Kernel Maximum Autocorrelation Factor and Minimum Noise Fraction Transformations

Allan Aasbjerg Nielsen

Abstract—This paper introduces kernel versions of maximum autocorrelation factor (MAF) analysis and minimum noise fraction (MNF) analysis. The kernel versions are based upon a dual formulation also termed Q-mode analysis in which the data enter into the analysis via inner products in the Gram matrix only. In the kernel version, the inner products of the original data are replaced by inner products between nonlinear mappings into higher dimensional feature space. Via kernel substitution also known as the kernel trick these inner products between the mappings are in turn replaced by a kernel function and all quantities needed in the analysis are expressed in terms of this kernel function. This means that we need not know the nonlinear mappings explicitly. Kernel principal component analysis (PCA), kernel MAF, and kernel MNF analyses handle nonlinearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space. Three examples show the very successful application of kernel MAF/MNF analysis to: 1) change detection in DLR 3K camera data recorded 0.7 s apart over a busy motorway, 2) change detection in hyperspectral HyMap scanner data covering a small agricultural area, and 3) maize kernel inspection. In the cases shown, the kernel MAF/MNF transformation performs better than its linear counterpart as well as linear and kernel PCA. The leading kernel MAF/MNF variates seem to possess the ability to adapt to even abruptly varying multi and hypervariate backgrounds and focus on extreme observations.

Index Terms—Dual formulation, kernel maximum autocorrelation factor (MAF), kernel minimum noise fraction (MNF), kernel substitution, kernel trick, orthogonal transformations, Q-mode analysis.

I. INTRODUCTION

BASED upon work by Pearson [1] in 1901, Hotelling [2] in 1933 introduced principal component analysis (PCA). PCA is often used for linear orthogonalization or compression by dimensionality reduction of correlated multivariate data, see [3] for a comprehensive description of PCA and related techniques. An interesting dilemma in the reduction of dimensionality of data is the desire to obtain simplicity for better understanding, visualization and interpretation of the data on the one hand, and the desire to retain sufficient detail for adequate representation on the other hand.

Switzer and Green [4] introduced maximum autocorrelation factor (MAF) analysis. In [5], MAF analysis was used and in [6] PCA was used to detect change in images consisting of simple differences between corresponding spectral bands acquired at two points in time. Green *et al.* [7] introduced the minimum noise fraction (MNF) transformation. Both the MAF and the MNF transformations contain spatial elements and they are, therefore, (conceptually) better suited for spatial data than PCA.

Schölkopf *et al.* [8] introduced kernel PCA. Lai and Fyfe [9] described kernel canonical correlation analysis (CCA), and Bach and Jordan [10] described kernel independent component analysis (ICA) based upon kernel CCA. Good general references for kernel methods are [11], [12]. Kernel methods among many other subjects are described in [13], [14]. In [15], kernel PCA is used for change detection in univariate image data. In [16], the kernel MAF transformations is applied to analyze irregularly spaced geochemical data.

In this paper, Section II sketches linear PCA primarily to motivate the reparameterization used in dual or Q-mode analysis. Based upon the usual combination of the dual formulation and kernel substitution [8], [12] kernel versions of MAF and MNF analyses are introduced in Sections III and IV, respectively. These transformations may be used for general feature generation as preprocessing to for example classification tasks. Here, we apply kernel PCA and MAF/MNF analysis to three examples. The data used in the examples are described in Section V. In the first two examples, we detect change over time in remotely sensed images. This is done by finding the projections along the eigenvectors for data consisting of simple band-by-band differences of coregistered, calibrated variables which represent the same spectral bands covering the same geographical region acquired at two different time points, one on airborne camera data over a busy motorway , [17], and another one on airborne, hyperspectral scanner data over a small agricultural region [18]. In a third example, the methods are applied to hyperspectral image data in maize kernel inspection [19]. Results are described in Section VI. Section VII concludes and an Appendix gives details on maximizing a Rayleigh quotient with special attention to the case where the matrix in the denominator is not full rank.

The kernel versions of PCA and MAF/MNF handle nonlinearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space. For the kernel MAF/MNF variates to be used in for example denoising we must look into the so-called preimage problem, [20]–[22]. This deals with the complicated problem of mapping back from the feature space defined implicitly by the kernel function to the original variable space. This is not described here.

II. PCA

Let us consider a data set (for example an image) with n observations (or pixels) and p variables (or spectral bands) organized as a matrix X with n rows and p columns; each column contains measurements over all observations from one variable and each row consists of a vector of measurements x_i^T from p variables for a particular observation, the superscript T denotes the transpose

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}. \tag{1}$$

X is often called the data matrix. Without loss of generality we assume that the variables in the columns of X have mean value zero.

A. Primal Formulation

In ordinary (primal also known as R-mode) PCA we analyze the sample variance-covariance matrix $S = X^T X/(n-1) = 1/(n-1)\sum_{i=1}^{n} x_i x_i^T$ which is p by p. If $X^T X$ is rank $r \le \min(n,p)$ this will lead to r nonzero eigenvalues λ_i and r orthogonal or mutually conjugate unit length eigenvectors u_i ($u_i^T u_j = 0, i \ne j$; $u_i^T u_i = 1$) from the eigenvalue problem

$$\frac{1}{n-1} X^T X u_i = \lambda_i u_i. \tag{2}$$

We see that the sign of u_i is arbitrary. To find the principal component scores for an observation x we project x onto the eigenvectors, $x^T u_i = u_i^T x$. The variance of these scores is $u_i^T S u_i = \lambda_i u_i^T u_i = \lambda_i$ which is maximized by solving the eigenvalue problem, see the Appendix with matrix B there equal to the identity matrix.

B. Dual Formulation

To get to the dual formulation (also known as Q-mode analysis) multiply both sides of (2) from the left with X

$$\frac{1}{n-1} X X^T (X u_i) = \lambda_i (X u_i) \tag{3}$$

or

$$\frac{1}{n-1} X X^T v_i = \lambda_i v_i \tag{4}$$

with v_i proportional to $X u_i$, $v_i \propto X u_i$, which is normally not normed to unit length if u_i is. The elements in the so-called Gram[1] matrix $X X^T$ are the inner products $x_i^T x_j$ of the rows in X. Now multiply both sides of (4) from the left with X^T

$$\frac{1}{n-1} X^T X (X^T v_i) = \lambda_i (X^T v_i) \tag{5}$$

to show that $u_i \propto X^T v_i$ is an eigenvector of S with the same eigenvalue λ_i. We scale these eigenvectors to unit length assuming that v_i are unit vectors ($1 = v_i^T v_i \propto u_i^T X^T X u_i = (n-1)\lambda_i u_i^T u_i = 1$)

$$u_i = \frac{1}{\sqrt{(n-1)\lambda_i}} X^T v_i. \tag{6}$$

[1]Named after Danish mathematician Jørgen Pedersen Gram (1850–1916).

We see that $X^T X/(n-1)$ and $X X^T/(n-1)$ have the same r nonzero eigenvalues λ_i and that their eigenvectors are related by $u_i = X^T v_i/\sqrt{(n-1)\lambda_i}$ and $v_i = X u_i/\sqrt{(n-1)\lambda_i}$. This result is closely related to the Eckart–Young theorem [23], [24].

III. MAF ANALYSIS

In MAF analysis first suggested in [4], we maximize the autocorrelation of linear combinations, $a^T x(r)$, of zeroth-mean original (spatial) variables, $x(r)$, see also [5], [25], [26]. $x(r)$ is a multivariate observation at location r and $x(r+\Delta)$ is an observation of the same variables at location $r+\Delta$; Δ is a spatial displacement vector.

The autocovariance R of a linear combination $a^T x(r)$ of zeroth-mean $x(r)$ is

$$R = \text{Cov}\{a^T x(r), a^T x(r+\Delta)\} \tag{7}$$
$$= a^T \text{Cov}\{x(r), x(r+\Delta)\}a \tag{8}$$
$$= a^T C_\Delta a \tag{9}$$

where C_Δ is the covariance between $x(r)$ and $x(r+\Delta)$. Assuming or imposing second-order stationarity of $x(r)$, C_Δ is independent of location, r. Introduce the multivariate difference $x_\Delta(r) = x(r) - x(r+\Delta)$ with variance-covariance matrix $S_\Delta = 2S - (C_\Delta + C_\Delta^T)$ where $S = X^T X/(n-1)$ is the variance-covariance matrix of x. Since ($a^T C_\Delta a$ is a scalar)

$$a^T C_\Delta a = (a^T C_\Delta a)^T \tag{10}$$
$$= a^T C_\Delta^T a \tag{11}$$
$$= a^T (C_\Delta + C_\Delta^T) a/2 \tag{12}$$

we obtain

$$R = a^T (S - S_\Delta/2)a. \tag{13}$$

To get the autocorrelation ρ of the linear combination we divide the covariance by its variance $a^T S a$

$$\rho = 1 - \frac{1}{2} \frac{a^T S_\Delta a}{a^T S a} \tag{14}$$
$$= 1 - \frac{1}{2} \frac{a^T X_\Delta^T X_\Delta a}{a^T X^T X a} \tag{15}$$

where X is the data matrix defined previously and X_Δ is a similarly defined matrix for x_Δ with zeroth-mean columns. C_Δ above equals $X^T X_\Delta/(n-1)$.

