

Kernel principal component analysis for change detection

Allan A. Nielsen^a and Morton J. Canty^b

^aTechnical University of Denmark
DTU Space – National Space Institute
DK-2800 Kgs. Lyngby, Denmark

^bResearch Center Juelich
Institute of Chemistry and Dynamics of the Geosphere
D-52425 Juelich, Germany

ABSTRACT

Principal component analysis (PCA) is often used to detect change over time in remotely sensed images. A commonly used technique consists of finding the projections along the two eigenvectors for data consisting of two variables which represent the same spectral band covering the same geographical region acquired at two different time points. If change over time does not dominate the scene, the projection of the original two bands onto the second eigenvector will show change over time. In this paper a kernel version of PCA is used to carry out the analysis. Unlike ordinary PCA, kernel PCA with a Gaussian kernel successfully finds the change observations in a case where nonlinearities are introduced artificially.

Keywords: Orthogonal transformations, dual formulation, Q-mode analysis, kernel substitution, kernel trick.

1. INTRODUCTION

Based on work by Pearson¹ in 1901, Hotelling² in 1933 introduced principal component analysis (PCA). PCA is often used for linear orthogonalization or compression by dimensionality reduction of correlated multivariate data, see Jolliffe³ for a comprehensive description of PCA and related techniques. An interesting dilemma in 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.

Wiemker et al.⁴ describe iterated PCA to change detection in data consisting of two variables which represent the same spectral band covering the same geographical region acquired at two different time points. Schölkopf et al.⁵ introduce kernel PCA. Shawe-Taylor and Cristianini⁶ is an excellent reference for kernel methods in general. Bishop⁷ and Press et al.⁸ describe kernel methods among many other subjects.

The kernel version of PCA handles 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 this paper we shall apply kernel PCA to detect change over time in remotely sensed images by finding the projections along the eigenvectors for data consisting of two variables which represent the same spectral band covering the same geographical region acquired at two different time points. If change over time does not dominate the scene, the projection of the original two bands onto the second eigenvector from an ordinary PCA will show change over time. For kernel PCA change may be depicted by (a) higher order component(s).

Further author information:

A.A.N.: Presently located at DTU Informatics – Department of Informatics and Mathematical Modelling, Richard Petersens Plads, Building 321, E-mail aa@space.dtu.dk, <http://www.imm.dtu.dk/~aa>, Tel +45 4525 3425, Fax +45 4588 1397.

M.J.C.: E-mail m.canty@fz-juelich.de, http://www.fz-juelich.de/ste/remote_sensing, Tel +49 (0)2461 614885, Fax +49 (0)2461 612518.

2. PRINCIPAL COMPONENT ANALYSIS

Let us consider an image with n observations or pixels and p spectral bands organized as a matrix X with n rows and p columns; each column contains measurements over all pixels from one spectral band and each row consists of a vector of measurements x_i^T from p spectral bands for a particular observation

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

The superscript T denotes the transpose. X is sometimes called the data matrix or the design matrix. Without loss of generality we assume that the spectral bands in the columns of X have mean value zero.

2.1 Primal Formulation

In ordinary (primal also known as R-mode) PCA we analyze the 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 full rank $r = \min(n, p)$ this will lead to r non-zero eigenvalues λ_i and r orthogonal or mutually conjugate unit length eigenvectors u_i ($u_i^T u_i = 1$) from the eigenvalue problem

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

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, $x^T u_i$. 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.

2.2 Dual Formulation

In the dual formulation (also known as Q-mode analysis) we analyze $X X^T / (n - 1)$ which is n by n and which in image applications can be very large. Multiply both sides of Equation 2 from the left with X

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

or

$$\frac{1}{n-1} X X^T v_i = \lambda_i v_i \quad (4)$$

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. Now multiply both sides of Equation 4 from the left with X^T

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

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

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

We see that if $X^T X$ is full rank $r = \min(n, p)$, $X^T X / (n - 1)$ and $X X^T / (n - 1)$ have the same r non-zero 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^{9,10} theorem.

