

Kernel principal component and maximum autocorrelation factor analyses 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

Kernel versions of the principal components (PCA) and maximum autocorrelation factor (MAF) transformations are used to postprocess change images obtained with the iteratively re-weighted multivariate alteration detection (MAD) algorithm. It is found that substantial improvements in the ratio of signal (change) to background noise (no change) can be obtained especially with kernel MAF.

1. INTRODUCTION

Principal component analysis (PCA) [1,2] has often been used to detect change over time in remotely sensed images. A commonly used technique consists of finding the projections along the eigenvectors for data consisting of pair-wise (perhaps generalized) differences between corresponding spectral bands covering the same geographical region acquired at two different time points. In this paper kernel versions of the principal component (PCA) and maximum autocorrelation factor (MAF) transformations [3,4] are used to carry out the analysis. An example is given based on bi-temporal Landsat-5 TM imagery. For an application of kernel PCA to change detection with nonlinear data, see [5].

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/st/remote_sensing, Tel +49 (0)2461 614885, Fax +49 (0)2461 612518.

2. KERNEL PCA AND KERNEL MAF

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 from a spectral band and each row consists of a vector of measurements from p spectral bands for a particular observation, x_i . Without loss of generality let us assume that the spectral bands in the columns of X have mean value zero. In ordinary (R-mode or primal) PCA we analyse the variance-covariance matrix $X^T X / (n - 1)$ which is p by p . Here T denotes the transpose. If $X^T X$ is full rank this will lead to p non-zero eigenvalues and p orthogonal or mutually conjugate unit length eigenvectors.

In the Q-mode or dual formulation [5] we analyze $XX^T / (n - 1)$ which is n by n and can be very big. If $X^T X$ is full rank $X^T X / (n - 1)$ and $XX^T / (n - 1)$ have the same p non-zero eigenvalues λ_i and their eigenvectors (u_i for $X^T X / (n - 1)$ and v_i for $XX^T / (n - 1)$) are related by

$$u_i = X^T v_i / \sqrt{(n - 1)\lambda_i} \quad \text{and} \quad v_i = X u_i / \sqrt{(n - 1)\lambda_i}.$$

This leads to the desired

$$u_i^T u_i = v_i^T v_i = 1.$$

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 (so-called Gram) matrix XX^T consist of inner products of the multivariate observations in the columns of X , namely $x_i^T x_j$. We may now replace these inner products in XX^T by the inner products of some nonlinear mapping ϕ of x into higher (even infinite) dimensional feature space, i.e.,

$$x_i^T x_j \rightarrow \phi(x_i)^T \phi(x_j).$$

Applying kernel substitution, also known as the kernel trick [7,8], we may even replace the inner products $\phi(x_i)^T \phi(x_j)$ with some kernel function $\kappa(x_i, x_j)$ without explicit reference to the mapping $\phi(x)$. The only prerequisite is that κ must be a so-called Mercer kernel, i.e., the n by n kernel matrix with elements $\kappa(x_i, x_j)$ must be symmetric and positive semidefinite. Moreover, projection of new observations x along the principal axes in the mapped feature space and other quantities needed to do the analysis, may be expressed in terms of the kernel function only. For example, the kernel matrix $\tilde{\kappa}$ corresponding to column-centered feature space observations is given by

$$\tilde{\kappa} = \kappa - \kappa \cdot 1_{nn} / n - 1_{nn} \cdot \kappa / n + 1_{nn} \cdot \kappa \cdot 1_{nn} / n^2,$$

where 1_{nn} is an $n \times n$ matrix of ones. Popular choices 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) that depend on the Euclidean distance between x_i and x_j only, $\kappa(x_i, x_j) = \kappa(\|x_i - x_j\|)$.

The kernel version of PCA handles non-linearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space. The variance of projections along the i th eigenvector u_i in the transformed feature space (the kernel principal components) is given by

$$\text{Var}(u_i^T \phi(x)) = \frac{\lambda_i}{n - 1},$$

where λ_i is the corresponding i th eigenvalue of the (centered) kernel matrix.

As with the kernel PCA, we can use the combination of the dual formulation and the kernel trick to obtain an implicit non-linear mapping for the MAF transform. The variance of the kernel MAFs, again for centered kernel matrices, is given by

$$\text{Var}(u_i^T \phi(x)) = \frac{1}{n-1}.$$

A detailed account of the kernel MAF transform is given in [9].

3. DATA ANALYSIS

In the case of the kernel versions, to be able to treat the large amounts of pixels (order $10^6 - 10^8$) it is necessary to sub-sample the image and use a small portion only (in the order $n = 10^3$ training pixels) to find the eigenvectors onto which we then project the kernelized versions of the entire image. The n sampled pixel vectors are then referred to as *training data* and the calculation of the centered kernel matrix and its diagonalization constitute the *training phase*. This sub-sampling potentially avoids problems that may arise from the spatial autocorrelation inherent to image data.