For regularly sampled spatial data, i.e., ordinary digital image data, often a one-pixel horizontal shift, Δ_h, to obtain $x(r+\Delta_h)$ is used to estimate S_{Δ_h} and a one-pixel vertical shift, Δ_v, to obtain $x(r+\Delta_v)$ is used to estimate S_{Δ_v}. S_Δ is then a pool of the two. Alternatively, the two one-pixel shifts may be used to estimate $x_\Delta(r) = x(r) - [x(r+\Delta_h) + x(r+\Delta_v)]/2$. Since this differencing is not possible for samples located in the first row or column, such samples are removed from these calculations.

A. Primal Formulation

To maximize ρ in (15) we must minimize the Rayleigh quotient $a^T X_\Delta^T X_\Delta a/(a^T X^T X a)$ or maximize its inverse. This is done by solving a symmetric generalized eigenvalue problem, see Appendix.

Unlike linear PCA, the result from linear MAF analysis is scale invariant: if x_i is replaced by some matrix transformation Tx_i corresponding to replacing X by XT^T, the result is the same.

B. Dual Formulation and Kernelization

As in PCA described in Section II we get to the dual formulation by reparameterizing and setting $a \propto X^T b$

$$\rho = 1 - \frac{1}{2} \frac{b^T X X_\Delta^T X_\Delta X^T b}{b^T X X^T X X^T b}. \tag{16}$$

With Φ an n by $q(q \geq p))$ matrix

$$\Phi = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix} \tag{17}$$

with nonlinear mappings $\phi(x)$ into high-dimensional feature space of the rows in X, and Φ_Δ an n by q matrix with nonlinear mappings of the rows in X_Δ, this in the kernelized version becomes

$$\rho = 1 - \frac{1}{2} \frac{b^T \Phi \Phi_\Delta^T \Phi_\Delta \Phi^T b}{b^T \Phi \Phi^T \Phi \Phi^T b} \tag{18}$$

$$= 1 - \frac{1}{2} \frac{b^T K_\Delta K_\Delta^T b}{b^T K^2 b}. \tag{19}$$

Replacing $\Phi \Phi^T$—which has elements $\phi(x_i)^T \phi(x_j)$, i.e., we need not know the mapping ϕ explicitly, all we need are the inner products—with kernel matrix K with elements $\kappa(x_i, x_j)$ and $\Phi \Phi_\Delta^T$ with K_Δ is known as the kernel trick or kernel substitution. Nonsymmetric $K_\Delta = \Phi \Phi_\Delta^T$ here has elements $\kappa(x_i, x_{\Delta j})$, $i, j = 1, \ldots, n$. Both K and K_Δ are centered versions of the kernel matrices, i.e., the mean values of their columns are zero [8], [12], [13], [15], [16]. Since we want $1 = a^T X^T X a \propto b^T X X^T X X^T b \to b^T \Phi \Phi^T \Phi \Phi^T b = b^T K^2 b = 1$ in this case we have $a = \Phi^T b$ and $b = \Phi a$, i.e., unlike the case with kernel PCA in kernel MAF analysis we do not divide the dual eigenvector by the square root of the corresponding eigenvalue, [8], [12], [13], [15], [16].

The dual formulation linear MAF analysis inherits the scale invariance of the primal formulation problem. This is not the case for the nonlinear mapping $\phi(x) \to \phi(Tx) \neq T\phi(x)$ and, therefore, it is not the case for the kernel version either. However, for radial basis function (RBF) kernels, i.e., $\kappa(x_i, x_j) = \kappa(\|x_i - x_j\|)$, any transformation by an orthogonal matrix Q (i.e., a rotation) of the original coordinate system will not influence the result of the analysis, $(Qx_i)^T Qx_j = x_i^T Q^T Qx_j = x_i^T x_j$.

C. Regularization and Kernelization

Because of the versatility of the model in kernel CCA where we find two different sets of linear combinations (the eigenvectors), regularization there is always needed, [9], [12]. We see

that in kernel MAF analysis where we find one linear combination only, we need not regularize. Still we may wish to regularize and if so one of several possible alternative versions of the primal formulation is

$$\rho = 1 - \frac{1}{2} \left[\frac{a^T X^T X a}{a^T [(1-k) X_\Delta^T X_\Delta + k I_p] a} \right]^{-1} \tag{20}$$

which in the dual version becomes

$$\rho = 1 - \frac{1}{2} \left[\frac{b^T X X^T X X^T b}{b^T [(1-k) X X_\Delta^T X_\Delta X^T + k X X^T] b} \right]^{-1} \tag{21}$$

which in turn kernelizes to

$$\rho = 1 - \frac{1}{2} \left[\frac{b^T K^2 b}{b^T [(1-k) K_\Delta K_\Delta^T + k K] b} \right]^{-1}. \tag{22}$$

D. Implementation Issues

Here the kernel version of the symmetric generalized eigenvalue problem is solved by maximizing the inverse of the Rayleigh quotient in (19), see also the Appendix for a more thorough description. This may be done by writing symmetric $K_\Delta K_\Delta^T$ as a product of matrix square roots

$$K^2 b = \lambda K_\Delta K_\Delta^T b \tag{23}$$

$$= \lambda \left(K_\Delta K_\Delta^T \right)^{1/2} \left(K_\Delta K_\Delta^T \right)^{1/2} b. \tag{24}$$

The problem now rewrites to

$$\left(K_\Delta K_\Delta^T \right)^{-1/2} K^2 \left(K_\Delta K_\Delta^T \right)^{-1/2} \left[\left(K_\Delta K_\Delta^T \right)^{1/2} b \right]$$
$$= \lambda \left[\left(K_\Delta K_\Delta^T \right)^{1/2} b \right] \tag{25}$$

which is a symmetric ordinary eigenvalue problem. The Appendix shows how this may be done when $K_\Delta K_\Delta^T$ is not full rank.

A very different implementation issue is the following: K and K_Δ are n by n where n is the number of observations (pixels) which in image analysis can be prohibitively large. In this case, we may subsample the image and carry out the kernel eigenvalue analysis on these samples only. These samples are termed the training data. To obtain a transformed version of the entire image we then project all pixels, which we call the test data, mapped by ϕ onto the primal eigenvectors. Hence, we need to calculate K_{12} which is a centered version of n_1 mapped training observations in Φ_1 kernelized with $n_2 = n$ mapped test observations in Φ_2. The image will typically be too large to hold the kernel matrix K_{12} in computer memory and to carry out centering and kernelization of the test data with the training data in memory. Therefore, we may need to center and kernelize in smaller chunks at a time, for example row by row. Choosing to center the test data with the test data mean will force us to calculate K_{12} twice—once for centering and once for calculating the projections, see the following—resulting in increased execution time. Because of the necessary looping over rows this is especially true for vector and matrix oriented programming

environments such as Matlab and IDL. This part of the code is well suited for parallel implementation or for running on the GPU processor on a graphics card.

The subsampling potentially facilitates avoiding problems arising from the spatial autocorrelation inherent to image data.

E. Projections Onto Eigenvectors

To find the kernel MAFs from the generalized eigenvalue problem in (19) or a regularized version, we project a mapped x onto the primal eigenvector a_i

$$\phi(x)^T a_i = \phi(x)^T \Phi^T b_i \tag{26}$$
$$= \begin{bmatrix} \kappa(x, x_1) & \kappa(x, x_2) & \cdots & \kappa(x, x_n) \end{bmatrix} b_i \tag{27}$$

or in matrix notation for the training observations $\Phi A = KB$ (A is a matrix with a_i in the columns and B is a matrix with b_i in the columns). The variance of this projection is

$$\mathrm{Var}\left\{ a_i^T \phi(x) \right\} = a_i^T \Phi^T \Phi a_i / (n-1) \tag{28}$$
$$= b_i^T \Phi \Phi^T \Phi \Phi^T b_i / (n-1) \tag{29}$$
$$= b_i^T K^2 b_i / (n-1) \tag{30}$$
$$= 1/(n-1). \tag{31}$$

IV. MNF ANALYSIS

In MNF analysis first suggested in [7], we minimize the noise fraction (NF) or equivalently maximize the signal-to-noise ratio (SNR) of linear combinations, $a^T x(r)$, of zeroth-mean original (spatial) variables, $x(r)$. NR and SNR will be defined shortly. Here we write the total $x(r)$ as a sum of a signal part, $x_S(r)$, and a noise part, $x_N(r)$

$$x(r) = x_S(r) + x_N(r). \tag{32}$$

These two parts are considered to be uncorrelated. Therefore, the variance-covariance matrix S of x may be written as a sum of the signal and noise dispersions

$$S = S_S + S_N. \tag{33}$$

Assuming or imposing second-order stationarity of $x_S(r)$ and $x_N(r)$, S_S and S_N are independent of location, r.

