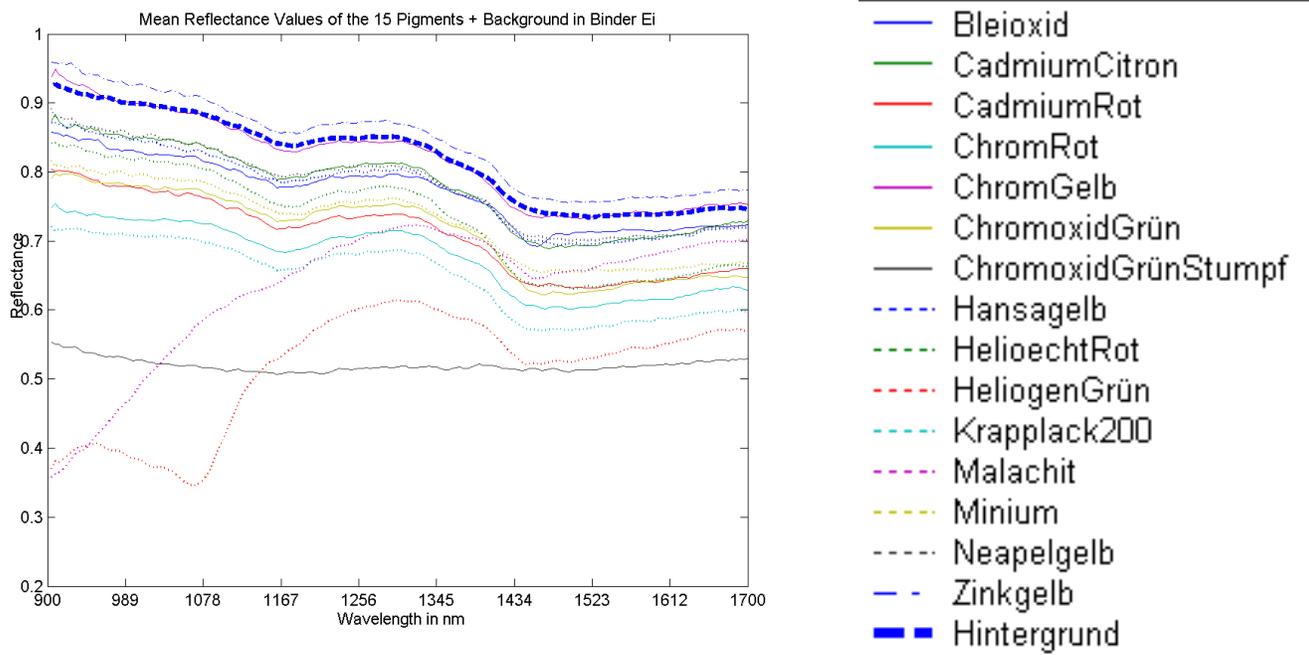


Figure 3: Mean Value of the Pigments Spectral Distribution for Binder Egg

2.1 Testpanels, Data and Goals

The basis of our work is composed of hyperspectral images of different color pigments and binders taken from 15 testpanels. The length of one panel is about 200mm at a width of about 90mm and on each of them one specific color pigment is applied to three rectangular areas whereas in each area the color pigment is mixed with one of the three binders *Scotch Glue*, *Lineseed Oil* and *Egg Tempera*. Figure 4 shows the testpanels for Chrome Red and Chromium Oxide Green as an example. Therein the three separate areas, which differ in the binder used are readily identifiable.

Every 12 bit hyperspectral image file is given in ASCII format and contains one line of a testpanel with a spatial resolution of about 40 pixels recorded in 180 different wavelengths from 900nm to 1700nm. Altogether there are 15 panels corresponding to 15 different color pigments. Our first approach was not to distinguish between the three types of binder but only between the 15 color pigments, which would imply that there are 15 classes with about 25 to 30 pixels per class and 180 features. Unfortunately after our first research into the spectral curves of the pigments it became evident that the varieties between the types of binder within one and the same class of color were too big to neglect them. Finally we went for three classification tasks with different setups.

- Task 1: The first setup consists of the image pixels from all panels according to a color pigment put together into one single matrix. The first classification goal is to correctly identify the 44 classes composed of 15 color pigments and three binders minus one because the panel of pigment *Malachite* lacks of binder *Lineseed Oil* since *Malachite* can hardly be mixed with *Lineseed Oil*.

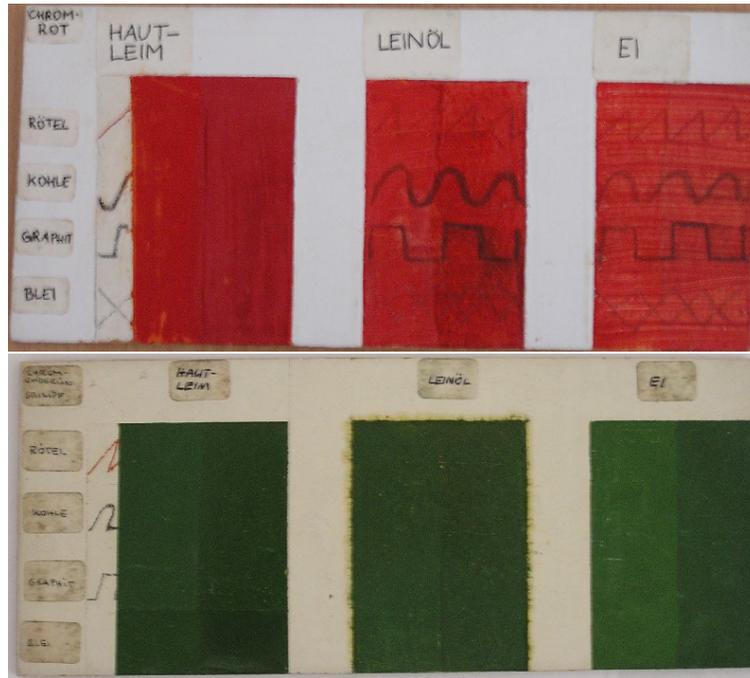


Figure 4: 200x90mm Testpanels for Chrome Red and Chromium Oxide Green

- Task 2: The second setup equals the first but this time the aim of the classification step is to distinguish between the three classes of binder.
- Task 3: For the last setup the matrix from one and two is split up into three smaller matrices, where each one only contains image pixels according to one specific binder. I.e. the first matrix defines 15 classes composed of the 15 color pigments mixed with binder *Scotch Glue* and so on. This time we want to identify the different color pigments (15 within *Scotch Glue* and *Egg Tempera* and 14 within binder *Linseed Oil*) separately for each binder.

So in fact the number of pixels per class averages out at approximately ten for tasks one and three and at approximately 133 for task two.

All algorithms were implemented in MATLAB, the images were stored in two-dimensional matrices where each row corresponds to the line from the testpanel recorded in a specific wavelength. Each column represents a pixel vector consisting of the 180 reflectance values recorded per image point.

In addition to the 15 images of testpanels a whiteboard and a blackboard reference image were provided and used for calibration purposes in the first step. In the calibration step the blackboard reference image, which represents the noise of the CCD sensor was first subtracted from every image file. Afterwards the images were calibrated to compensate for absorption bands due to varying sensitivity of the sensor as well as for variations in reflectance values because of uneven illumination. This was done by first subtracting the blackboard reference from the whiteboard reference and then dividing each image by the new whiteboard reference image. I.e. $I_c = (I - B) / (W - B)$ where I_c , I , B and W are

the calibrated image, the original image, the blackboard reference and the whiteboard reference respectively.

2.2 NIR - Spectral Imaging System

In this section the technical specifications of the spectrograph (Table 1) and the illumination system used to produce the near-IR spectroscopic imaging data for our experiment are depicted.

Spectrograph:

Spectral Range	900 - 1700nm
Dispersion	120 nm/mm
Spectral Resolution	12 nm
Image Size	6.6 (spectral) x 8.8 (spatial) mm
Spatial Resolution	15 line-pairs/mm
Numerical Aperture	F/2.8
Slit Width	80 μm
Effective Slit Length	9.8 mm
Total Efficiency	> 50%, independent of polarization
Stray light	< 0.5%

Table 1: Technical Specifications of the Spectrograph

Illumination:

NIR-radiator (HeLeN) with reflector for focusing.

Intensity of radiation: > 70% in the range of 900 - 1700nm

With this imaging system one line of every testpanel was recorded in 180 wavelengths ranging from 900nm to 1700nm. The size of one pixel in an image is approx. 4.5mm, which causes the small number of about 40 pixels per testpanel and increases the need for a feature reduction step.

3 Feature Reduction

Classification cost (computational complexity and memory requirements) increases with the number of features used to describe pixel vectors in multispectral space. Both in training and in application the time needed for computation of e.g. the distance between two pixel vectors, the covariance matrix of a training set or the product of a vector and a covariance matrix is dependent on the size of the vector/matrix and thus dependent on the number of features used. Moreover the memory requirements also depend on the size (and hence the number of features) of vectors and matrices. For maximum

likelihood classification for example the cost increase with features is quadratic. Therefore it is sensible economically to ensure that no more features than necessary are used for classification. Features which do not aid discrimination, by contributing little to the separability of spectral classes, should be discarded since they will represent a cost burden. Furthermore an increase in the number of spectral bands demands for an increase in the number of training pixels as well to ensure reliable estimates of class statistics like mean value or covariance matrix. Thus, adding more spectral bands is not helpful unless more training pixels per class are available to train supervised classifiers for example [8].