An obvious advantage of the dual formulation is the case where $n < p$. Another advantage even for $n \gg p$ is due to the fact that the elements of the matrix $G = X X^T$, which is known as the Gram* matrix, consist of inner products of the multivariate observations in the rows of X , $x_i^T x_j$.

* named after Danish mathematician Jørgen Pedersen Gram (1850-1916)

2.3 Regularization

If $X^T X$ is singular or near singular we often replace it by $(1 - k)X^T X + kI_p$ where k is a small positive number and I_p is the p by p unit matrix. It is easily seen that regularization in the primal and dual formulations with the same k leads to the same non-zero eigenvalues for $(1 - k)X^T X + kI_p$ and $(1 - k)X X^T + kI_n$, and to eigenvectors related as above. In the latter case I_n of course is the n by n unit matrix.

2.4 Kernel Formulation

We now replace x by $\phi(x)$ which maps x nonlinearly into a typically higher dimensional feature space. As an example consider a two-dimensional vector $[z_1 \ z_2]^T$ being mapped into $[z_1 \ z_2 \ z_1^2 \ z_2^2 \ z_1 z_2]^T$. This maps the original two-dimensional vector into a five-dimensional feature space so that for example a linear decision rule becomes general enough to differentiate between all linear and quadratic forms including ellipsoids.

The mapping by ϕ takes X into Φ which is an n by q ($q \geq p$) matrix

$$\Phi = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix}. \quad (7)$$

For the moment we assume that the mappings in the columns of Φ have zero mean. In this higher dimensional feature space $C = \Phi^T \Phi / (n - 1) = 1 / (n - 1) \sum_{i=1}^n \phi(x_i) \phi(x_i)^T$ is the variance-covariance matrix and for PCA we get the primal formulation

$$\frac{1}{n - 1} \Phi^T \Phi u_i = \lambda_i u_i \quad (8)$$

where we have re-used the symbols λ_i and u_i from above.

For the corresponding dual formulation we get

$$\frac{1}{n - 1} \Phi \Phi^T v_i = \lambda_i v_i \quad (9)$$

where we have re-used the symbol v_i from above. As above the non-zero eigenvalues for the primal and the dual formulations are the same and the eigenvectors are related by

$$u_i = \frac{1}{\sqrt{(n - 1)\lambda_i}} \Phi^T v_i \quad (10)$$

and $v_i = \Phi u_i / \sqrt{(n - 1)\lambda_i}$.

Here $\Phi \Phi^T$ plays the same role as the Gram matrix above and has the same size, namely n by n (so introducing the nonlinear mappings in ϕ does not make the eigenvalue problem in Equation 9 bigger).

2.4.1 Kernel Substitution

Applying kernel substitution also known as the kernel trick we replace the inner products $\phi(x_i)^T \phi(x_j)$ in $\Phi \Phi^T$ with a kernel function $\kappa(x_i, x_j) = \kappa_{ij}$ which could have come from some unspecified mapping ϕ . In this way we avoid the explicit mapping ϕ of the original variables. We obtain

$$K v_i = (n - 1) \lambda_i v_i \quad (11)$$

where $K = \Phi \Phi^T$ is an n by n matrix with elements $\kappa(x_i, x_j)$. To be a valid kernel K must be symmetric and positive semi-definite, i.e., its eigenvalues are non-negative. Normally the eigenvalue problem is formulated without the factor $n - 1$

$$K v_i = \lambda_i v_i. \quad (12)$$

This gives the same eigenvectors v_i and eigenvalues $n - 1$ times greater. In this case $u_i = \Phi^T v_i / \sqrt{\lambda_i}$ and $v_i = \Phi u_i / \sqrt{\lambda_i}$.