The noise fraction NF is here defined as the ratio of the variance of the noise and the variance of the total, so for a linear combination $a^T x(r)$ of zeroth-mean $x(r)$ we get

$$\mathrm{NF} = \frac{a^T S_N a}{a^T S a}. \tag{34}$$

Similarly, for the SNR, which is here defined as the ratio of the variance of the signal and the variance of the noise, we get

$$\mathrm{SNR} = \frac{a^T S_S a}{a^T S_N a}. \tag{35}$$

This gives the relation $\mathrm{NF} = 1/(\mathrm{SNR}+1)$ or $\mathrm{SNR} = 1/\mathrm{NF}-1$, i.e., by maximizing SNR we minimize NF.

In [4] it is argued that if the signal is highly correlated and the noise is uncorrelated we have $S_N \simeq S_\Delta / 2$, see Section III.

For regularly sampled spatial data, i.e., ordinary digital image data, often the difference between the actual value at a location and the mean of the values in a 3×3 neighborhood, or the residual from a local regression in a 3×3 neighborhood to a plane or a paraboloid is used to estimate $x_N(r)$. This may be formulated as a filtering problem [27]. Since this filtering is not possible for samples located in the first and last row or column, such samples are removed from these calculations.

A. Primal Formulation

We wish to maximize

$$\frac{1}{\mathrm{NF}} = \frac{a^T S a}{a^T S_N a} \tag{36}$$
$$= \frac{a^T X^T X a}{a^T X_N^T X_N a} \tag{37}$$

where X is the n by p data matrix and X_N is a similarly defined matrix for $x_N(r)$ with zeroth-mean columns.

B. Dual Formulation and Kernelization

As in kernel MAF analysis described in Section III we get to the dual formulation by reparameterizing and setting $a \propto X^T b$

$$\frac{1}{\mathrm{NF}} = \frac{b^T X X^T X X^T b}{b^T X X_N^T X_N X^T b} \tag{38}$$

which in the kernelized version becomes

$$\frac{1}{\mathrm{NF}} = \frac{b^T \Phi \Phi^T \Phi \Phi^T b}{b^T \Phi \Phi_N^T \Phi_N \Phi^T b} \tag{39}$$
$$= \frac{b^T K^2 b}{b^T K_N K_N^T b} \tag{40}$$

where Φ_N is an n by $q(q \geq p)$ matrix with mappings of X_N and nonsymmetric $K_N = \Phi \Phi_N^T$ has elements $\kappa(x_i, x_{Nj})$, $i,j = 1, \ldots, n$. We see that this is completely equivalent to the previously MAF analysis with X_N here playing the role of X_Δ there.

V. DATA

To illustrate the techniques we give three cases. The first two are change detection examples, one is based upon RGB data from the DLR 3K camera system [28], [29], and another is based upon hyperspectral data from the HyMap scanner [30]. Because of the nature of the data in the first case we have a good impression of the "truth," see the following. The case on HyMap data is chosen because the dual formulation inherent to kernel methods as described here is well suited for hypervariate data. The third case shows an application to maize kernel inspection.

A. DLR 3K Camera Data

The images used were recorded with the airborne DLR 3K camera system [28], [29] from the German Aerospace Center, DLR. This system consists of three commercially available 16 megapixel cameras arranged on a mount and a navigation unit with which it is possible to record time series of images covering large areas at frequencies up to 3 Hz. The 1000 rows by 1000 columns example images acquired 0.7 s apart cover a busy motorway. Fig. 1 shows the RGB images at time points one (a) and two (b). The data at the two time points were orthoprojected using

(a)

(b)

Fig. 1. DLR 3K camera images as RGB acquired 0.7 s apart; note the movements of the cars on the motorway. (a) Time point one. (b) Time point two.

GPS/IMU measurements and a DEM. For flat terrain like here one pixel accuracy was obtained. In these data, the change occurring between the two time points will be dominated by the movement of the cars on the motorway. Undesired, apparent change will occur due to the movement of the aircraft and the different viewing positions at the two time points, see also and [17].

B. Hyperspectral HyMap Data

In this example we use all spectral bands of 400 rows by 270 columns 5-m pixels HyMap [30] data covering a small agri-

Fig. 2. DLR 3K camera simple difference image as RGB.

cultural area near Lake Waging-Taching in Bavaria, Germany. HyMap is an airborne, hyperspectral instrument which records 126 spectral bands covering most of the wavelength region from 438 to 2483 nm with 15–20 nm spacing. Fig. 7 shows HyMap bands 27 (828 nm), 81 (1 648 nm) and 16 (662 nm) as RGB, 30 June 2003 8:43 UTC (a) and 4 August 2003 10:23 UTC (b). The data at the two time points were radiometrically calibrated and orthorectified using GPS/IMU measurements, a DEM and ground control points. One pixel accuracy was obtained. These data are dealt with in [18], [31], and [32].

C. Maize Kernel Data

Here we use a line scan camera to obtain 149 rows by 370 columns 153 band (900–1 700 nm) hyperspectral images of maize samples comprised of the front and back side of eight kernels on a black background acquired as two separate images appended into one. The kernels are not fresh from harvest and, hence, they have a very low water content and in addition they are free from any infections. Many cereals in general share the same compounds and basic structure. In our case of maize, a single kernel can be divided into many different constituents on the macroscopic level. In general, the structural components of cereals can be divided into three classes denoted "Endosperm," "Germ," and "Pedicel," see Fig. 10. These components have different functions and compounds leading to different spectral profiles. This is described in more detail in [19].

VI. RESULTS AND DISCUSSION

To be able to carry out kernel PCA and MAF/MNF analysis on the large amount of pixels typically present in Earth observation and other image data, we subsample the image and use a small portion termed the training data only. We use typically in the order 10^3 randomly sampled training pixels to find the eigenvectors onto which we then project the entire image termed
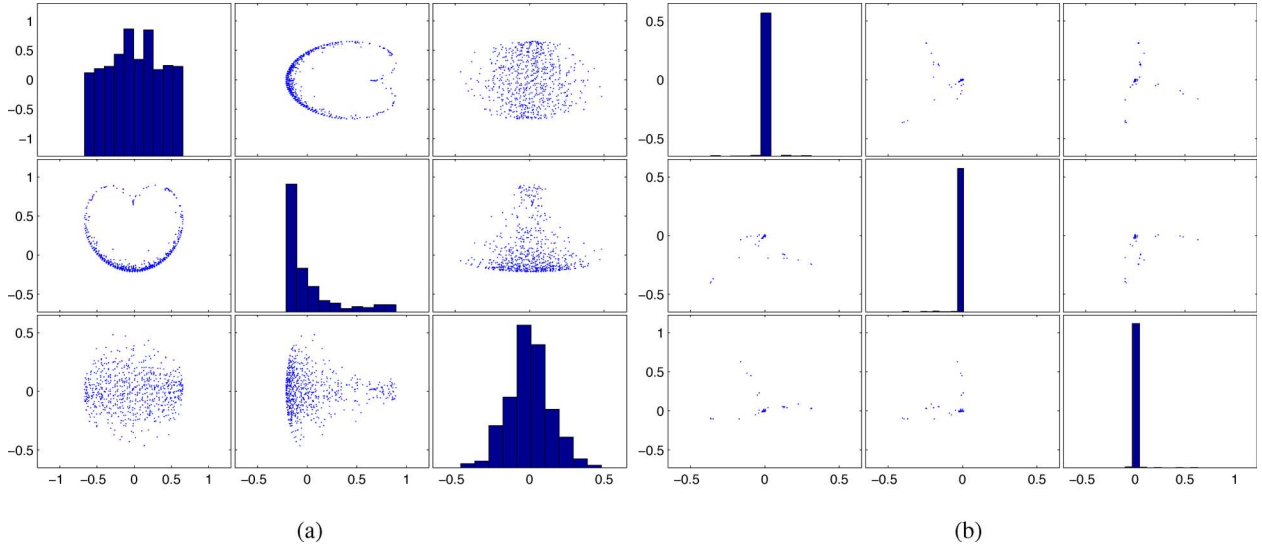
Fig. 3. Scatterplots and histograms of (a) the first three kernel PCs and (b) kernel MAFs.

the test data kernelized with the training data. With this subsampling we potentially avoid problems that may arise from the spatial autocorrelation inherent to image data. A Gaussian kernel $\kappa(x_i, x_j) = \exp(-(1)/(2)\|x_i - x_j\|^2/\sigma^2)$ with σ equal to the mean distance between the observations in feature space is used.