There are two main forms of feature reduction. One is called feature selection and deals with the removal of least effective features by means of separability measures. One such measure for instance is the Jeffries-Matusita (JM) distance [1] which computes the degree of separability for two classes on the basis of first and second order class statistics. The second form of feature reduction is to transform the pixel vectors into a new coordinate system in which the features that can be removed are made more evident [8].

In view of the small number of per class pixels in our experiment it is essential for classification to reduce the large number of 180 features first. Therefore three different linear transformations are applied to the pixel vectors independently from each other to obtain the best result. Feature selection algorithms are not to be used for two reasons. First the results of a distribution-based separability measure like the JM distance are without meaning when using a non-parametric classifier like *k-nearest neighbor* and secondly the fact that 180 bands should be reduced to about five features or less implies a computationally very expensive process [8].

The three transformations applied to our test data include PCA, block-based PCA and LDA. These techniques will be explained in detail in the next three subsections.

3.1 Principal Component Analysis

The position of pixel points in hyperspectral space can be described by vectors, whose components are the individual spectral responses in each band. Little correlation between the components of the vector denotes that all components are necessary to describe where the pixel lies in space. On the other hand a high degree of correlation between two bands means that one of these two components is almost sufficient to predict where a pixel lies in space. In other words, an increase or decrease in either component suggests a corresponding increase or decrease in the other [8]. For a two dimensional example see Figure 5.

For highly correlated data like our test data this means that not all features are necessary for discrimination of the spectral distributions. Thus in a first approach we carried out a principal components transformation to the data to obtain an uncorrelated data set. The principal components transformation maps the feature vectors into a new, uncorrelated coordinate system or vector space. Moreover it produces a space in which the data has most variance along its first axes, the next largest variance along a second mutually orthogonal axes, and so on. The later principal components, in general, show little variance and it is thus assumed that they only represent noise and are not necessary for reconstruction of the original data. From this it follows that they can also be omitted for classification.

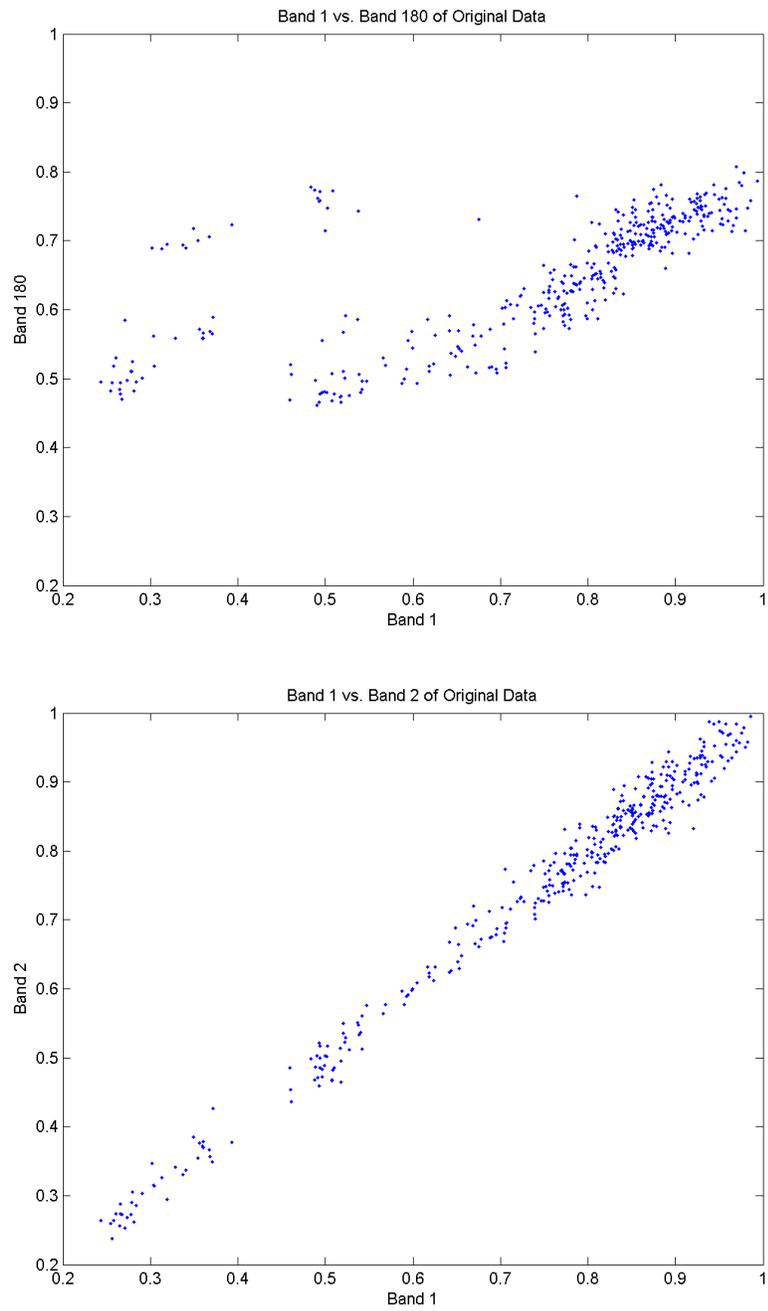


Figure 5: Two Low Correlated Bands on Top vs. Two High Correlated Bands on Bottom

It is fundamental to the development of the principal components transformation to find a new coordinate system in which the covariance matrix is diagonal, i.e. the data can be represented without correlation. So the first step is to compute the covariance matrix Σ_x of the original pixel data x space.

$$\Sigma_x = \frac{1}{N-1} \sum_{n=1}^N (x_n - m)(x_n - m)^T = \varepsilon\{(x - m)(x - m)^T\}, \quad (1)$$

where

$$m = \varepsilon\{x\} = \frac{1}{N} \sum_{n=1}^N x_n \quad (2)$$

is the expectation value or mean pixel vector and the superscript T denotes the vector transpose. The task now is to find a transformation matrix G such that

$$y = Gx \quad (3)$$

holds under the constraint that the covariance matrix of the transformed pixel vectors in y space is diagonal, i.e. the off-diagonal elements are zero. The covariance matrix in y space is

$$\Sigma_y = \varepsilon\{(y - m_y)(y - m_y)^T\}$$

with

$$m_y = \varepsilon\{y\} = \varepsilon\{Gx\} = Gm_x$$

and therefore

$$\Sigma_y = \varepsilon\{(Gx - Gm_x)(Gx - Gm_x)^T\}$$

which can be written as

$$\Sigma_y = G\varepsilon\{(x - m_x)(x - m_x)^T\}G^T, \text{ i.e.}$$

$$\Sigma_y = G\Sigma_x G^T \quad (4)$$

Because of the constraint that Σ_y has to be diagonal, G can be recognized as a transposed matrix consisting of normalized eigenvectors of Σ_x , sorted corresponding to decreasing eigenvalues. The eigenvectors in G form an orthonormal basis and G performs a coordinate transformation. In the new coordinate system the data have the largest variance along the first axes corresponding to the eigenvector that belongs to the largest eigenvalue [8].

In our data, after PCA, over 90 percent of variance can be found along the first principal axes. This result implies that the first principal axes alone could be enough to separate our classes, if they can be discriminated at all.

3.2 Linear Discriminant Analysis

Another transformation applied to our test data for feature reduction is *linear discriminant analysis* (LDA) or *discriminant feature extraction* (DAFE), which is often used for dimension reduction in classification problems. It is also called a parametric feature extraction method since it uses the mean vector and covariance matrix of each class. In contrast to PCA that aims to preserve as much variance as possible in the reduced features the purpose of LDA is to find a transformation matrix, such that the separability of the transformed data is maximized [2]. Figure 6 [6] depicts the difference between PCA and LDA for a two dimensional three-class problem where the x -axes corresponds to the first principal axes whereas the y -axes accords to the first LDA axes.

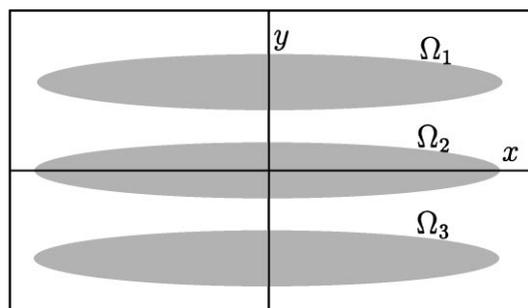


Figure 6: Behavior of PCA vs. LDA

For LDA usually within-class, between-class and mixture scatter matrices are used to formulate a measure of class separability [2].

The most commonly used method is to obtain as transformation matrix the set of vectors d_n such that the ratio of the between-class variance to the within-class variance is maximized. This criterion can be expressed as

$$C = \frac{d_n^T B d_n}{d_n^T W d_n} \quad (5)$$

In our case, the between-class covariance matrix B and the within-class covariance matrix W are defined as

$$B = \sum_{i=1}^K P_i (m_i - m) (m_i - m)^T$$

and

$$W = \sum_{i=1}^K P_i \Sigma_i$$

where m_i , Σ_i and P_i are the mean vector, covariance matrix and *a priori* probability of the i -th class respectively and m is the total mean. Using criterion (5) the best discriminant vector d_1 is given by

$$W^{-1} B d_1 = \lambda d_1.$$

Here, d_1 is the eigenvector of $W^{-1}B$ corresponding to the largest eigenvalue. In classical LDA for K classes and N bands there exist at most $K-1$ eigenvectors if $N > K-1$, which implies that LDA provides separability with reduced dimensionality. In our experiment we do not use classical LDA but a slightly modified method that produces an orthogonal set of vectors which have been shown [14] to provide a better discrimination ratio than the classical discriminant vectors [5].