2.4.2 Basic Properties

Several basic properties including the norm in feature space, the distance between observations in feature space, the norm of the mean in feature space, centering to zero mean in feature space, and standardization to unit variance in feature space, may all be expressed in terms of the kernel function without using the mapping by ϕ explicitly.^{6,7}

2.4.3 Projections onto Eigenvectors

To find the kernel principal component scores from the eigenvalue problem in Equation 12 we project a mapped x onto the primal eigenvector u_i

$$\phi(x)^T u_i = \phi(x)^T \Phi^T v_i / \sqrt{\lambda_i} \quad (13)$$

$$= \phi(x)^T \begin{bmatrix} \phi(x_1) & \phi(x_2) & \cdots & \phi(x_n) \end{bmatrix} v_i / \sqrt{\lambda_i} \quad (14)$$

$$= \begin{bmatrix} \phi(x)^T \phi(x_1) & \phi(x)^T \phi(x_2) & \cdots & \phi(x)^T \phi(x_n) \end{bmatrix} v_i / \sqrt{\lambda_i} \quad (15)$$

$$= \begin{bmatrix} \kappa(x, x_1) & \kappa(x, x_2) & \cdots & \kappa(x, x_n) \end{bmatrix} v_i / \sqrt{\lambda_i}, \quad (16)$$

or in matrix notation $\Phi U = K V \Lambda^{-1/2}$ (U is a matrix with u_i in the columns, V is a matrix with v_i in the columns and $\Lambda^{-1/2}$ is a diagonal matrix with elements $1/\sqrt{\lambda_i}$), i.e., also the projections may be expressed in terms of the kernel function without using ϕ explicitly.

The variance of this projection is

$$\text{Var}\{u_i^T \phi(x)\} = u_i^T C u_i \quad (17)$$

$$= u_i^T \Phi^T \Phi u_i / (n - 1) \quad (18)$$

$$= v_i^T \Phi \Phi^T \Phi \Phi^T v_i / ((n - 1) \lambda_i) \quad (19)$$

$$= v_i^T K K v_i / ((n - 1) \lambda_i) \quad (20)$$

$$= \lambda_i / (n - 1). \quad (21)$$

If the mapping by ϕ is not column centered the variance of the projection $u_i^T \phi(x)$ must be adjusted by subtraction of $n/(n - 1)$ times the squared mean of the projection, i.e., we must subtract $n/(n - 1)$ times ($\mathbf{1}_n$ here is an n by 1 vector of ones divided by n)

$$(\mathbb{E}\{u_i^T \phi(x)\})^2 = (u_i^T \bar{\phi})^2 \quad (22)$$

$$= (u_i^T \Phi^T \mathbf{1}_n)^2 \quad (23)$$

$$= (v_i^T \Phi \Phi^T \mathbf{1}_n)^2 / \lambda_i \quad (24)$$

$$= (v_i^T K \mathbf{1}_n)^2 / \lambda_i \quad (25)$$

$$= \lambda_i (v_i^T \mathbf{1}_n)^2 \quad (26)$$

from the variance in Equation 21. $v_i^T \mathbf{1}_n$ is the mean value of the elements in vector v_i .

Kernel PCA is a so-called memory-based method: from Equation 16 we see that if x is a new data point that didn't go into building the model, i.e., finding the eigenvectors and -values, we need the original data x_1, x_2, \dots, x_n as well as the eigenvectors and -values to find scores for the new observations. This is not the case for ordinary PCA where we don't need the training data to project new observations.

2.4.4 Some Popular Kernels

Popular choices for the kernel function are stationary kernels that depend on the vector difference $x_i - x_j$ only (they are therefore invariant under translation in feature space), $\kappa(x_i, x_j) = \kappa(x_i - x_j)$, and homogeneous kernels also known as radial basis functions (RBFs) that depend on the Euclidean distance between x_i and x_j only, $\kappa(x_i, x_j) = \kappa(\|x_i - x_j\|)$. Some of the most often used RBFs are ($h = \|x_i - x_j\|$)

- multiquadric: $\kappa(h) = (h^2 + h_0^2)^{1/2}$,

- inverse multiquadric: $\kappa(h) = (h^2 + h_0^2)^{-1/2}$,
- thin-plate spline: $\kappa(h) = h^2 \log(h/h_0)$ (which tends to 0 for h tending to 0), or
- Gaussian: $\kappa(h) = \exp(-\frac{1}{2}(h/h_0)^2)$,

where h_0 is a scale parameter to be chosen. Generally, h_0 should be chosen larger than a typical distance between samples and smaller than the size of the study area. Other kernels often used (which are not RBFs) are

- linear: $\kappa(x_i, x_j) = x_i^T x_j$,
- power: $\kappa(x_i, x_j) = (x_i^T x_j)^p$,
- polynomial: $\kappa(x_i, x_j) = (x_i^T x_j + h_0)^p$.