A. DLR 3K Data

Since there are only three spectral bands in this example we can see all simple difference images between acquisitions at time points two and one simultaneously. Fig. 2 shows an RGB image of these simple differences stretched linearly from mean minus to mean plus three standard deviations. Because of the 0.7 s between acquisitions the change detected is dominated by the movements of the cars on the motorway. Also, the movement of the airplane gives rise to edge effects where height differences on the ground occur. We see that the no-change background appears quite noisy.

Fig. 3 shows scatter plots and histograms of the first three kernel PCs (a) and kernel MAFs (b) for the ∼1000 training samples. We see that the histograms for the kernel MAFs are very narrow and that many more samples are concentrated in the center of the scatter plots for the kernel MAFs. This reflects the fact that no-change pixels have values very close to zero and change pixels are very different from zero. Thus, the kernel MAFs give a much better discrimination between change and no-change. Also, the scatter plots for kernel MAF as opposed to those for kernel PCA much like independent components show linear structures indicating a better separation of interesting structure in the kernel MAF components.

Fig. 4 shows kernel principal components 1–3 (a) and kernel MAFs 1–3 (b). Fig. 5 shows kernel MNFs 1–3 of simple band-by-band difference images as RGB, with noise estimated as the difference from the mean in a 3×3 window (a) and kernel MNFs 1–3 with noise estimated as the residual from a quadratic surface in a 3×3 window (b). The three bands are stretched linearly between mean minus and plus three standard deviations. In this representation, no-change areas will appear as grayish and change regions will appear in saturated colors

(including black and white). From the images, we see that all methods detect the change represented by the moving cars and that all cars stand out clearly as being detected. Ordinary linear PCA or MAF analysis (not shown) does not give this beautiful discrimination between change and no-change regions. The most conspicuous difference between kPCA on the one hand and the kMAF/kMNF methods on the other hand is the much less noisy no-change background offered by the kernel MAF/MNF transformations corresponding to a lower rate of false detection for these methods. Also, kMAF and the two different noise models for kMNF give rise to different coloring of the change pixels reflecting the fact that different kernel components contain different change information. It is, however, the same change pixels that get the extreme coloring irrespective of the noise model used.

In this case a more relevant quantitative evaluation of the change detection methods than the often applied comparison with manually labelled "ground truth," is a comparison of background noise suppression with good noise suppression corresponding to fewer false alarms. Kernel MAF/MNF analyses suppress the noisy no-change background much more successfully than ordinary PCA, ordinary MAF and kernel PCA. The ratio between variances of ordinary PC 1 and kernel MAF 1 (both scaled to unit variance) calculated in the 100×100 pixels top-right no-change region of the images is 85.7 corresponding to 19.3 dB. For ordinary MAF 1 and kernel MAF 1 the ratio is 271.9 corresponding to 24.3 dB, and for kernel PC 1 and kernel MAF 1 376.5 corresponding to 25.7 dB.

The eigenvalue based model SNR $= 2a^T X^T X a/ (a^T X_\Delta X_\Delta^T a) - 1$ for the first three ordinary MAFs are 14.24, 9.30, and 5.83 corresponding to 11.5 dB, 9.7 dB, and 7.7 dB. For the first three kernel MAFs the corresponding numbers for SNR $= 2 b^T K^2 b/(b^T K_\Delta K_\Delta^T b) - 1$ are 85.0 dB, 75.4 dB, and 72.5 dB, all drastic improvements. For kernel MNF analysis with the noise estimated as the difference between the center pixel and the mean in a 3×3 window yielding similar change results, the corresponding numbers for SNR $= b^T K^2 b/(b^T K_N K_N^T b) - 1$ are 85.5 dB, 82.0 dB, and
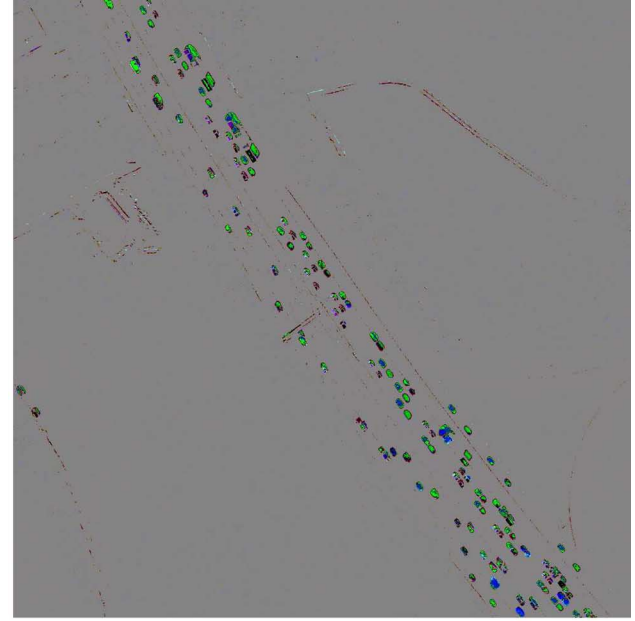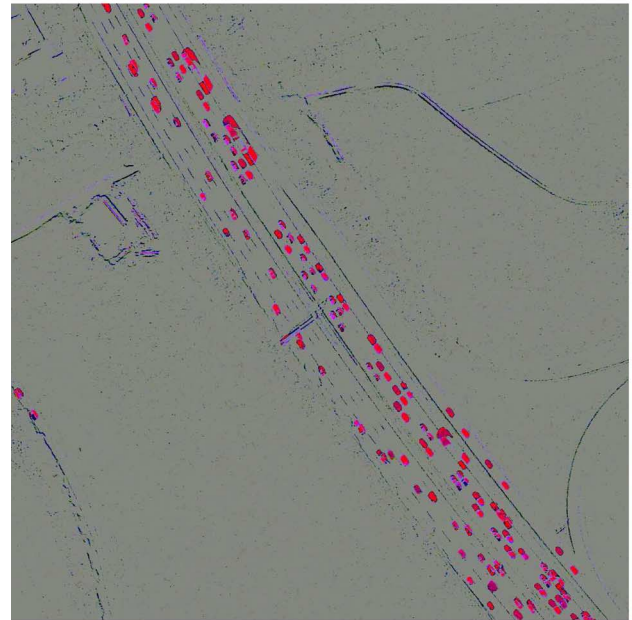
(a)

(a)

(b)
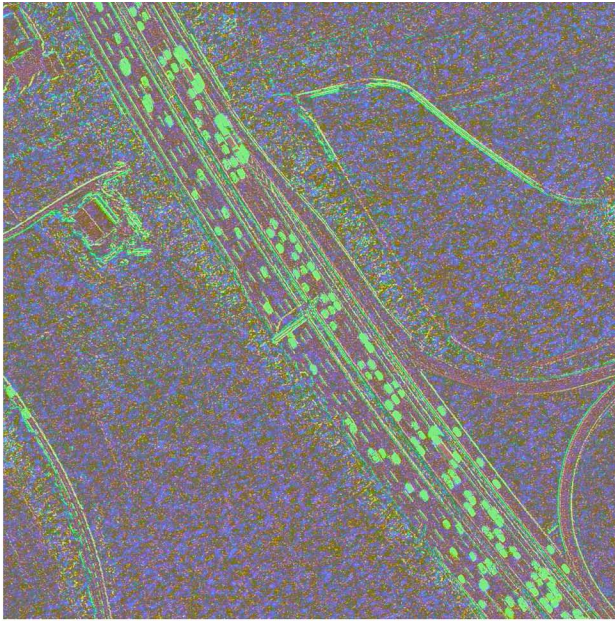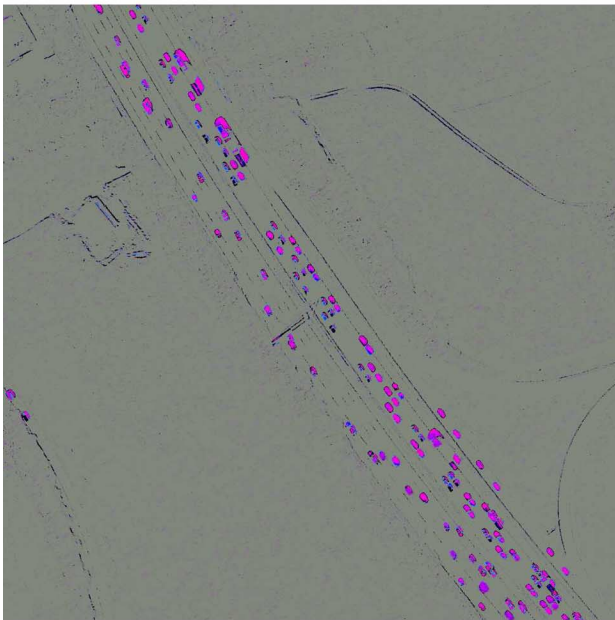
(b)

Fig. 4. (a) Kernel principal components 1–3 and (b) kernel MAFs 1–3 of the three simple difference images as RGB. All bands are stretched linearly between mean (which is zero) minus and plus three standard deviations.

