# Kernel Principal Component Analysis for the Construction of the Extended Morphological Profile.

M. Fauvel*, J. Chanussot† and J. A. Benediktsson◇

*MISTIS, INRIA Rhônes Alpes & Laboratoire Jean Kuntsmann, Grenoble, France
†GIPSA-lab, Signal & Image Dept., Grenoble Institute of Technology - INPG, France
◇Dept. of Electrical and Computer Engineering, University of Iceland, Iceland

mathieu.fauvel@inrialpes.fr

The Morphological Profile (MP) has been proposed for the classification of remote sensing images with very high spatial resolution and reduced spectral information, such as panchromatic IKONOS data [1, 2]. It consists of a granulometry with two advanced morphological filters, the geodesic opening and the geodesic closing [3]. The improvement in terms of classification accuracies was clearly demonstrate over several experiments and now the MP is a well know tool of the remote sensing community, especially for the analysis of urban area.

The extension of the MP, namely the Extended Morphological Profile (EMP), to multispectral or hyperspectral data is not straightforward. Because of the multi-valued nature of pixels, the morphological operators which require a total ordering relation cannot be applied. Plaza et al. have proposed an extension to the morphological transformation in order to integrate the spectral and the spatial information from the hyperspectral data [4]. A simpler approach was proposed in [5]. The authors proposed to use the Principal Component Analysis (PCA) in order to reduce the spectral channels to a few number of features, the principal components (PC) and to apply the MP on each PCs. Then the EMP is built by the concatenation of all MPs. But it was found that too much spectral information was lost during the PCA.

To overcome this shortcoming, it was proposed to fuse the original spectral channels together with the EMP [6]. Supervised feature reduction algorithms, such as Decision Boundary Feature Reduction or Decision Boundary Feature Reduction, were used to reduce the redundancy of the fused features. Improved classification accuracies and substantial reduction of the training time were obtained.

In this paper, an alternative approach to data fusion is discussed. The PCA is optimal for the purpose of representation under some simple assumptions: The $n$ observed variable, *i.e.*, the spectral channels $\mathbf{x}$, result from a linear transformation of $m$ latent variables Gaussianly distributed and thus it is possible to recover the latent variable, *i.e.*, the principal components, from the observed one by solving the following eigenvalue problem:

$$\lambda \mathbf{v} = \Sigma_{\mathbf{x}} \mathbf{v}, \text{ subject to } \|\mathbf{v}\|_2 = 1 \tag{1}$$

where $\Sigma_{\mathbf{x}} = \mathbb{E}\left[\mathbf{x}_c \mathbf{x}_c^T\right] \approx \frac{1}{\ell-1}\sum_{i=1}^{\ell}\left(\mathbf{x}^i - \mathbf{m}_x\right)\left(\mathbf{x}^i - \mathbf{m}_x\right)^T$ and $\mathbf{x}_c$ is the centered vector $\mathbf{x} \in \mathbb{R}^n$. The PCA only relies on second order statistics and theoretical limitations for hyperspectral data analysis have been pointed out in [7]. Since the PCA does not handle all the spectral information, another unsupervised feature extraction is proposed, namely the Kernel PCA (KPCA) [8]. The KPCA is a non-linear version of the PCA solving a similar eigenvalue problem in a different feature space:

$$\lambda \boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}, \text{ subject to } \|\boldsymbol{\alpha}\|_2 = \frac{1}{\lambda} \tag{2}$$

where $\mathbf{K}$ is the kernel matrix constructed as follows:

$$\mathbf{K} = \begin{pmatrix} k(\mathbf{x}^1, \mathbf{x}^1) & \dots & k(\mathbf{x}^1, \mathbf{x}^\ell) \\ k(\mathbf{x}^2, \mathbf{x}^1) & \dots & k(\mathbf{x}^2, \mathbf{x}^\ell) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}^\ell, \mathbf{x}^1) & \dots & k(\mathbf{x}^\ell, \mathbf{x}^\ell) \end{pmatrix}. \tag{3}$$