As an example consider the polynomial kernel function $\kappa(x, x') = (x^T x' + h_0)^2$ with two-dimensional $x = [z_1 \ z_2]^T$ and $x' = [z'_1 \ z'_2]^T$. We obtain

$$\kappa(x, x') = (x^T x' + h_0)^2 \tag{27}$$

$$= (z_1 z'_1 + z_2 z'_2 + h_0)^2 \tag{28}$$

$$= z_1^2 z_1'^2 + z_2^2 z_2'^2 + h_0^2 + 2z_1 z'_1 z_2 z'_2 + 2z_1 z'_1 h_0 + 2z_2 z'_2 h_0 \tag{29}$$

$$= [h_0 \ \sqrt{2h_0} z_1 \ \sqrt{2h_0} z_2 \ z_1^2 \ z_2^2 \ \sqrt{2} z_1 z_2] [h_0 \ \sqrt{2h_0} z'_1 \ \sqrt{2h_0} z'_2 \ z_1'^2 \ z_2'^2 \ \sqrt{2} z'_1 z'_2]^T \tag{30}$$

$$= \phi(x)^T \phi(x'). \tag{31}$$

We see that the kernel function maps the two-dimensional vector into six dimensions which (apart from the constant in the first dimension and the specific weighting) corresponds to the mapping mentioned in Section 2.4. For many kernels this decomposition back into $\phi(x)^T \phi(x')$ is not possible.

Its important to realize that the information content in the original data is conveyed to a kernel method through the choice of kernel only (and possibly through a labeling of the data; this is not relevant for kernel PCA). For example, since kernel methods are implicitly based on inner products, any rotation by an orthogonal matrix Q of the original coordinate system will not influence the result of the analysis, $(Qx_i)^T Qx_j = x_i^T Q^T Qx_j = x_i^T x_j$.

3. DATA

The images used were recorded with the airborne DLR 3K-camera system^{11,12} 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 600 by 600 pixel sub-images acquired 0.7 seconds apart cover a busy motorway near Munich in Bavaria, Germany. Figure 1 (left) shows the the image at time point 1 as red and at time point 2 as cyan.

A nonlinear version of the data is constructed by raising the data at time point 2 to the power of three and normalizing its variance to that of the data at time point 1.

For both real data and data with the artificial nonlinearity the only real change on the ground is very likely to be the movements of the vehicles on the motorway.

4. RESULTS AND DISCUSSION

To be able to carry out kernel PCA on the large amounts of pixels typically present in Earth observation data, we sub-sample the image and use a small portion termed the training data only. We typically use in the order 10^3 training pixels (here $\sim 2,000$) to find the eigenvectors onto which we then project the entire image termed the test data kernelized with the training data. This sub-sampling potentially avoids problems that may arise from the spatial autocorrelation inherent to image data. Figure 1 (right) shows the positions of the training pixels. A Gaussian kernel $\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$ with σ equal to three times the mean distance between the observations in feature space is used.