Fig. 5. (a) Kernel MNFs 1–3 with noise estimated as the difference from the mean in a 3×3 window and (b) kernel MNFs 1–3 with noise estimated as the residual from a quadratic surface in a 3×3 window of the three simple difference images as RGB. All bands are stretched linearly between mean (which is zero) minus and plus three standard deviations.

74.5 dB. For kernel MNF analysis with the noise estimated as the residual from fitting a quadratic surface in a 3×3 window again yielding similar change results with a slightly more noisy visual appearance of the background, the corresponding numbers are 100.6 dB, 87.9 dB, and 86.7 dB.

For the training data alone the kernel MAF/MNFs are uncorrelated. The correlation matrix for the entire image for the first three kernel MAFs is shown in Table I for the DLR 3K data (left). Although not zero, the correlations between different kernel MAF modes are still fairly small.

TABLE I
CORRELATION MATRICES FOR ENTIRE IMAGE FOR FIRST THREE KERNEL MAFS
FOR THE DLR 3K CASE (LEFT) AND THE HYMAP CASE (RIGHT)

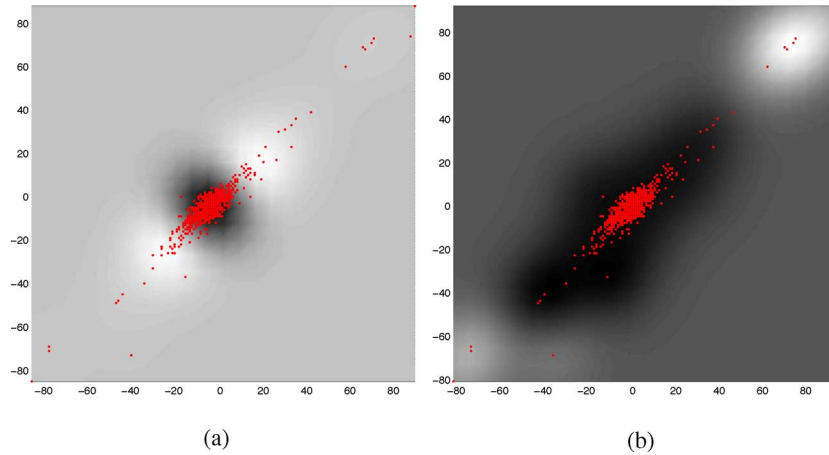| 1.0000	−0.1411	−0.1800	1.0000	0.0989	−0.1517
−0.1411	1.0000	0.1615	0.0989	1.0000	−0.2193
−0.1800	0.1615	1.0000	−0.1517	−0.2193	1.0000

(a)   (b)

Fig. 6. Scatter plots for training data of red and green bands differences at the two time points on backgrounds of contours for (a) kernel PC 2 and (b) kernel MAF 1, DLR 3K data.

(a)   (b)

Fig. 7. HyMap bands 27 (828 nm), 81 (1 648 nm) and 16 (662 nm) as RGB. (a) 30 June 2003 8:43 UTC. (b) 4 August 2003 10:23 UTC.

The generation of three kernel MAF/MNFs for the entire image based upon ∼1 000 random training samples calculated by Matlab code based upon the `eigs` function with a simple for-loop-over-rows implementation for the test data takes around 7.9 min on a 32-b, 2.00 GHz Intel Core 2 CPU laptop with 2.00 GB, 998 MHz memory.

Two-Band Example: Let us redo the analysis, this time skipping the blue band leaving two spectral bands, red and green.

This leads to change patterns very similar to the ones described in the three-band case mentioned previously and more importantly it gives us the possibility to produce a scatter plot of the simple differences for red versus simple differences for green for the training samples on the background of contours for the individual kernel MAF and PC variates. Fig. 6 shows examples of such plots where no-change pixels are located near the origo and extreme change pixels successfully identified by especially
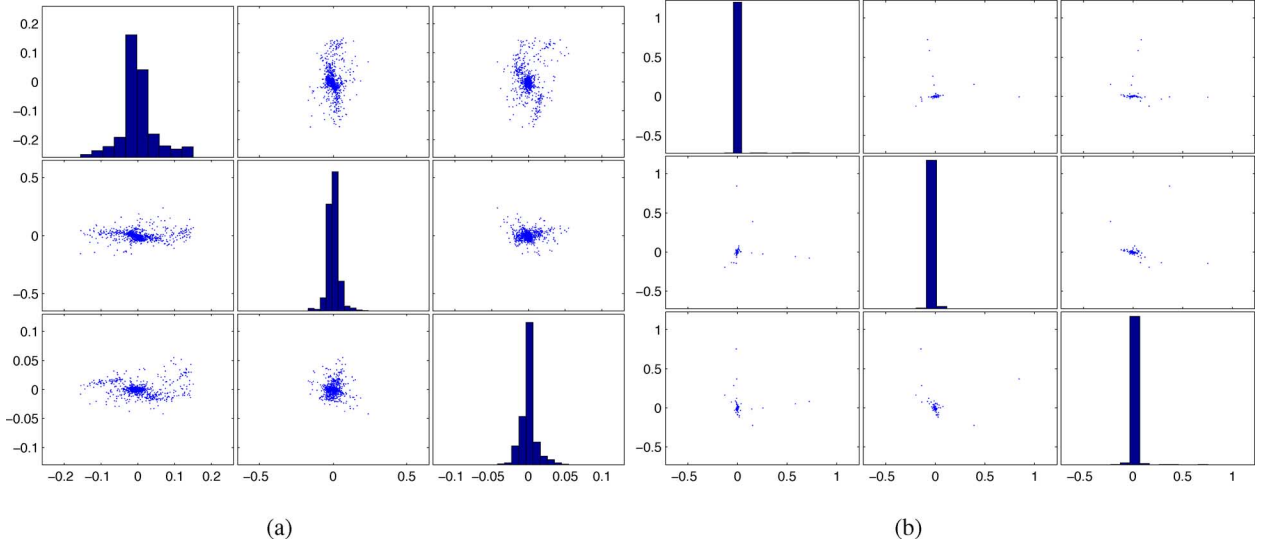
Fig. 8. Scatterplots and histograms of the (a) first three kernel PCs and (b) kernel MAFs, HyMap data.

kMAF analysis are located towards the end of the "line," i.e., the elongated point cloud in the scatter plot. Note, that the value of the kMAF variate 1 is very small (dark graytones) for the no-change observations and very large (bright graytones) for the extreme change pixels, and that the sign of kMAF variate 1 is the same for values around $(-70, -70)$ and $(70, 70)$ in the scatter plot showing that kMAF 1 is an excellent change/no-change discriminator. This effect is not obtained here with kernel PCA and can never be obtained with a linear technique.

B. HyMap Data

In this example all band-wise differences of the 126 spectral bands of the HyMap are used. Fig. 8 shows scatter plots and histograms of the first three kernel PCs (a) and kernel MAFs (b) for the ∼1000 training samples. As in the previous case, we see that the histograms for the kernel MAFs are very narrow and that many more samples are concentrated in the center of the scatter plots for the kernel MAFs. This reflects the fact that no-change pixels have values very close to zero and change pixels are very different from zero. Thus, the kernel MAFs give a much better discrimination between change and no-change.

Fig. 9 shows kernel principal components 1–3 (a) and kernel MAFs 1–3 (b) of simple band-by-band difference images as RGB. All bands are stretched linearly between mean minus and plus three standard deviations. In this representation, no-change areas will appear as grayish and change regions will appear in saturated colors (including black and white). The change detected over the five weeks is due to growth of the main crop types such as maize, barley and wheat. On pastures, which are constantly being grazed, in forest stands and in the lake to the south, no change is observed. Furthermore, both solar elevation and azimuth have changed which gives rise to edge effects where abrupt height differences on the ground occur.

We see that both types of kernel analysis emphasize change and that unlike kernel PCA, kernel MAF analysis seems to focus on the most conspicuous changes and that it gives a very strong discrimination between change and no-change regions.

The eigenvalue based model SNR $= 2a^T X^T X a/(a^T X_\Delta X_\Delta^T a) - 1$ for the first three ordinary MAFs are 40.70, 33.00, and 22.58 corresponding to 16.1 dB, 15.2 dB and 13.5 dB. For the first three kernel MAFs the corresponding numbers for SNR $= 2 \; b^T K^2 b/(b^T K_\Delta K_\Delta^T b) - 1$ are 78.5 dB, 74.1 dB and 70.3 dB, as in the previous case all drastic improvements. For kernel MNF analysis with the noise estimated as the difference between the center pixel and the mean in a 3×3 window yielding similar change results with a slightly more noisy appearance (not shown), the corresponding numbers for SNR $= b^T K^2 b/(b^T K_N K_N^T b) - 1$ are 81.2 dB, 77.3 dB, and 75.9 dB. For kernel MNF analysis with the noise estimated as the residual from fitting a quadratic surface in a 3×3 window again yielding similar change results with a slightly more noisy appearance (not shown), the corresponding numbers are 93.9 dB, 90.5 dB, and 85.0 dB.