After diagonalization, the *generalization phase* involves the projection of each image pixel vector along the eigenvectors of the kernel matrix. For an image with N pixels there are therefore $n \times N$ kernels involved, generally much too large an array to be held in memory, so that it is advisable to read in and project the image pixels row-by-row.

Kernel PCA and MAF are so-called memory-based methods: whereas ordinary PCA (and MAF) 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 (and MAF) needs the original training data as well as the eigenvectors to handle new data (and for kernel PCA also the eigenvalues).

4. RESULTS

Both satellite images and airborne digital photographs are used in the following to illustrate the effects of kernelization for enhancing change signals.

4.1 Satellite imagery

Two six-band LANDSAT 5 TM images acquired over irrigation fields in Nevada on successive passes of the satellite in August-September 1991 (the thermal band is omitted) with 1,000 by 1,000 28.5 m pixels were registered to one another with sub-pixel accuracy. Then they were processed with the iteratively re-weighted MAD (IR-MAD) algorithm [10] in order to discriminate change. The MAD image was post-processed with both ordinary and kernel versions of PCA and MAF analysis.

Kernel MAF was found to suppress the noisy no-change background much more successfully than ordinary MAF, see Figure 1 showing the MAD image post-processed with ordinary MAF and kernel MAF transformations. In the latter case a Gaussian kernel $\kappa(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ with the scale parameter $\gamma = 0.01$ was applied. The change signals are due almost entirely to differing

vegetation reflection characteristics. Kernel MAF suppresses the noisy no-change background much more successfully than ordinary MAF.

The ratio between variances of the ordinary MAF component 1 and the kernel MAF component 1 (both scaled to unit variance) calculated in a no-change region of the images is 140 corresponding to 21.5 dB. Kernel MAF analysis also outperforms both linear and kernel PCA here (not shown).

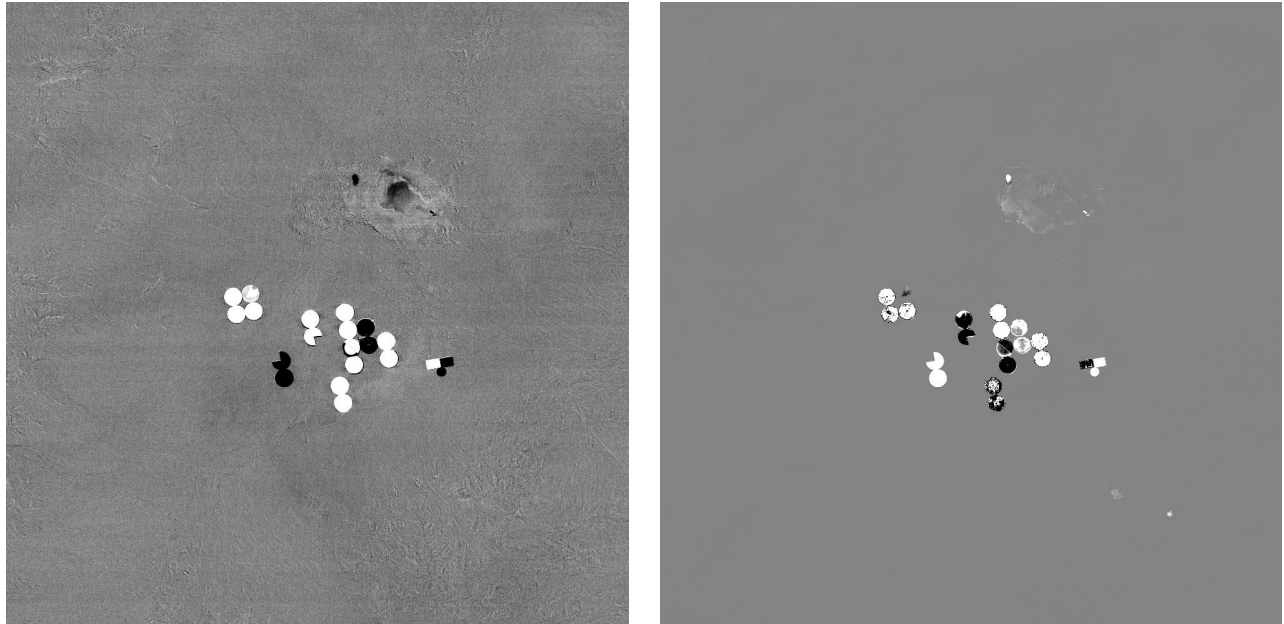


Figure 1. MAF band 1 (left), kernel MAF band 1 (right) for the LANDSAT-5 TM bitemporal image. Both are stretched linearly from mean (middle gray) minus to plus two standard deviations.

4.2 Aerial imagery

Aerial color photographs were recorded with the airborne DLR 3K-camera system from the German Aerospace Center (DLR) [11], a system consisting of three off-the-shelf cameras arranged on a mount with one camera looking in the nadir direction and two cameras tilted approximately 35° across track. Two 1000 by 1000 pixel sub-images acquired 0.7 seconds apart and covering a busy motorway near Munich, Germany, were registered to one another with subpixel accuracy. The only physical changes on the ground are due to the motion of the automobiles.