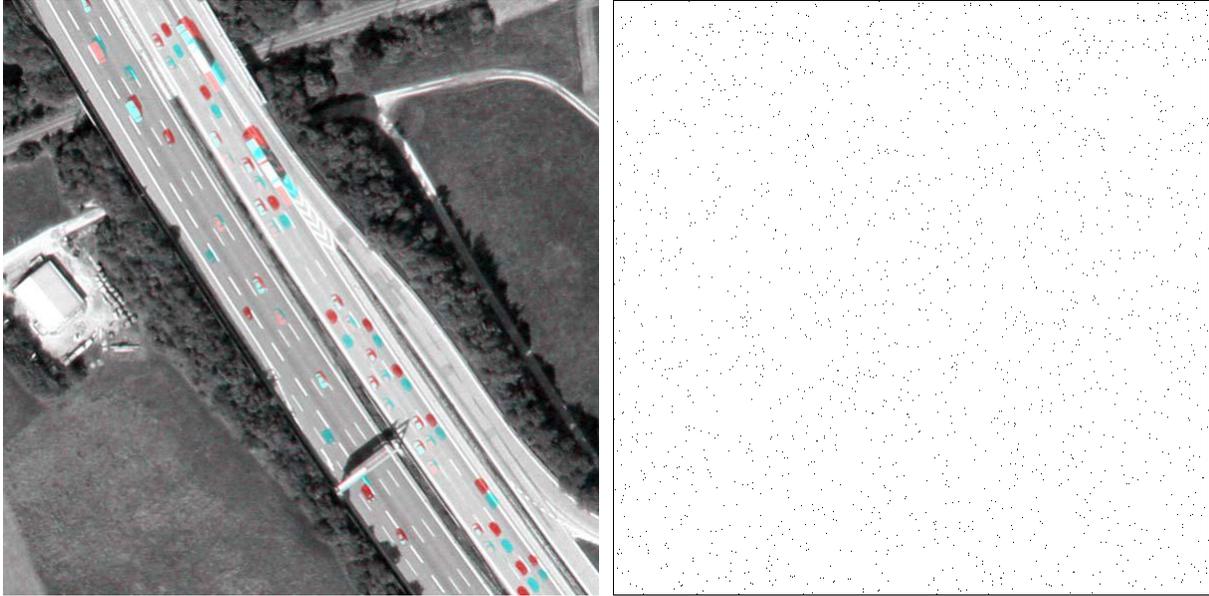


Figure 1. Image from time point 1 as red and time point 2 as cyan (left), $\sim 2,000$ samples used to solve the eigenvalue problem (right).

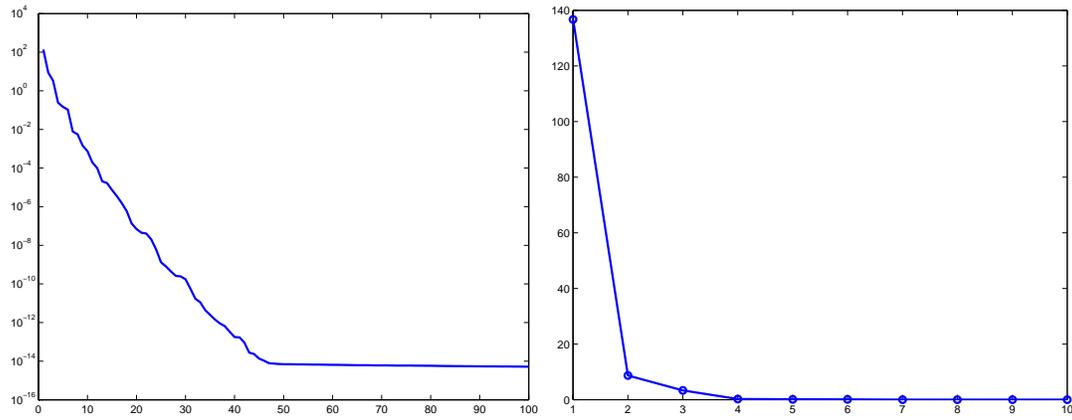


Figure 2. Eigenvalues for kernel PCA of original data.

For the ordinary PCA there are two eigenvalues only; these are 3731.86 and 84.37. Figure 2 shows eigenvalues for kernel PCA of the original data, logarithms of the first 100 eigenvalues (left) and the 10 first eigenvalues (right). For the artificial nonlinear data the eigenvalues are very similar. Although the dimensionality of the implicitly mapped data is in principle infinite, the data seem to reside in a sub-space with dimensionality around 45.

Figure 3 shows scatterplots of the $\sim 2,000$ training pixels at times 1 and 2 on backgrounds of contours for projections onto PCs 2 for the original data (left) and for the data with an artificial nonlinearity (right).

Figure 4 shows scatterplots of the $\sim 2,000$ training pixels at times 1 and 2 on backgrounds of contours for projections onto kernel PCs 1 (left), 2 (middle), and 3 (right) for the original data.

Figure 5 shows scatterplots of the $\sim 2,000$ training pixels at times 1 and 2 on backgrounds of contours for projections onto kernel PCs 1 (left), 2 (middle), and 3 (right) for the data with an artificial nonlinearity.