As in the previous example ordinary linear PCA or MAF analysis (not shown) does not give this beautiful discrimination between change and no-change regions.

For the training data alone the kernel MAF/MNFs are uncorrelated. The correlation matrix for the entire image for the first three kernel MAFs is shown in Table I for the HyMap data (right). Although not zero, the correlations between different kernel MAF modes are still fairly small.

The generation of three kernel MAF/MNFs for the entire image based upon ∼1000 random training samples calculated by Matlab code based upon the `eigs` function with a simple for-loop-over-rows implementation for the test data takes around 80 s on a 32-b, 2.00 GHz Intel Core 2 CPU laptop with 2.00 GB, 998 MHz memory.

C. Maize Kernel Data

The 153 band (900–1 700 nm, i.e., the near infrared region) hyperspectral images of the maize samples consist of the front and back sides of the kernels on a black background in two separate images appended into one image, the front in the left half image and the back in the right half image.
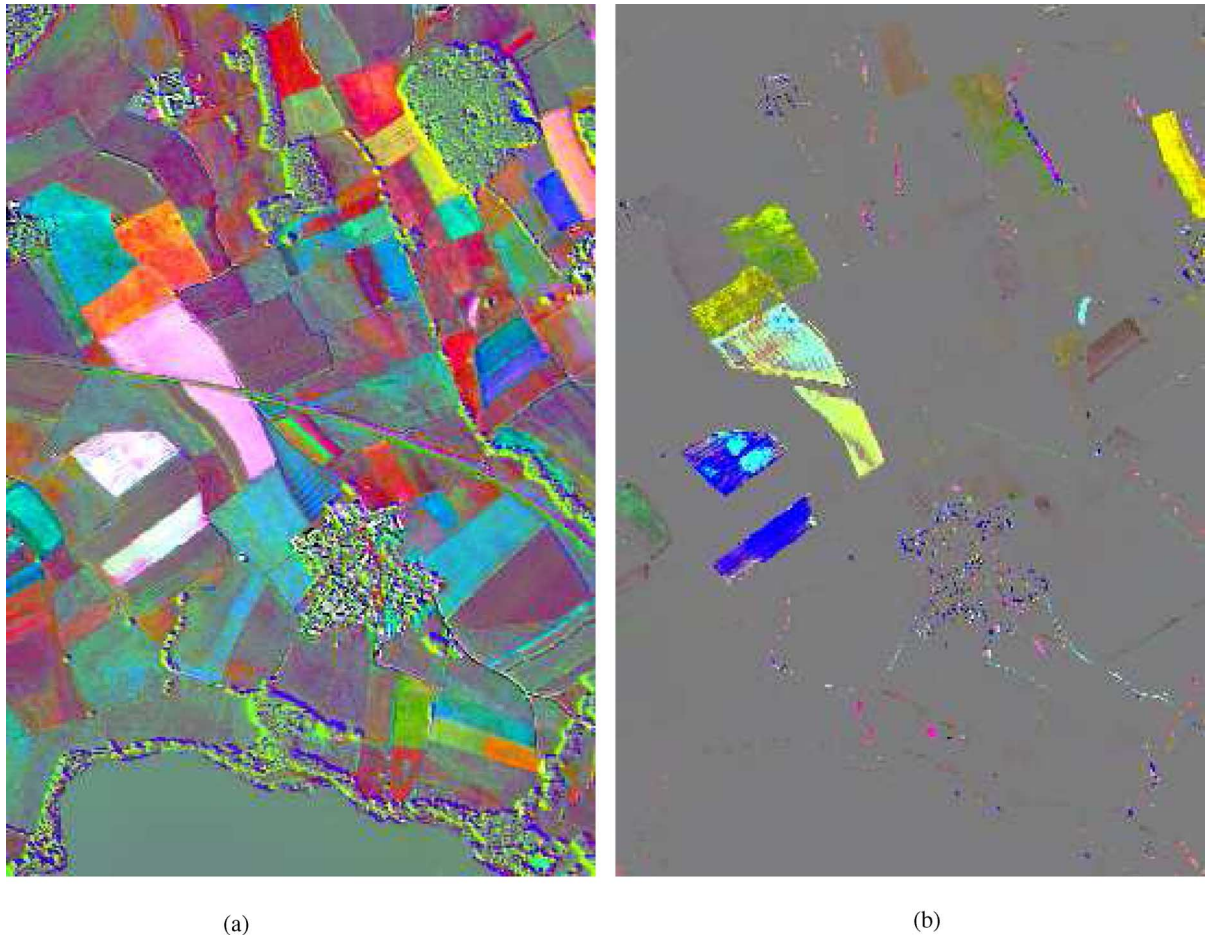
Fig. 9. (a) Kernel principal components 1–3 and (b) kernel MAFs 1–3 of 126 simple difference images as RGB. All bands are stretched linearly between mean (which is zero) minus and plus three standard deviations.
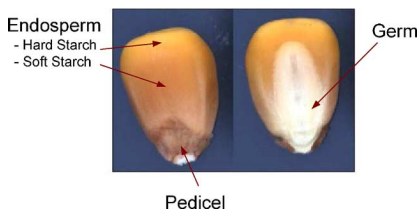

Fig. 10. Maize kernel constituents, front (left) and back side (right), the color image is constructed as an RGB combination of NIR bands 150, 75, and 1.

In this case we calculate kernel components based upon $\sim$3 000 training samples. Fig. 11 shows linear principal components and linear MAFs of front (left half image) and back sides (right half image) of eight maize kernels.

Fig. 12 shows kernel principal components and kernel MAFs of front (left half image) and back sides (right half image) of eight maize kernels. Note the superior ability of the kernel MAF variates to adapt to even very abrupt changes such as image edges (between the front and back images) and shadows in a hyperspectral background and to label different maize kernel structural components with the same color.
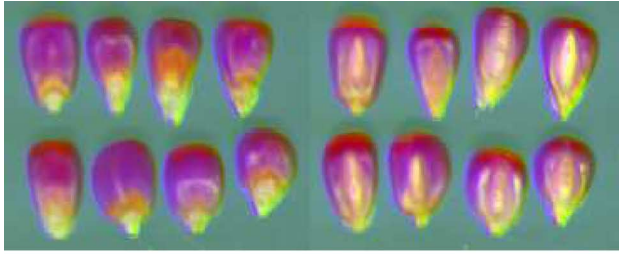
Autocorrelations for the first three linear MAFs are 0.9940, 0.9842 and 0.9685. Autocorrelations for the first three kernel MAFs are all 1 to seven decimal places, all higher than achieved by the linear analysis.

The generation of six kernel MAFs for the entire combined front and back image based upon $\sim$3 000 random training samples calculated by Matlab code based upon the `eigs` function with a simple for-loop-over-rows implementation for the test data takes around 11.5 min on a 32-b, 2.00 GHz Intel Core 2 CPU laptop with 2.00 GB, 998 MHz memory.
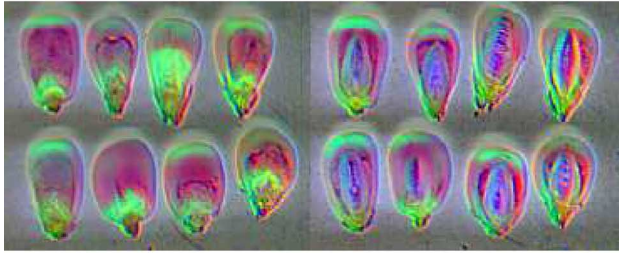
VII. CONCLUSIONS