The first eigenvector is obtained directly by maximizing criterion (5), which is well known as the Fisher linear discriminant. For the computation of the remaining discriminant vectors the same criterion

$$C = \frac{d_n^T B d_n}{d_n^T W d_n}$$

must be maximized with regard to the following constraints:

$$d_1^T d_n = d_2^T d_n = \dots = d_{n-1}^T d_n = 0$$

and

$$d_n^T W d_n = 1.$$

In doing so we obtain an orthogonal set of vectors which provides both a greater discrimination ratio and a better mean error probability than the classical transformation that is W -orthogonal because of the constraint $d_n^T W d_n = 1$. The resulting criterion to be maximized for vector d_n is

$$\left(I - W^{-1} [D^{(n-1)}]^T [S^{(n-1)}]^{-1} D^{(n-1)} \right) W^{-1} B d_n = \lambda d_n. \quad (6)$$

From this follows that d_n is the eigenvector of

$$M = \left(I - W^{-1} [D^{(n-1)}]^T [S^{(n-1)}]^{-1} D^{(n-1)} \right) W^{-1} B \quad (7)$$

according to the largest eigenvalue. I being the identity matrix, S being a matrix with $S_{ij}^{(n-1)} = d_i^T W^{-1} d_j$ and

$$D^{(n-1)} = \begin{bmatrix} d_1^T \\ d_2^T \\ \vdots \\ d_{n-1}^T \end{bmatrix}.$$

This method furthermore permits to obtain at most N orthogonal features. For a detailed derivation and declaration of this criterion see [5]. The composition of the obtained discriminant vectors gives the resulting transformation matrix, which transforms the original data into a new space where the classes are best separable along the first axes then along the second axes and so on [5, 7].

3.3 Block-Based PCA

A serious problem with hyperspectral data is the need to estimate class statistics (mean value, covariance matrix) both for feature reduction methods and classification algorithms.

The problem lies in the relative small number of pixels per class compared to the number of spectral bands, which makes the estimated statistics, especially the higher order statistics, unreliable. To avoid this problem the number of per class pixels should be increased to at least ten times the dimensionality of the data. Another way to overcome this drawback lies in the nature of most hyperspectral data. In general, correlations between neighboring bands are higher than for bands further apart and highly correlated bands appear in groups. Therefore the correlation matrix is roughly block diagonal.

By assuming that the subgroups of bands within each block are independent of those in other subgroups, feature reduction techniques can be applied to each group separately. In this way the dimensionality of the data can be reduced to the number of bands in the largest block, thus decreasing the number of pixels per class, required to ensure reliable class statistics. In our data set there are three blocks of very high correlation as can be seen in Figure 7. The correlation matrix has a block diagonal form where the bright areas correspond to high correlation.

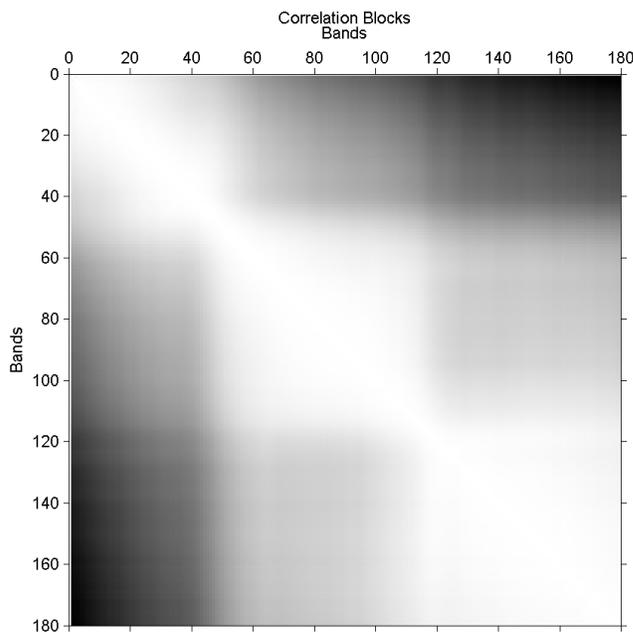


Figure 7: Grayscale Image of the Correlation Matrix of Our Test Data

In block-based principal component transformation the original data set is first partitioned into K subgroups according to the highly correlated bands, identified either visually (Figure 7) or by the use of an edge detection algorithm on the correlation matrix. Then the principal components transformation is conducted separately on each subgroup and finally feature selection is carried out on each of the transformed data sets. Feature selection can be done either by using variance information, i.e. take the first n principal components of each block, or by applying a separability measure to every transformed data set and then taking those components which contribute most to discrimination. The selected features can be regrouped and transformed again until the desired data reduction

ratio is achieved for classification purposes [8].

3.4 Feature Reduction Results

In this section we will present the results of the feature reduction techniques when applied to the single tasks as listed in Section 2.1. For task one and task two the results of PCA and block-based PCA are the same because the setup is the same and both methods do not use class information. For LDA however the results of tasks one and two differ because of their diversity in class structure.

3.4.1 PCA

The principal components transformation was conducted by using the MATLAB function *pcacov* that uses the covariance matrix of the data to compute a set of eigenvectors and eigenvalues, which minimize the correlation in the transformed set of vectors. Table 2 shows the percentage of the total variance for the first ten principal components for tasks one and two.

1	92.099	6	0.0161
2	6.724	7	0.0123
3	0.608	8	0.0105
4	0.357	9	0.0081
5	0.0531	10	0.0060

Table 2: Percentage of Total Variance for the First Ten PC's

As one can see in Table 2 more than 98 percent of the variance is covered by the first two principal components and the first three together account for 99.43 percent of total data variance. This gives rise to the assumption that the later components contribute very little information and could therefore be negligible for classification purposes. This becomes evident in Figure 8 too, where we can see a plot of the first three principal components in red, green and blue respectively. Figure 9 shows an image of the correlation matrix of the transformed data which reveals no correlation between bands in contrast to Figure 7.

Figure 10 shows a scatterplot of the first three features of the pc-transformed data for task 1. In this and in all following scatterplots each class is represented by one marker style and color. Again one can see that they are not correlated but the strong blending of the classes implies no good classification result. The scatterplot for task two in Figure 11 also reveals one big cluster as the transformation is the same as for task 1. But because in this case there are only three classes better classification results seem to be possible.

For task three we obtain three results because we must conduct the transformation on each matrix separately. In Table 3 the percentage of total variance of the first three components are listed for every binder. Again these components account for more than 99 percent of total variance. As for task one and task two the covariance matrices for task three are diagonal and the first three principal components should describe the structure of the original data nearly without loss of information. In Figures 12, 13 and 14 we see

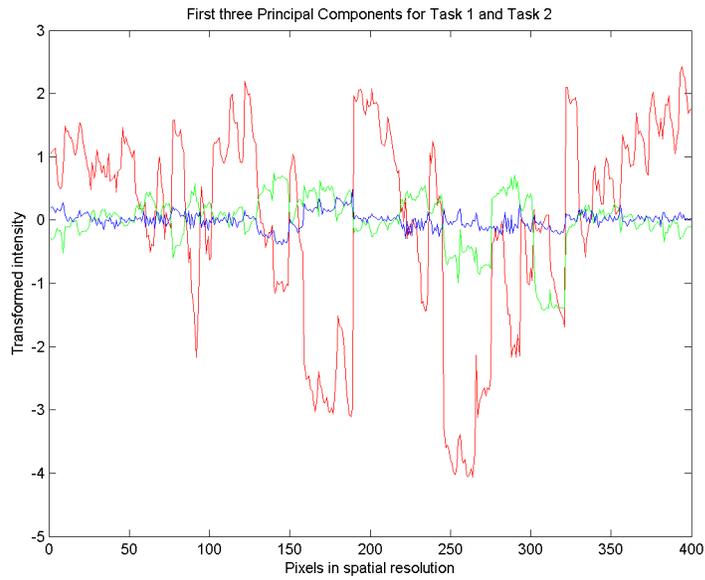


Figure 8: First Principal Component in Red, the Second in Green and the Third in Blue