We see that the change for the original data is nicely depicted by PC 2, Figure 3 (left). With kernel PCA change is depicted by PC 3, Figures 4 and 5, right. The contours for kernel PC 3 for the original data are nearly linear, Figure 4 right. In Figure 5 (right) the no-change pixels nicely follow the contours of kernel PC 3. This is not the case for the (non-kernel) PC 2 in Figure 3 (right).

Figure 6 shows scores for kernel PCs 3 for the original data (left) and the artificially nonlinear data (right). Although some details in the no-change background (middle-gray pixels) differ, the over-all impression is that the same good discrimination between change (very dark and very bright pixels) and no-change is obtained for both cases.

The results will depend on the choice of kernel, the choice of the scale parameter, and the actual training samples used to build the kernel change detector.

5. CONCLUSIONS AND FUTURE

In the dual formulation of PCA 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 PCA is to express the inner products between the mappings in terms of a kernel function to avoid the explicit use of the mappings. Both the eigenvalue problem, the centering to zero mean and the projections onto eigenvectors to find kernel PC scores may be expressed by means of the kernel function. Kernel PCA handles 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.

Kernel PCA with a Gaussian kernel $\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$ is used for detecting change in data consisting of two variables which represent the same spectral band covering the same geographical region acquired at two different time points. Unlike ordinary PCA kernel PCA successfully finds the change observations in a case where nonlinearities are introduced artificially.

Kernel PCA is a so-called memory-based method: where ordinary PCA handles new observations by projecting them onto the eigenvectors found based on the training data, because of the kernelization of the new observations with the training observations, kernel PCA needs the original data as well as the eigenvectors and λ -values to handle new data.

Inspired by the success of ordinary canonical correlation analysis (CCA) to multivariate change detection¹³⁻¹⁵ and normalization over time^{16,17} the application of kernel CCA to these subjects should be investigated.

Inspired by Wiemker et al.⁴ an iterative scheme may be built into the kernel PCA change detector.

ACKNOWLEDGMENTS

Thanks to Dr. Peter Reinartz and colleagues at the German Aerospace Center (DLR) at Oberpfaffenhofen, Germany, for letting us use the airborne data.

This work was carried out partly within the project Global Monitoring for Security and Stability (GMOSS) which is a Network of Excellence in the Aeronautics and Space Priority of the Sixth Framework Programme funded by the European Commission's Directorate General Enterprise and Industry, see <http://gmoss.jrc.it>.

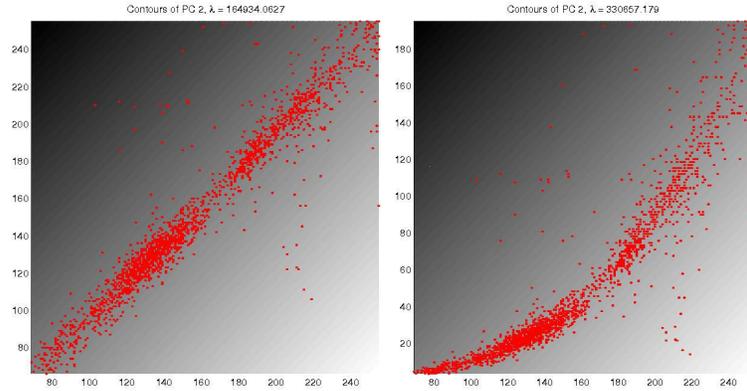


Figure 3. Scatterplots of training data from time points 1 and 2 on contours of projections onto principal components 2 for original data (left) and for data with artificial nonlinearity (right).