In the dual formulation of PCA and MAF/MNF analyses the data enter into the problem as inner products between the observations. These inner products may be replaced by inner products between mappings of the measured variables into higher order feature space. The idea in kernel orthogonalization is to express the inner products between the mappings in terms of a kernel function to avoid the explicit specification of the mappings. Both the eigenvalue problem, the centering to zero mean and the projections onto eigenvectors to find kernel scores may be expressed by means of the kernel function. Kernel orthogonalization methods handle nonlinearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space.
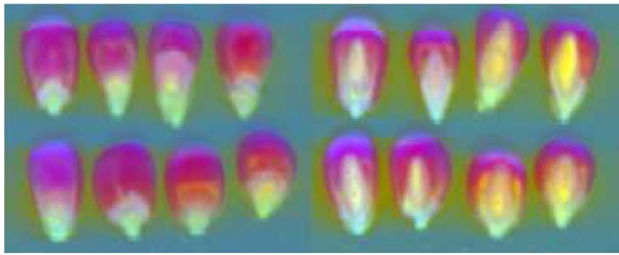
In all three examples given, kernel orthogonalization with a Gaussian kernel $\kappa(x_i, x_j) = \exp(-1/2\|x_i - x_j\|^2/\sigma^2)$ is used. Generally, the scale parameter σ should be chosen equal to a typical distance between observations in feature space and
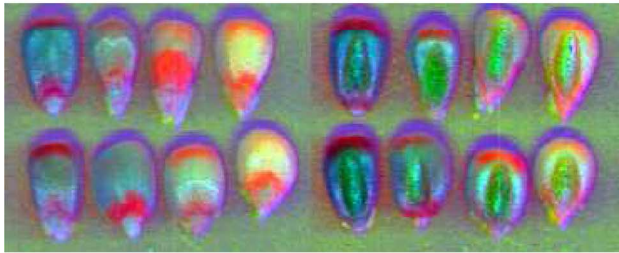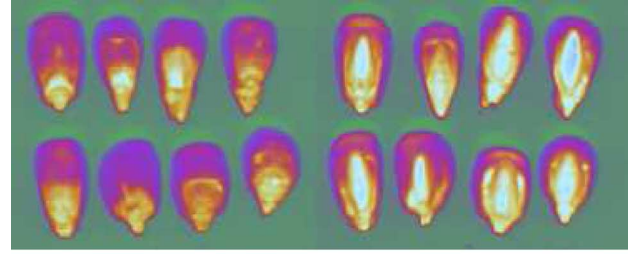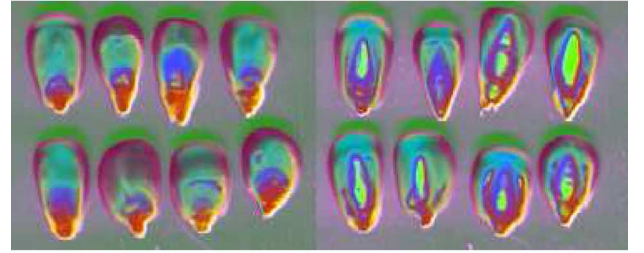
Fig. 11.   Linear principal components of front (left half image) and back sides (right half image) of eight maize kernels shown as RGB combination of factors (1, 2, 3) and (4, 5, 6) [(a) and (b)], and corresponding linear MAFs [(c) and (d)]. (a) PC1, PC2, and PC3. (b) PC4, PC5, and PC6. (c) MAF1, MAF2, and MAF3. (d) MAF4, MAF5, and MAF6.

Fig. 12.   Kernel principal components of front (left half image) and back sides (right half image) of eight maize kernel shown as RGB combination of factors (1, 2, 3) and (4, 5, 6) [(a) and (b)], and corresponding kernel MAFs [(c) and (d)]. (a) kPC1, kPC2, and kPC3. (b) kPC4, kPC5, and kPC6. (c) kMAF1, kMAF2, and kMAF3. (d) kMAF4, kMAF5, and kMAF6.

smaller than the largest distance. Here it is chosen simply as the mean value of the relevant distances. The sensitivity of the kernel methods to the scale parameter σ, image noise, and the subsampling applied are all good subjects for future work.
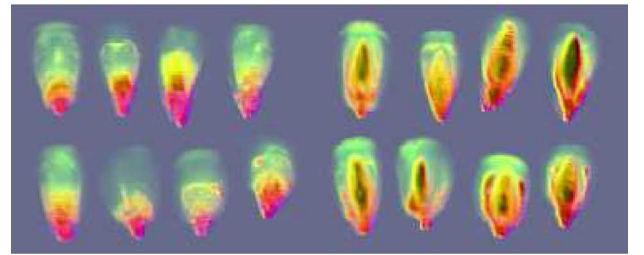
In two of three examples given, kernel orthogonalization with a Gaussian kernel is used for detecting change in coregistered, calibrated simple band-by-band difference DLR 3K camera and hyperspectral HyMap images. Unlike ordinary linear PCA or MAF/MNF analyses, especially kernel MAF/MNF analysis gives a strong discrimination between change and no-change regions with a conspicuously better suppression of both noise
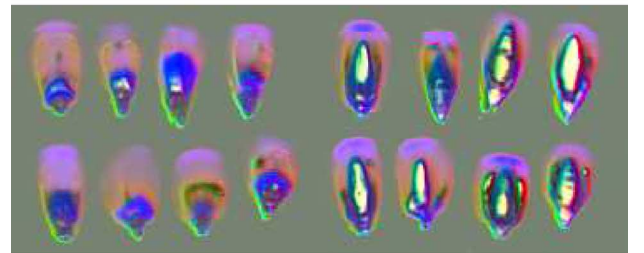
and signal in the no-change background. The simple differencing is meaningful for calibrated or normalized data only. If the data available is not of this nature, generalized differences as described in [32]–[35] may be applied.

In the DLR 3K camera case the ratio between variances in a known no-change region between leading components of ordinary PCA, ordinary MAF analysis and kernel PCA on the one hand and kernel MAF analysis on the other, drastically favor kernel MAF analysis. The ratios range from 19.3 dB to 25.7 dB.

Also the eigenvalue based model SNRs improve drastically going from linear MAF to kernel MAF analysis. For the first

three MAF variates in the DLR 3K camera case the SNR improves from 11.5 dB to 85.0 dB, 9.7 dB to 75.4 dB, and 7.7 dB to 72.5 dB, respectively. Corresponding numbers for the hyperspectral HyMap case are from 16.1 dB to 78.5 dB, 15.2 dB to 74.1 dB, and 13.5 dB to 70.3 dB. Using kernel MNF analysis these model SNRs improve further but the visual impression of the change results (not shown for the HyMap case) are not necessarily better than for kernel MAF.

The maize kernel example shows the superior ability of the kernel MAF variates to adapt to shadow, to even very abrupt changes such as image edges in a hyperspectral background and to label different maize kernel parts with the same color.

All three examples indicate that generally, the kernel MAF/MNF variates seem to possess the ability to adapt to even abruptly varying multi- or hypervariate backgrounds and focus on extreme observations.

Kernel PCA and kernel MAF/MNF analysis are so-called memory-based methods: where ordinary PCA and MAF/MNF analysis handle new observations by projecting them onto the eigenvectors found based upon the training data, because of the kernelization of the new observations with the training observations, kernel PCA and MAF/MNF analysis need the original data as well as the eigenvectors (and for PCA the eigenvalues) to handle new data.

The kernel MAF and MNF transformations are expected to be very useful in many other application areas for example in medical, industrial and astronomical image analysis involving both panchromatic and univariate data.

Matlab code to carry out the analyses described may be found on the author's homepage. IDL/ENVI and Python versions may be found on Dr. Morton J. Canty's homepage.

APPENDIX

Here we want to maximize the Rayleigh quotient

$$\lambda(w) = \frac{w^T A w}{w^T B w} \qquad (41)$$

for both A and B n by n, symmetric and positive semidefinite. In this case, both the numerator and the denominator are nonnegative. Also, λ is unchanged by a rescaling of w. To maximize λ we differentiate with respect to w

$$\frac{\partial \lambda}{\partial w} = 2 \frac{w^T B w\, A w - w^T A w\, B w}{(w^T B w)^2}. \qquad (42)$$

Setting this gradient to zero we get the symmetric generalized eigenvalue problem

$$A w = \frac{w^T A w}{w^T B w} B w = \lambda B w. \qquad (43)$$

The second derivative of λ or the Hessian, $H(\lambda) = \partial^2 \lambda / (\partial w \partial w^T)$, is

$$H = 2 \frac{(w^T B w\, A + 2 A w\, w^T B)(w^T B w)^2}{(w^T B w)^4}$$
$$- 2 \frac{(2 B w\, w^T A + w^T A w\, B)(w^T B w)^2}{(w^T B w)^4}$$
$$- 2 \frac{(w^T B w\, A w - w^T A w\, B w) 2 w^T B w\, 2 w^T B}{(w^T B w)^4}. \qquad (44)$$

We have $A w\, w^T B = (B w\, w^T A)^T$ but the two terms are not symmetric so they do not cancel. However, for the solution in (43), i.e., at the stationary points where the gradient is zero, the two terms do cancel, the last line above is zero, and we get

$$H = 2 \frac{A - \lambda B}{w^T B w} \qquad (45)$$

at these points. For B full rank we have the following: For the largest eigenvalue λ_1, $A - \lambda_1 B$ has nonpositive eigenvalues, i.e., H is negative semidefinite corresponding to a maximum for the Rayleigh quotient. For the smallest eigenvalue λ_n, $A - \lambda_n B$ has nonnegative eigenvalues, i.e., H is positive semidefinite corresponding to a minimum for the Rayleigh quotient. The eigenvalues for $A - \lambda_i B$, $i = 2, \ldots, n - 1$ are both negative, zero and positive, these correspond to saddle points.