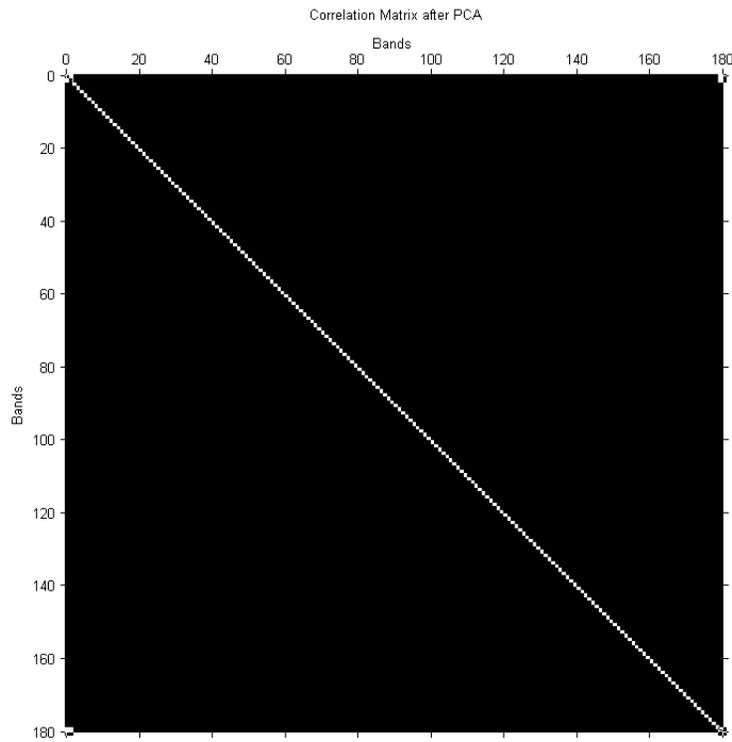


Figure 9: Correlation Matrix of the 180 Features After PCA Transformation

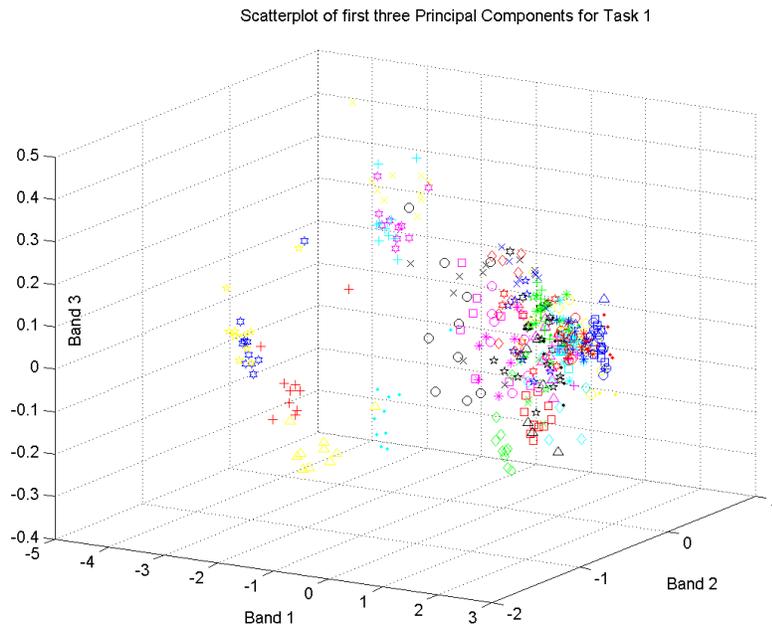


Figure 10: Scatterplot of the First Three Bands After PC Transformation of Data for Task1

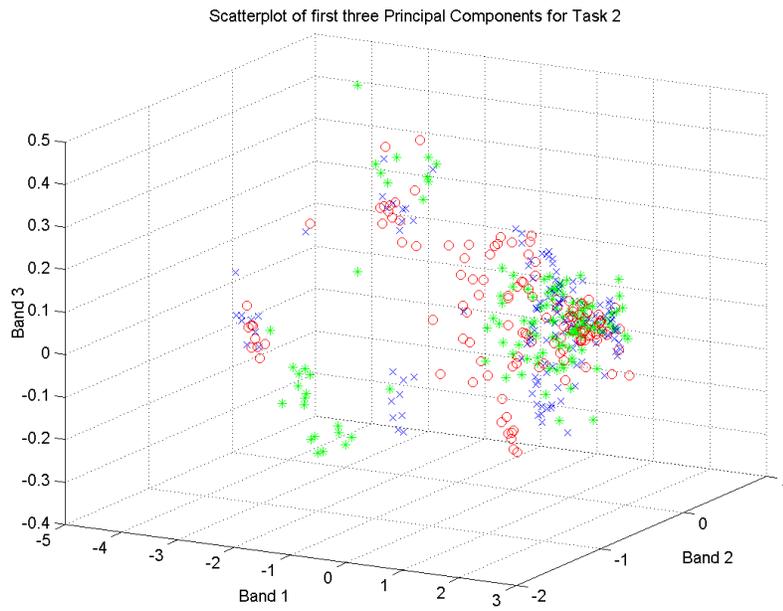


Figure 11: Scatterplot of the First Three Bands After PC Transformation of Data for Task2

the scatterplots of the first three features of the pc-transformed data for each binder and once again one can spot an area where the data points cluster but the mixture of data points is not as strong as in task one.

Binder 1	91.026	7.839	0.579
Binder 2	95.963	3.206	0.573
Binder 2	90.371	8.235	0.627

Table 3: Percentage of Total Variance of the First Three PC's of Each Binder

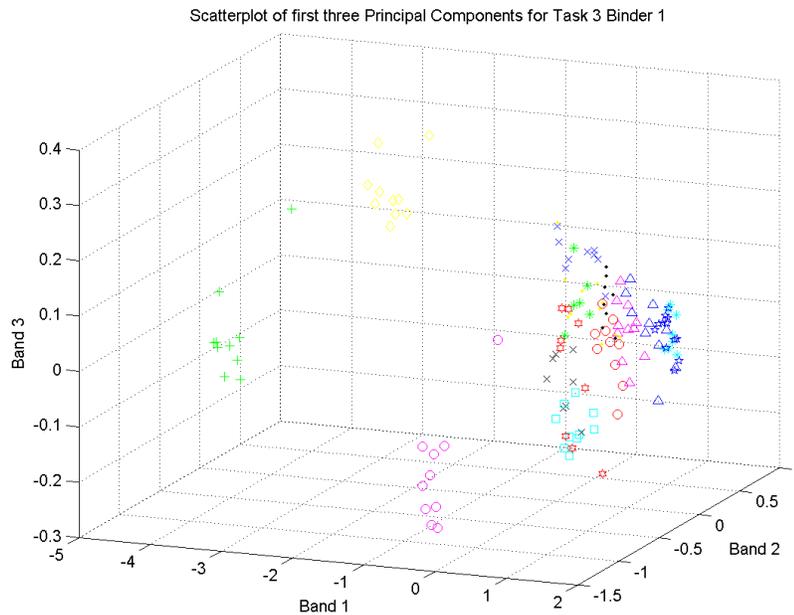


Figure 12: Scatterplot of the First Three Bands After PC Transformation of Data for Task3 Binder1

3.4.2 Block-Based PCA

According to Section 3.3 three major blocks of high correlation can be identified within our data set. The first block covers the first 50 bands, the second the next 68 and the third block the last 62 bands. After dividing the whole data set into these groups, principal component transformation was applied to each block separately. Table 4 shows the amount of data variance in percentage according to the first principal component of each group for tasks one and two.

In Figure 15 the first three principal components of the first block for tasks one and two are depicted as an example. Due to the vast portion of variance we decided to take the first component of every group and to put them together for classification. In Figure 16 we

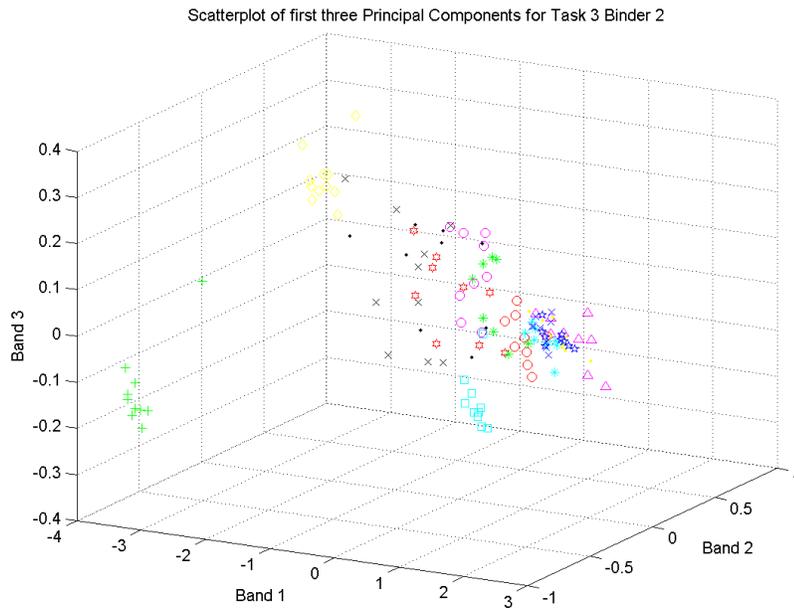


Figure 13: Scatterplot of the First Three Bands After PC Transformation of Data for Task3 Binder2

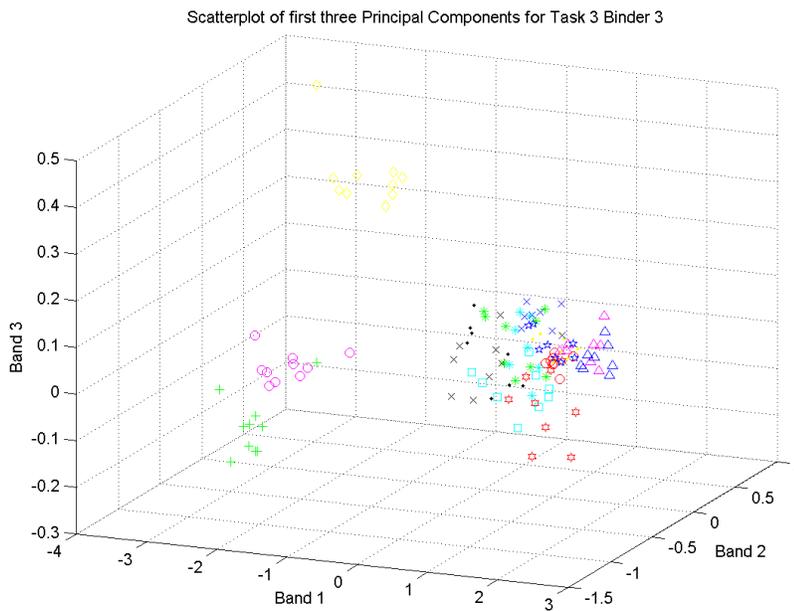


Figure 14: Scatterplot of the First Three Bands After PC Transformation of Data for Task3 Binder3

Block	Variance
1	98.507
2	98.795
3	99.5

Table 4: Percentage of Total Variance for the First PC of Each Block

can see a scatterplot of the first band of each of the three blocks after PCA was conducted separately for every block. Although the correlation between the original blocks is low the correlation of the three components is quite high compared to the result of the original PCA. This plot also reveals a mixture of classes as in case of PCA.