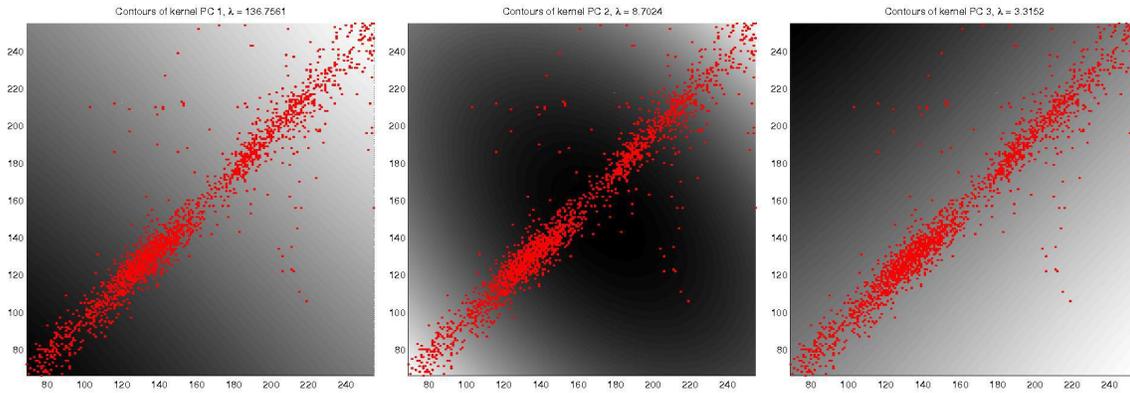


Figure 4. Scatterplots of training data from time points 1 and 2 on contours of projections onto kernel principal components 1 (left), 2 (middle) and 3 (right) for original data.

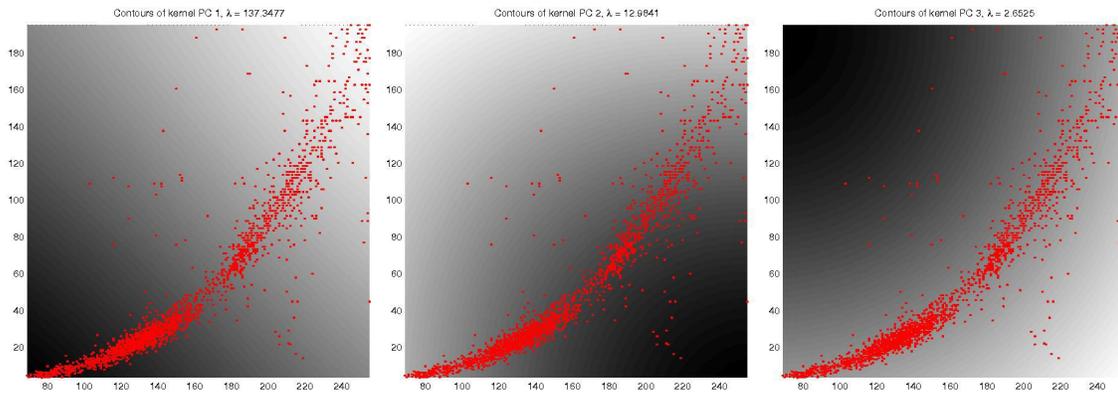


Figure 5. Scatterplots of training data from time points 1 and 2 on contours of projections onto kernel principal components 1 (left), 2 (middle) and 3 (right) for data with artificial nonlinearity.