Again the images were processed with the IR-MAD algorithm and post-processed with ordinary and kernel MAF, see Figure 2. In this case, kernel MAF component 5 exhibited the smallest no-change variance: The ratio between variances of the ordinary MAF component 1 and kernel MAF component 5 (both scaled as before to unit variance) calculated in a no-change region of the images is 12.

5. CONCLUSIONS

We have applied both kernel PCA and kernel MAF nonlinear postprocessing to difference images obtained with the (linear) IR-MAD transformation and have achieved, in the case of kernel MAF, quite

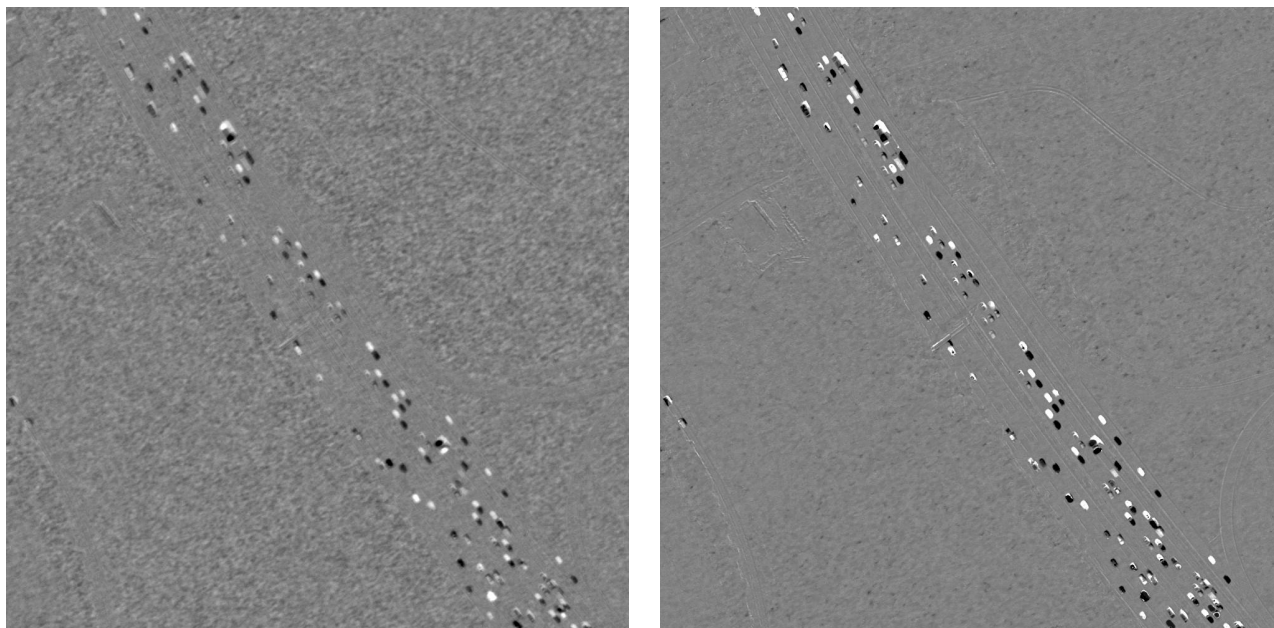


Figure 2. MAF band 1 (left), kernel MAF band 5 (right) for the traffic scene. Both are stretched linearly from mean (middle gray) minus to plus five standard deviations.

substantial enhancements of the ratio of change signal to background no-change noise. Further investigations will be needed in order to better understand this phenomenon, particularly the role played by the choice of different kernel functions and their parameters.

Acknowledgement

The authors would like to express their thanks to the German Aerospace Center (DLR) for permission to use the traffic scene images.

References

- [1] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 6, no. 2, pp. 559–572, 1901.
- [2] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441 and 498–520, 1933.
- [3] P. Switzer and A. A. Green, "Min/max autocorrelation factors for multivariate spatial imagery," Tech. Rep. 6, Department of Statistics, Stanford University, 1984.
- [4] P. Switzer and S. E. Ingebritsen, "Ordering of time-difference data from multispectral imagery," *Remote Sensing of Environment*, vol. 20, pp. 85–94, 1986.
- [5] A. A. Nielsen and M. J. Canty, "Kernel principal component analysis for change detection," SPIE Europe Remote Sensing Conference, vol. 7109A, Cardiff, Great Britain, 15–19 September 2008, Internet <http://www.imm.dtu.dk/pubdb/p.php?5667>.

- [6] B. Schölkopf, A. Smola and K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [7] J. Shawe-Taylor and N. Christianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- [8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [9] A. A. Nielsen, “Kernel maximum autocorrelation factor and minimum noise fraction transformations,” submitted, 2009.
- [10] 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>.
- [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, editors, *Photogrammetric Image Analysis*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Service PIA07, Munich, Germany, 2007.