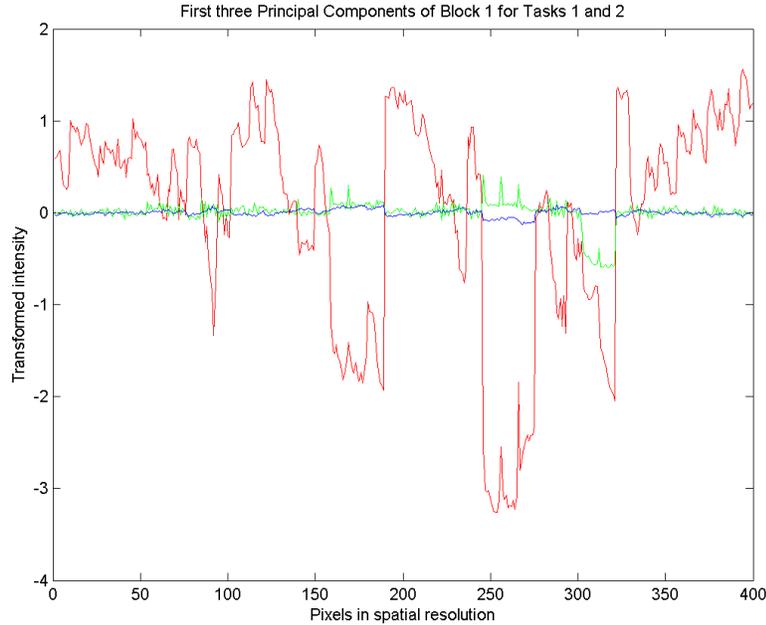


Figure 15: First Three Principal Components of Block1

Therefore the reduction to three features was not as effective as with PCA what could imply a worse classification result. As in case of pure PCA the result of block-based PCA for task two is the same as for task one but again the existence of only three classes promises a better outcome in the classification step as illustrated in Figure 17.

Table 5 shows the amount of data variance in percentage according to the first principal component of each group for task three. Because once more the variances of the first components are very high, we take the first PC of each block for every binder and put them together for classification. Figures 18, 19 and 20 show the scatterplots of the first component of each block for binder one, two and three. Again these results look more promising than those from task one and task two.

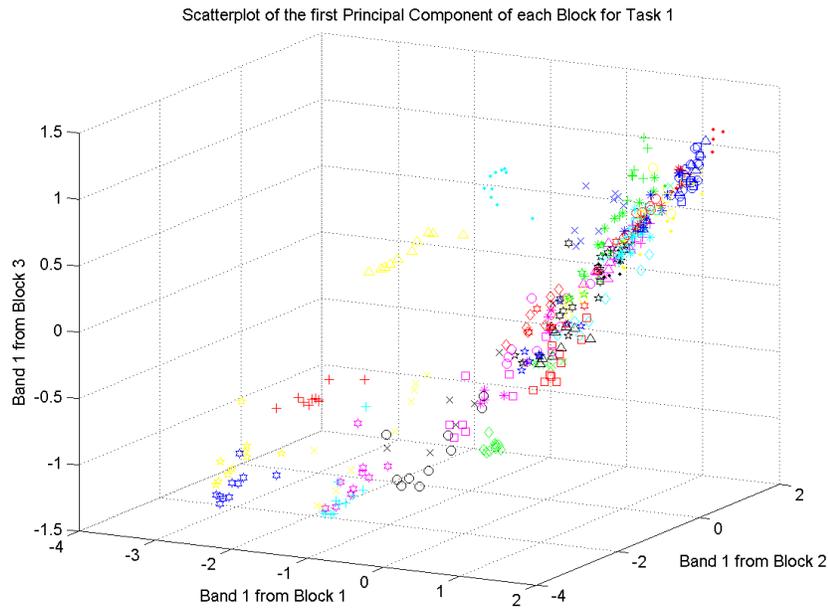


Figure 16: Scatterplot of the First Band of each Block After PC Transformation for Task1

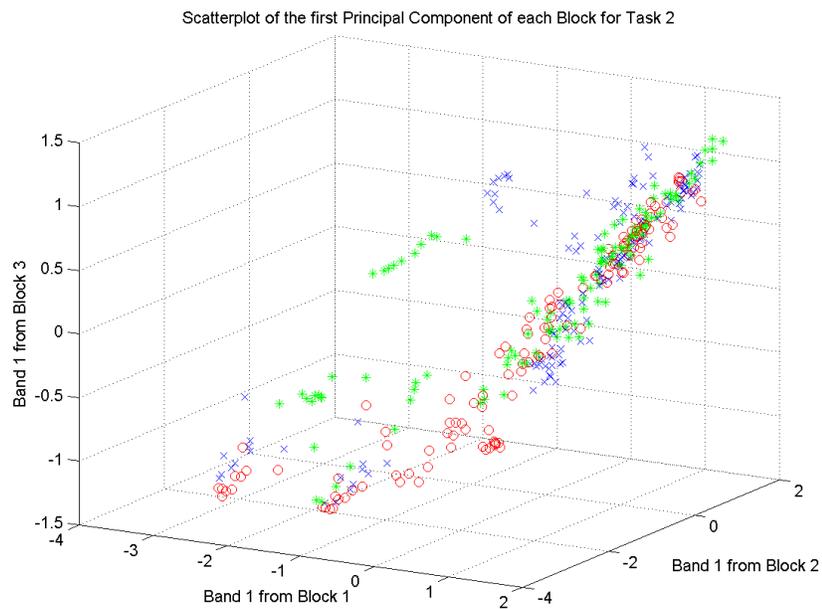


Figure 17: Scatterplot of the First Band of each Block After PC Transformation for Task2

	Block1	Block2	Block3
Binder 1	98.31	98.66	99.41
Binder 2	99.57	99.27	99.67
Binder 2	97.95	98.32	99.29

Table 5: Percentage of Total Variance of the First PC of Every Block in Each Binder

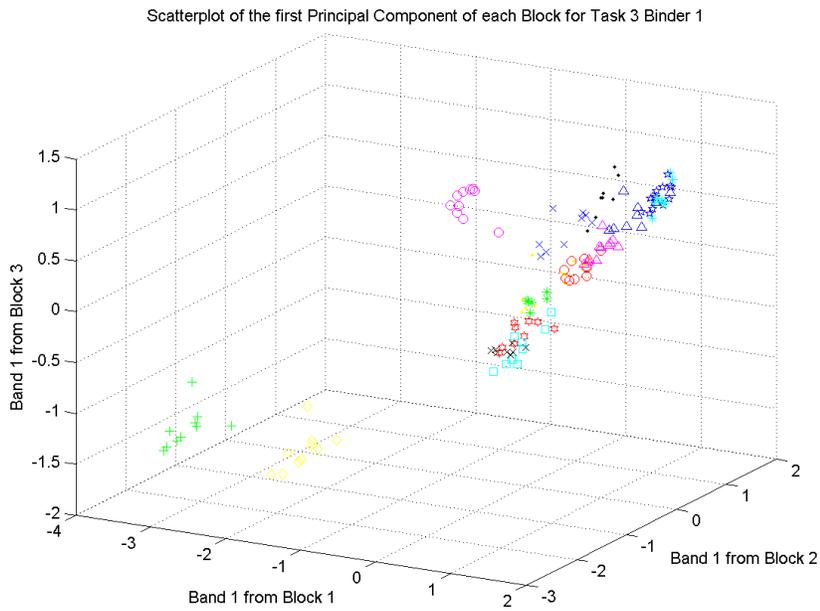


Figure 18: Scatterplot of the First Band of each Block After PC Transformation for Task3 Binder1

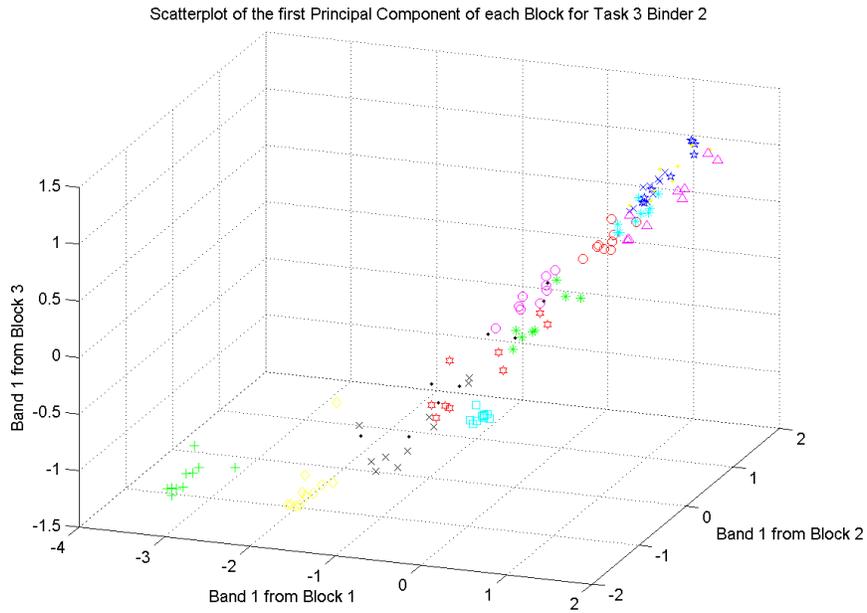


Figure 19: Scatterplot of the First Band of each Block After PC Transformation for Task3 Binder2