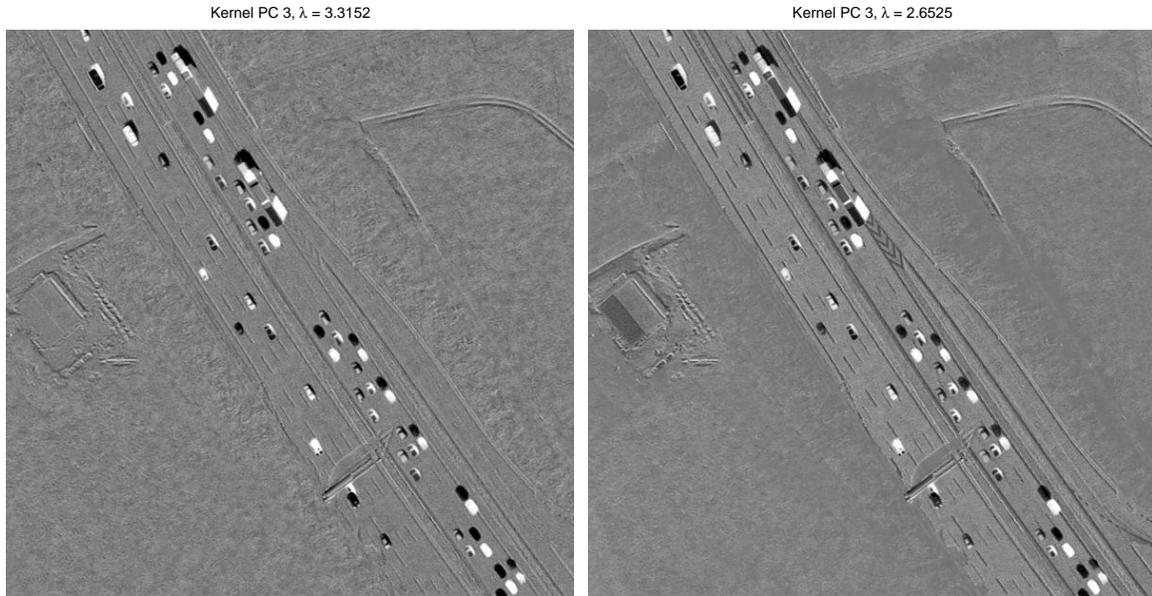


Figure 6. Kernel principal component 3 from original data (left), and for data with artificial nonlinearity (right).

REFERENCES

- [1] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine* **2**(6), 559–572 (1901).
- [2] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology* **24**, pp. 417–441 and pp. 498–520 (1933).
- [3] I. T. Jolliffe, *Principal Component Analysis*, second edition, Springer (2002).
- [4] R. Wiemker, A. Speck, D. Kulbach, H. Spitzer, and J. Beinlein, "Unsupervised robust change detection on multispectral imagery using spectral and spatial features," in *Proceedings from the Third International Airborne Remote Sensing Conference and Exhibition*, Copenhagen, Denmark, vol. I, 640–647 (1997).
- [5] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation* **10**(5), 1299–1319 (1998).
- [6] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press (2004).
- [7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).
- [8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, third edition, Cambridge University Press (2007).
- [9] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.
- [10] R. M. Johnson, "On a theorem stated by Eckart and Young," *Psychometrika*, vol. 28, no. 3, pp. 259–263, 1963.
- [11] F. Kurz, B. Charmette, S. Suri, D. Rosenbaum, M. Spangler, A. Leonhardt, M. Bachleitner, R. Stätter, and P. Reinartz, "Automatic traffic monitoring with an airborne wide-angle digital camera system for estimation of travel times," in U. Stilla, H. Mayer, F. Rottensteiner, C. Heipke, and S. Hinz (Eds.), *Photogrammetric Image Analysis, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Service, PIA07*, Munich, Germany (2007).
- [12] F. Kurz, R. Müller, M. Stephani, P. Reinartz, M. Schroeder, "Calibration of a wide-angle digital camera system for near real time scenarios," in C. Heipke, K. Jacobsen, M. Gerke (Eds.) *ISPRS Workshop, High Resolution Earth Imaging for Geospatial Information*, Hannover, Germany, ISSN 1682–1777 (2007).

- [13] 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 Sensing of Environment* **64**, 1–19 (1998), Internet <http://www.imm.dtu.dk/pubdb/p.php?1220>.
- [14] A. A. Nielsen, K. Conradsen, and O. B. Andersen, "A change oriented extension of EOF analysis applied to the 1996-1997 AVHRR sea surface temperature data," *Physics and Chemistry of the Earth* **27**(32–34), 1379–1386 (2002), Internet <http://www.imm.dtu.dk/pubdb/p.php?491>.
- [15] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Transactions on Image Processing* **16**(2), 463–478 (2007), Internet <http://www.imm.dtu.dk/pubdb/p.php?4695>.
- [16] M. J. Canty, A. A. Nielsen, and M. Schmidt, "Automatic radiometric normalization of multi-temporal satellite imagery," *Remote Sensing of Environment* **91**(3–4), 441–451 (2004), Internet <http://www.imm.dtu.dk/pubdb/p.php?2815>.
- [17] M. J. Canty and A. A. Nielsen, "Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation," *Remote Sensing of Environment* **112**(3), 1025–1036 (2008), Internet <http://www.imm.dtu.dk/pubdb/p.php?5362>.