The symmetric generalized eigenvalue problem may be solved by writing symmetric B as a product of matrix square roots

$$A w = \lambda B w \qquad (46)$$
$$= \lambda B^{1/2} B^{1/2} w \qquad (47)$$

where $B^{1/2} = P \Lambda^{1/2} P^T$ with P consisting of columns of eigenvectors and $\Lambda^{1/2}$ is a diagonal matrix of square roots of the eigenvalues of B. If B is full rank $r = n$ we retain all columns and all rows of both P and Λ. If B has rank $r < n$ we retain only the first r columns corresponding to the highest eigenvalues (but all rows) of P and only the r first rows and r first columns of Λ. Since $P^T P = I_r$ (and $P P^T = I_n$), this leads to the desired $B = P \Lambda^{1/2} P^T P \Lambda^{1/2} P^T = P \Lambda P^T$. The problem now rewrites to

$$(B^{-1/2} A B^{-1/2})(B^{1/2} w) = \lambda (B^{1/2} w) \qquad (48)$$

which is a symmetric ordinary eigenvalue problem. In this case, we may get the inverse for $B^{1/2}$ as $B^{-1/2} = (P \Lambda^{1/2} P^T)^{-1} = P \Lambda^{-1/2} P^T$ where $\Lambda^{-1/2}$ is an r by r diagonal matrix of inverse square roots of the eigenvalues.

Alternatively, we may maximize the Rayleigh quotient in (41) by multiplying from the left with the inverse B^{-1} or if needed the Moore-Penrose inverse B^-. In the latter case, we get

$$B^- A w = \lambda w \qquad (49)$$

which is a nonsymmetric ordinary eigenvalue problem. If B has rank $r < n$, its Moore-Penrose inverse is $B^- = (P \Lambda P^T)^{-1} = P \Lambda^{-1} P^T$ with Λ an r by r diagonal matrix of the nonzero eigenvalues of B ordered decreasingly and n by r matrix P consisting of columns of the corresponding eigenvectors. Λ^{-1} is an r by r diagonal matrix of inverse eigenvalues.

Following one of these lines of attack the eigenvectors are renormed so that the denominator of the Rayleigh quotient in (41) equals one, $w^T B w = 1$. Other natural renorming schemes are $w^T A w = 1$ (so that the numerator of the Rayleigh quotient equals one) or $w^T w = 1$ (to conserve Euclidean distance in the transformed space).

REFERENCES

[1] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philos. Mag.*, vol. 6, no. 2, pp. 559–572, 1901.

[2] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educat. Psych.*, vol. 24, pp. 417–441, 498–520, 1933.

[3] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002.

[4] P. Switzer and A. A. Green, Min/Max Autocorrelation Factors for Multivariate Spatial Imagery Dept. Stat., Stanford Univ., Stanford, CA, Tech. Rep. 6, 1984.

[5] P. Switzer and S. E. Ingebritsen, "Ordering of time-difference data from multispectral imagery," *Remote Sens. Environ.*, vol. 20, pp. 85–94, 1986.

[6] P. Gong, "Change detection using principal component analysis and fuzzy set theory," *Can. J. Remote Sens.*, vol. 19, no. 1, pp. 22–29, 1993.

[7] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-26, no. 1, pp. 65–74, Jan. 1988.

[8] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[9] P. L. Lai and C. Fyfe, "Kernel and nonlinear canonical correlation analysis," *Int. J. Neural Syst.*, vol. 10, no. 5, pp. 365–377, 2000.

[10] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, 2002.

[11] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.

[12] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[14] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[15] A. A. Nielsen and M. J. Canty, "Kernel principal component analysis for change detection," in *Proc. SPIE Eur. Remote Sens. Conf.*, Cardiff, U.K., Sep. 15–19, 2008, vol. 7109A, pp. 71090T-1–71090T-10.

[16] A. A. Nielsen, "A kernel version of maximum autocorrelation factor analysis for irregularly spaced data," *Mathematical Geosciences*, 2010, submitted for publication.

[17] A. A. Nielsen and M. J. Canty, "Kernel principal component and maximum autocorrelation factor analyses for change detection," in *Proc. SPIE Eur. Remote Sens. Conf.*, Berlin, Germany, 31 Aug.–3 Sep. 2009, vol. 7477, pp. 74770T-1–74770T-6.

[18] A. A. Nielsen, "Kernel methods in orthogonalization of multi- and hypervariate data," in *Proc. IEEE Int. Conf. Image Process.*, Cairo, Egypt, Nov. 7–11, 2009, pp. 3729–3732.

[19] R. Larsen, M. Arngren, P. W. Hansen, and A. A. Nielsen, "Kernel based subspace projection of near infrared hyperspectral images of maize kernels," in *Proc. 16th Scandinavian Conf. Image Anal.*, Oslo, Norway, Jun. 15–18, 2009, pp. 560–569.

[20] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature space," in *Proc. Adv. Neural Inf. Process. Syst. 11*, 1999, pp. 536–542.

[21] J. T. Kwok and I. W. Tsang, "The pre-image problem in kernel methods," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1517–1525, Nov. 2004.

[22] T. J. Abrahamsen and L. K. Hansen, "Input space regularization stabilizes pre-images for kernel PCA de-noising," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Grenoble, France, Sep. 2–4, 2009, pp. 1–6.

[23] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psykometrika*, vol. 1, pp. 211–218, 1936.

[24] R. M. Johnson, "On a theorem stated by Eckart and Young," *Psykometrika*, vol. 28, no. 3, pp. 259–263, 1963.

[25] A. A. Nielsen, "Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data" Ph.D. dissertation, Inf. Math. Model. Dept., Tech. Univ. of Denmark, Lyngby, 1994 [Online]. Available: http://www.imm.dtu.dk/pubdb/p.php?296

[26] A. A. Nielsen, K. B. Hilger, O. B. Andersen, and P. Knudsen, L. Bruzzone and P. Smits, Eds., "A temporal extension to traditional empirical orthogonal function analysis," in *Proc. MultiTemp2001*, Trento, Italy, Sep. 13–14, 2001, pp. 164–170.

[27] A. A. Nielsen, "An extension to a filter implementation of a local quadratic surface for image noise estimation," in *Proc. 10th Int. Conf. Image Anal. Process.*, Venice, Italy, Sep. 27–29, 1999, pp. 119–124.

[28] F. Kurz, B. Charmette, S. Suri, D. Rosenbaum, M. Spangler, A. Leonhardt, M. Bachleitner, R. Stätter, and P. Reinartz, U. Stilla, H. Mayer, F. Rottensteiner, C. Heipke, and S. Hinz, Eds., "Automatic traffic monitoring with an airborne wide-angle digital camera system for estimation of travel times," in *Proc. Photogram. Image Anal. Int. Archives Photogram. Remote Sens. Spatial Inf. Service*, Munich, Germany, 2007, pp. 83–88.

[29] F. Kurz, R. Müller, M. Stephani, P. Reinartz, and M. Schroeder, C. Heipke, K. Jacobsen, and M. Gerke, Eds., "Calibration of a wide-angle digital camera system for near real time scenarios," in *Proc. ISPRS Workshop, High Resolution Earth Imag. Geospatial Inf.*, Hannover, Germany, 2007.

[30] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, "The HyMap airborne hyperspectral sensor: The system, calibration, and performance," in *Proc. 1st EARSeL Workshop Imag. Spectroscopy*, Zürich, Switzerland, Oct. 6–8, 1998, pp. 37–42.

[31] A. A. Nielsen, A. Müller, and W. Dorigo, "Hyperspectral data, change detection and the MAD transformation," in *Proc. 12th Austral. Remote Sens. Photogram. Conf.*, Fremantle, Western Australia, Oct. 18–22, 2004, pp. 683–688.

[32] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.* vol. 16, no. 2, pp. 463–478, Feb. 2007.

[33] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: New approaches to change detection studies," *Remote Sens. Environ.* vol. 64, pp. 1–19, 1998.

[34] M. J. Canty, A. A. Nielsen, and M. Schmidt, "Automatic radiometric normalization of multitemporal satellite imagery," *Remote Sens. Environ.* vol. 91, no. 3–4, pp. 441–451, 2004.

[35] M. J. Canty and A. A. Nielsen, "Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation," *Remote Sens. Environ.* vol. 112, no. 3, pp. 1025–1036, 2008.
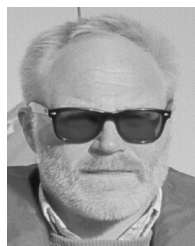
Allan A. Nielsen received the M.Sc. and Ph.D. degrees from the Technical University of Denmark, Lyngby, in 1978 and 1994, respectively.

He is currently an Associate Professor with the National Space Institute, Technical University of Denmark (DTU). He was with the Danish Defense Research Establishment from 1977 to 1978. He worked on energy conservation in housing with DTU's Thermal Insulation Laboratory from 1978 to 1985, with the section for image analysis in DTU's Department of Informatics and Mathematical Modelling (IMM) from 1985 to 2001, as a Part-Time Acting Professor from 1995 to 2001, and with the section for geoinformatics at IMM from 2001 to 2006. Since 1985, he has been working on several national and international projects on the development, implementation, and application of statistical methods and remote sensing in mineral exploration, mapping, geology, agriculture, environmental monitoring, oceanography, geodesy, and security funded by industries, the European Union, the Danish International Development Agency, and the Danish National Research Councils.