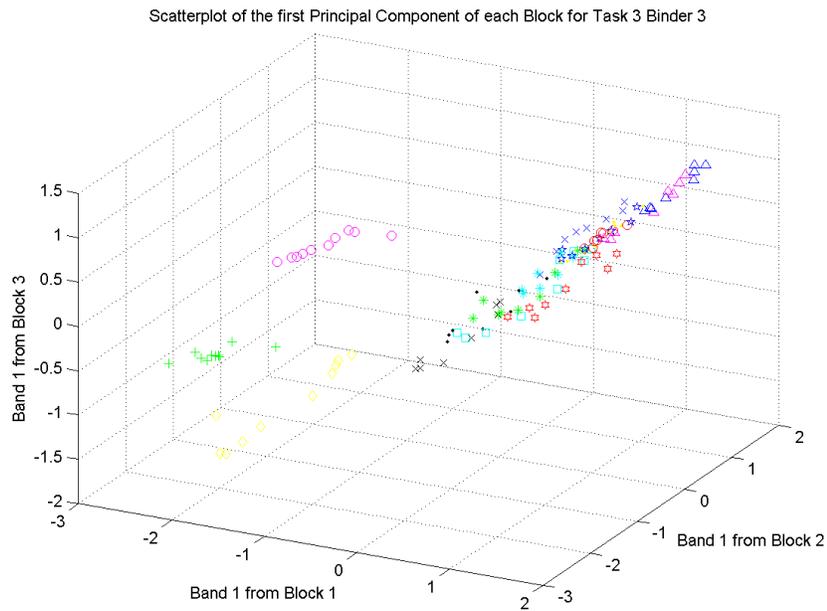


Figure 20: Scatterplot of the First Band of each Block After PC Transformation for Task3 Binder3

3.4.3 LDA

Figure 21 plots the first three components of the result of LDA, when applied to the data set of task one.

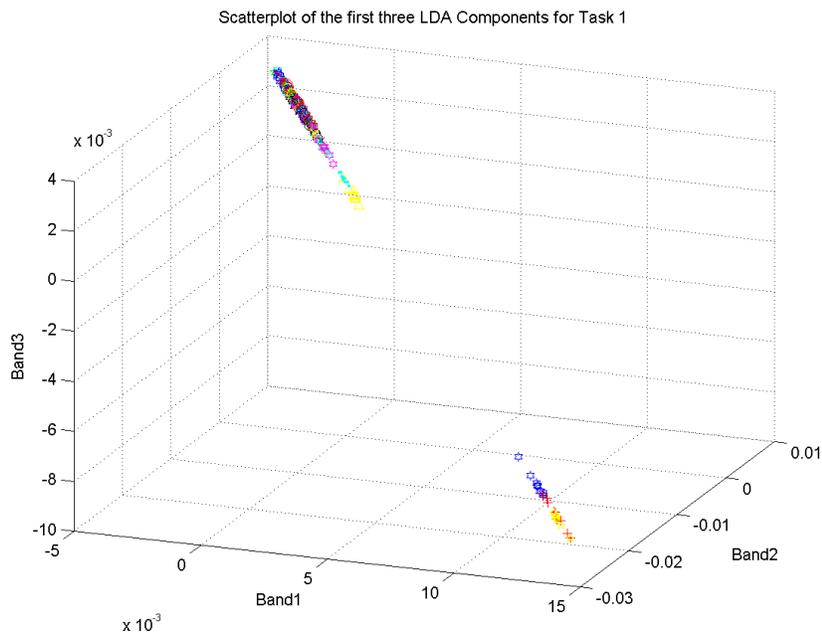


Figure 21: Scatterplot of the First Three Bands After LDA for Task1

Surprisingly the three components are highly correlated and no clusters can be recognized although the separation between classes should have been maximized by the LDA transformation. So the expected classification results for the first three LDA components in task one should be very bad. For task two the result is visually much better as can be seen in Figure 22. In this case the three classes are much more separable than the 44 classes in task one.

The outcomes of LDA transformation for task three are depicted in Figures 23, 24 and 25. They also look very promising with regard to nearest neighbor classification, especially the results for binder *Scotch Glue*.

In conclusion one can say that the classification accuracy in task two and three should be much better than in task one, when visually investigating the first three components of each feature reduction technique. According to the figures plotted in this section and with regard to nearest neighbor classification the best feature reduction results for task one can be achieved with PCA and block-based PCA. For task two and three however LDA seems to perform at least a little better than PCA and block-based PCA. In general it appears as if three features would not suffice to achieve good results in classification.

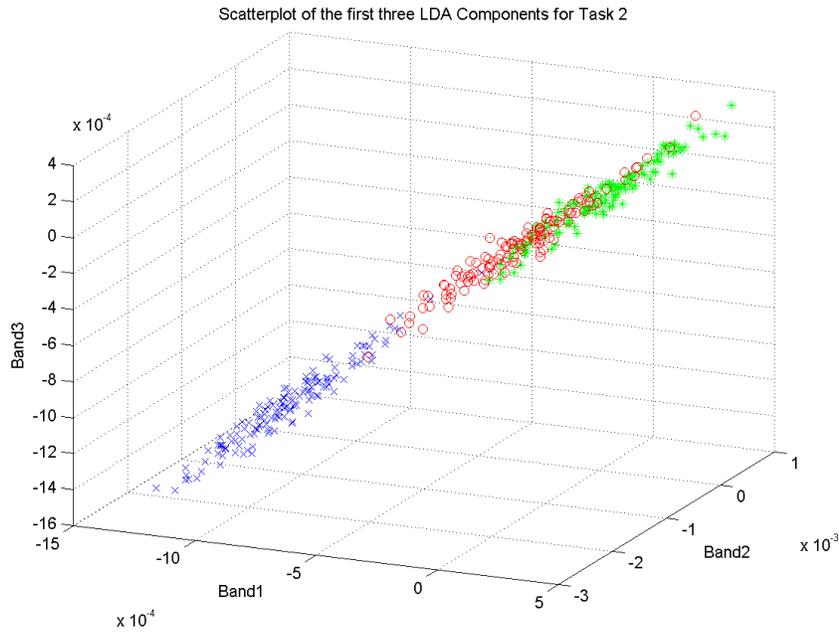


Figure 22: Scatterplot of the First Three Bands After LDA for Task2

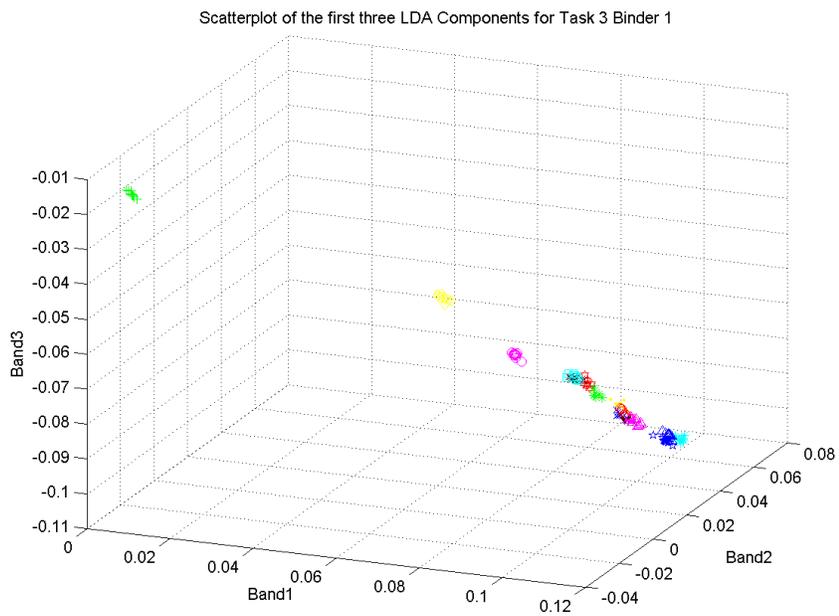


Figure 23: Scatterplot of the First Three Bands After LDA for Task3 Binder1

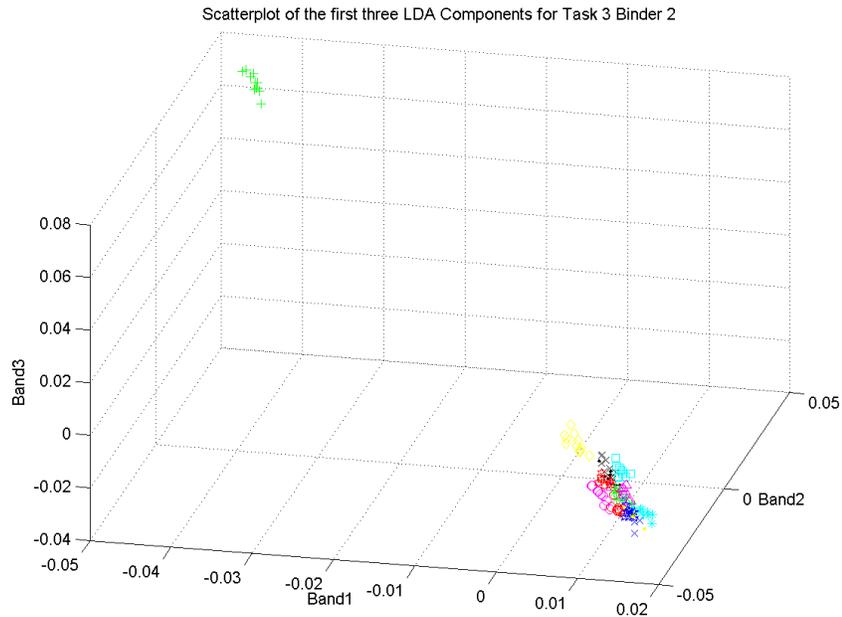


Figure 24: Scatterplot of the First Three Bands After LDA for Task3 Binder2

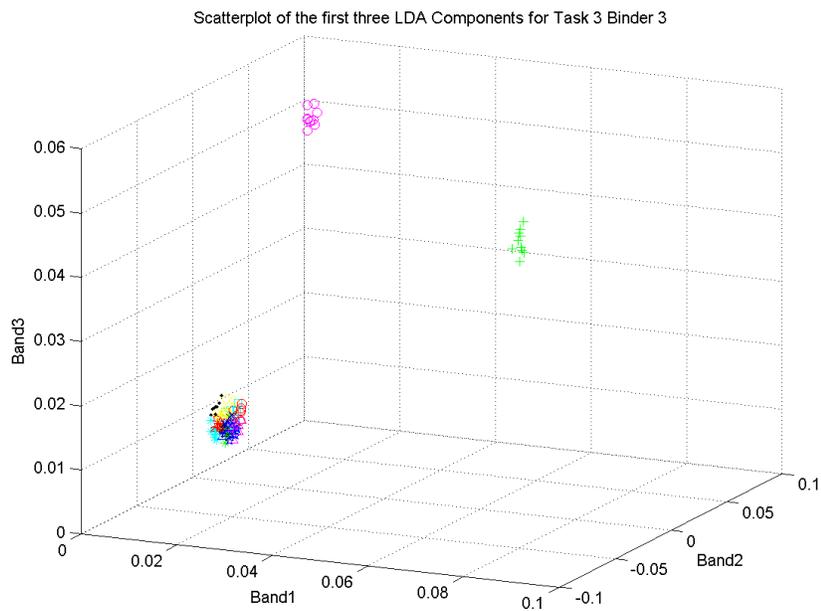


Figure 25: Scatterplot of the First Three Bands After LDA for Task3 Binder3

4 Classification

In this section we are going to describe both the *k-nearest neighbor* classification algorithm, which is then used to classify our test data and the distance measures used within the algorithm. For each task the classification method is applied both to the original data set and to the results of the feature reduction step as specified in Table 6 below. The outcomes of the classification process are finally depicted in the following section.

- Original data
- First three principal components
- First ten principal components
- First component of each PC-transformed correlation block put together to a single matrix
- First three elements of the LDA transformation
- First n features of LDA transformed data, where $n > 3$ and specific for each task to achieve a better result as with the first three features

Table 6: List of Feature Sets Used for Classification

4.1 K-Nearest Neighbor Classification

The *K-nearest neighbor* (KNN) algorithm is a non-parametric classification technique which has been shown to be very effective in statistical pattern classification applications. Non-parametric methods have the advantage of not needing an assumption of the class density function and therefore can achieve high classification accuracy in problems which have unknown or non-normal distributions [4].

The KNN algorithm assumes that pixels close to each other in feature space are likely to belong to the same class. It is based on a distance measure $d(x, y)$ that is applied to every pair of observations. The distances used in our case (*euclidean*, *standardized euclidean*, *cityblock* and *minkowski*) are described below. Suppose $T = \{t_1, \dots, t_r\}$ is the set of feature vectors used as training set labeled by their class labels $C = \{c_1, \dots, c_m\}$ and x is a vector whose class label is unknown. We then have to find the set S_x containing the k closest neighbors of x according to the distance measure d in T . Let c_i be the class label that appears most often in S_x then x is assigned the class label c_i . In case there is more than one class label with equal occurrence in S_x there are different decision rules for which one to take. In our case we decide for the class label of the vector nearest to x , which turned out to yield the best results [4, 13, 15].

Because of the small number of samples per class we do not divide the data set into training and test set and then classify the test set. Instead we use a leave-one-out cross

validation approach, i.e. we try to classify each vector of the data set separately by taking all remaining feature vectors as training set.

4.2 Distance Measures

For row vectors x_1, \dots, x_m of an m -by- n data matrix X the distances between two vectors x_r and x_s are defined as follows:

- *Euclidean Distance:*

$$d_{rs} = \sqrt{(x_r - x_s)(x_r - x_s)^T}$$

- *Standardized Euclidean Distance:*

$$d_{rs} = \sqrt{(x_r - x_s) D^{-1} (x_r - x_s)^T}$$

where D is the diagonal matrix with diagonal elements given by v_j^2 , which denotes the variance of the variable X_j over the m objects.

- *City Block Metric:*

$$d_{rs} = \sum_{j=1}^n |x_{rj} - x_{sj}|$$

- *Minkowski Metric:*

$$d_{rs} = \left(\sum_{j=1}^n |x_{rj} - x_{sj}|^p \right)^{\frac{1}{p}}$$

where p was chosen 4 because with $p = 4$ the best results were achieved.

5 Results

In this section we will provide the results of the classification step for each task. Classification was done using the *k-nearest neighbor* algorithm as described in Section 4.1. and the feature sets listed in Table 6. The results are presented as bar charts where *Original* stands for the 180 non-transformed features and the numbers behind *LDA* and *PCA* stand for the number of features used for classification.

In Figure 26 we can see the classification results for task one. The results are quite bad as expected after feature reduction, the maximum classification accuracy comes up to 45 percent, achieved with the first ten principal components.

For task two the results are much better as can be inferred from Figure 27. The classification accuracy of 99 percent achieved with the first ten LDA transformed features

corresponds to a correct classification of 396 samples out of 400. The best performance in classification was attained within task three. There the correctness for binder *Scotch Glue* reached a maximum of 99.3 percent, which conforms to only a single misclassified sample point (Figure 28). Though not as good as the output for binder *Scotch Glue* the results for binder *Linseed Oil* (Figure 29) and binder *Egg Tempera* (Figure 30) at 92.3 and 91.4 percent respectively are quite acceptable nevertheless.

Obviously it is not possible to discriminate the 44 classes of task one with a satisfying accuracy (at least with the setup and tools used in our experiment). A better way for classifying color pigments could therefore be a two step approach that consists of the identification of the binder in the first step and the classification of the pigments within each class of binder in the second step.

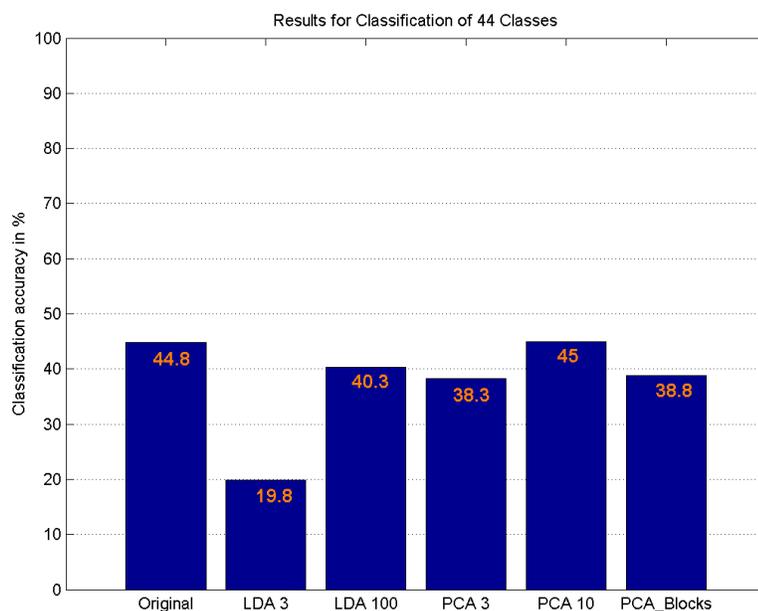


Figure 26: Classification Results for Task1

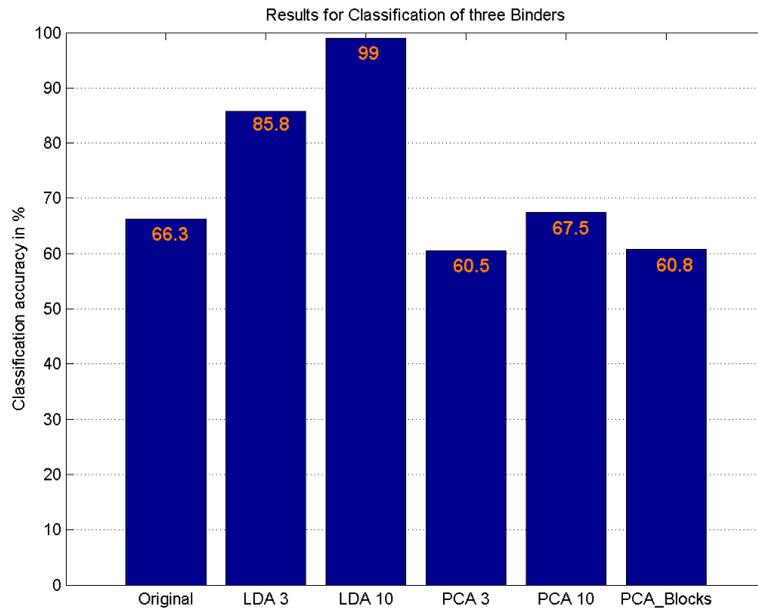


Figure 27: Classification Results for Task2

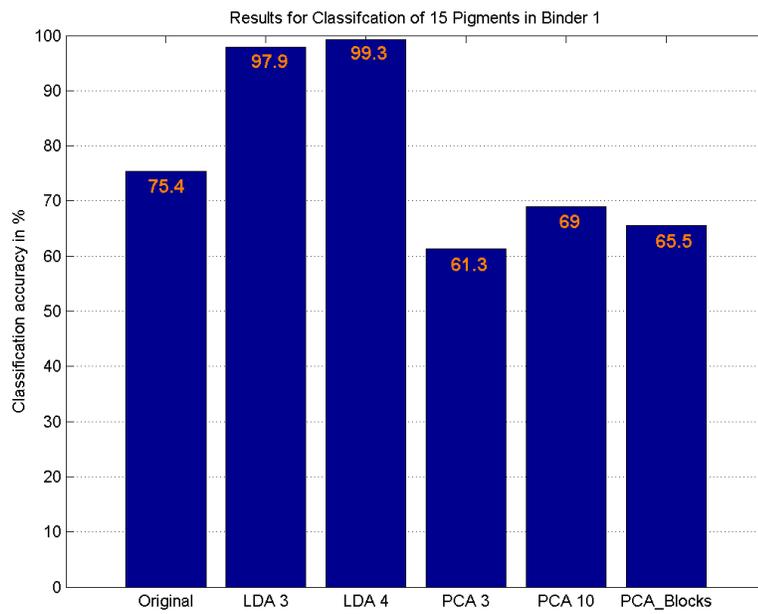


Figure 28: Classification Results for Task3 Binder1

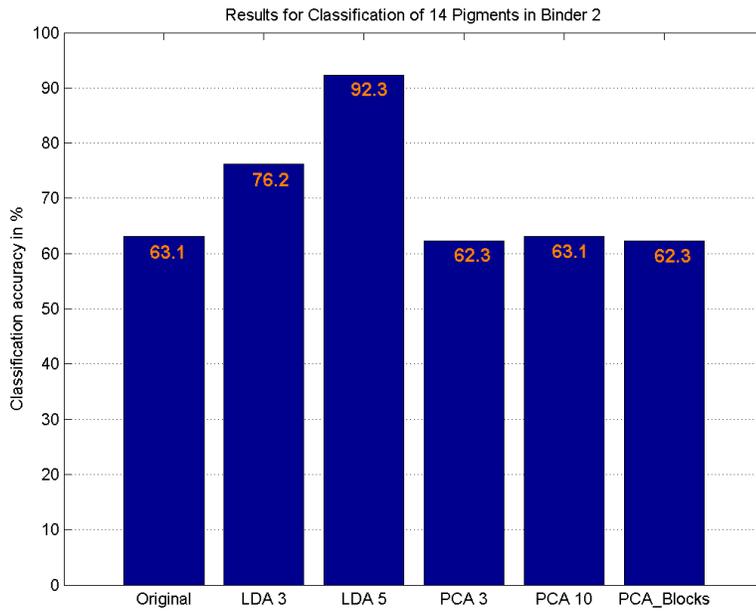


Figure 29: Classification Results for Task3 Binder2

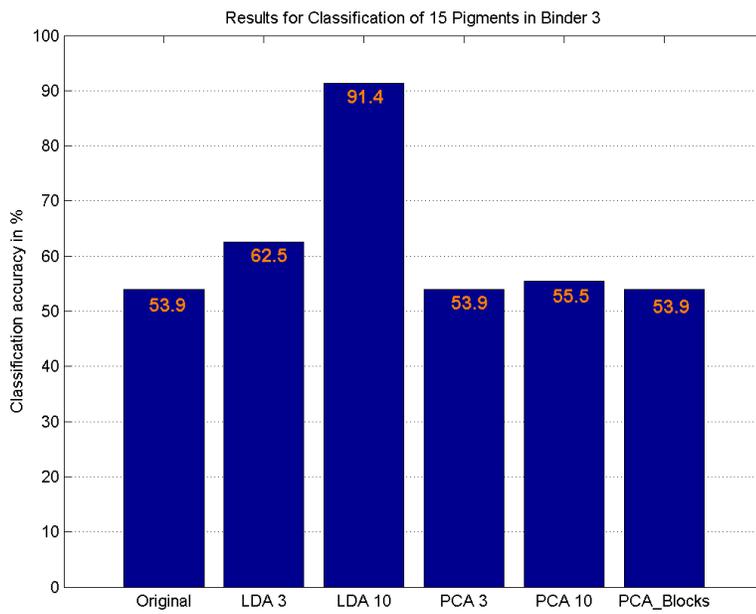


Figure 30: Classification Results for Task3 Binder3

6 Conclusion

The results of our experiment concerning the classification of color pigments showed that it is basically feasible to distinguish these pigments by means of hyperspectral images in the range from 900nm to 1700nm. After calibration of the images three feature reduction schemes were applied to the near-IR spectroscopic images to reduce the dimensionality of the data (180 features) to a smaller and more meaningful set before classification. While PCA and LDA are widely used transformations for multi- and hyperspectral data, block-based PCA emerged from the nature of hyperspectral data which often exhibits high correlation blocks. As classifier the *cross-validators k-nearest neighbor* algorithm was used to classify each sample vector against the remaining data set.

In the classification step it was possible to identify the 14 to 15 distinct pigments for each binder with an accuracy of more than 90 percent with the best ratio being 99.3 percent. The experiment furthermore revealed that the three binders influence the transparency of the color pigments with unequal strength and must therefore not be neglected. Due to this slightly differences is it possible to discriminate the three kinds of binder as well, what was done with an accuracy of 99 percent in our test data. Whereas the attempt to identify 44 classes consisting of the color pigments mixed with each binder was a disappointment because at most only 45 percent of the samples could be classified correctly. In four out of five cases the best results were achieved using LDA for feature reduction and only for the problem with 44 classes PCA performed best. In general, as one can see by means of the bar charts in Section 5 the originally 180 features can not be reduced to three features by attaining maximum classification accuracy at the same time.

The reason for the bad result in the 44 classes problem lies for the most part in the small number of pixels per class compared with the number of classes and features. Firstly the number of trainings pixel influences the reliability of class statistics required for LDA analysis and secondly small differences between two classes can be lost if the number of per class samples is too small. Therefore more images of each panel or a sensor with a higher spatial resolution could avoid the problem of the high misclassification ratio for task one.

Future Goals: For practical applications in painting analysis a more systematic evaluation of all constituents of a painting has to be performed. Different types of grounding as well as different materials used for underdrawings and color pigments in combination with diverse binders have to be distinguished due to their different interdependencies with respect to spectral response. Therefore a large database of all these configurations will be necessary for analysis of paintings by means of IR spectroscopy.

A further improvement could consist in segmenting the applications of a painting by using a scanning device that produces a spectral image of the painting both in the near infrared and in the visible range. The image can then be segmented into regions that contain parts of the underdrawing for classification of the type of drawing material used and into regions with no underdrawing for classification of the color pigments used. This would allow for a complete analysis of a painting including the extraction of strokes of underdrawings for the distinction between strokes on the surface and strokes of the underdrawing by integrating also the visual light range. In doing so one could get a useful analysis tool for art historians and restorers.

Acknowledgements

The author is grateful to Prof. Franz Mairinger for the contribution of the test panels and his expert knowledge in the field of spectral analysis. Special thanks go to the Department of Automation for Quality Control at Austrian Research Centers Seibersdorf for providing their hyperspectral acquisition system.

References

- [1] Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943.
- [2] Kuo B-C. and Landgrebe D.A. Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(5):1096–1105, 2004.
- [3] Anglos D., Balas C., and Fotakis C. Laser spectroscopic and optical imaging techniques in chemical and structural diagnostics of painted artwork. *American Laboratory*, 31:60–67, 1997.
- [4] Crimins F., Dimitri R., Klein T., Palmer N., and Cowen L. Higher dimensional approach for classification of lung cancer microarray data. *4th Annual Conference on Critical Assessment of Microarray Data*, Durham, USA, 2003. To appear in 'Methods of Microarray Data Analysis IV', published by Kluwer academic press.
- [5] Duchene J. and Leclercq S. An optimal transformation for discriminant and principal component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):978–983, 1988.
- [6] Hornegger J., Niemann H., and Risack R. Appearance-based object recognition using optimal feature transforms. *Pattern Recognition*, 33:209–224, 2000.
- [7] Ye J., Janardan R., Li Q., and Park H. Feature extraction via generalized uncorrelated linear discriminant analysis. In *Proceedings of the 21st International Conference on Machine Learning*, pages 895–902, Banff, Canada, 2004.
- [8] Richards J.A. and Jia X. *Remote Sensing Digital Image Analysis, An Introduction*. Springer Verlag, third edition, 1999.
- [9] Mansfield J.R., Sowa M.G., Majzels C., Collins C., Cloutis E., and Mantsch H.H. Near infrared spectroscopic reflectance imaging: Supervised vs. unsupervised analysis using an art conservation application. *Vibrational Spectroscopy*, 19(1):33–45, 1999.
- [10] Bacci M., Casini A., Cucci C., Picollo M., Radicati B., and Vervat M. Non-invasive spectroscopic measurements on the 'il ritratto della figliastra' by giovanni fattori: Identification of pigments and colourimetric analysis. *Journal of Cultural Heritage*, 4(4):329–336, 2003.

- [11] Hain M., Bartl J., and Jacko V. Multispectral analysis of cultural heritage artefacts. *Measurement Science Review*, 3(3), 2003.
- [12] Berns R.S. and Imai F.H. The use of multi-channel visible spectrum imaging for pigment identification. In *Proceedings of The 13th Triennial ICOM-CC meeting*, pages 217–222, Rio de Janeiro, 2002.
- [13] Baek S.J. and Sung K.-M. Fast k-nearest neighbour search algorithm for nonparametric classification. *Electronics Letters*, 36(21):1821–1822, 2000.
- [14] Okada T. and Tomita S. An optimal orthonormal system for discriminant analysis. *Pattern Recognition*, 18(2):139–144, 1985.
- [15] Jia X. and Richards J.A. Cluster-space classification: A fast k-nearest neighbour classification for remote sensing hyperspectral data. In *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pages 407–410, Greenbelt, USA, 2003.