The function $k$ is the core of the KPCA. It is a positive semi-definite function on $\mathbb{R}^n$ that introduces non-linearity into the processing. This is usually called a *kernel*. Using kernel, it is possible to implicitly compute the PCA in a feature space $\mathcal{H}$ where the data better fit the PCA model. This is the *kernel trick*: Noting the mapping $\Phi : \mathbb{R}^n \to \mathcal{H}$, the equality $k(\mathbf{x}^i, \mathbf{x}^j) = \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j)\rangle_{\mathcal{H}^n}$ holds for every kernels. Two classical kernels are 1) the Polynomial kernel $k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{R}} + q)^p$ where $q \in \mathbb{R}^+$, $p \in \mathbb{N}^+$ and 2) the Gaussian kernel $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right)$ where $\sigma \in \mathbb{R}^+$. In remote sensing, Gaussian kernel is typically used and previous work on KPCA have shown it to be appropriated [9].

Table 1: CLASSIFICATION RESULTS FOR THE ROSIS-03 DATA SET.

| | SVM & Linear kernel | | | SVM & Gaussian kernel | | | | |
|---|---|---|---|---|---|---|---|---|
| Feature | Raw | PCA | KPCA | Raw | PCA | KPCA | $EMP_{PCA}$ | $EMP_{KPCA}$ |
| Nb of features | 103 | 3 | 12 | 103 | 3 | 12 | 27 | 108 |
| OA | 76.40 | 78.32 | 78.22 | 79.48 | 78.38 | 79.81 | 92.04 | **96.55** |
| AA | 85.04 | 81.77 | 87.58 | 88.14 | 85.16 | 87.60 | 93.21 | **96.23** |
| $\kappa$ | 68.67 | 71.95 | 72.96 | 74.47 | 72.73 | 74.79 | 89.65 | **95.43** |

Therefore, it was chosen to compute the KPCA in these experiments. From the extracted kernel principal components (KPC), the EMP is constructed as usual: a MP is built for each KPC.

The data sets used in the experiments were two hyperspectral images, one from airborne sensor (ROSIS-03) and one from satellite sensor (HYDICE). The HYDICE data set, cover a high spectral range, from $0.4\mu m$ to $2.4\mu m$ and contains non-Gaussian noise.

The features extracted by the PCA and KPCA are first compared through the Mutual Information (MI). The MI is used to test independence between two variables and intuitively MI measures the information that the two variables share. A MI close to 0 indicates independence, while a high MI indicates dependence and consequently similar information. According to the MI measure, it is found that KCPA is more robust to the noise contains in the HYDICE data. The extracted KPCs have small MI while the PCs have high MI. Although uncorrelated, these features are still dependent. Retaining 95% of the variance as the criterion to select the remaining number of (K)PCs leads to stable number of KPCs, around 10 for each data sets. While with the PCA, for the ROSIS-03 image 3 PCs were selected and for the HYDICE image 40 PCs were selected: The non-Gaussian noise is distributed over several PCs.

The EMP was constructed using (K)PCs corresponding to 95% of the cumulative variance. A circular structuring element with a step size increment of 2 was used. Four openings and closings were computed for each (K)PC, resulting in an EMP of dimension $9 \times m$ ($m$ being the number of retained (K)PCs).

Results for the ROSIS-03 data set are presented in Table 1. The classification were obtained with a linear and non-linear SVM. The best results are obtained with the EMP built with KPCs and clearly outperforms the EMP built with PCs: the overall accuracy is 96.5% for the $EMP_{KPCA}$ and 92.0 % for the $EMP_{PCA}$. The Mc Nemmar test was applied to assess the statistical significance of difference between each result. It is found that KPCA performs significantly better than PCA: For both the linear and non-linear classification of KPCs and PCs and for the construction of the EMP. However, the difference between the non-linear classification of the raw data (the spectral channels) and the KPCs is not significant.

As a conclusion, the KPCA can extract more informative features than PCA for the construction of the EMP. Moreover, KPCA is robust to the noise contains in the hyperspectral remote sensing data. For these reasons, KPCA should be preferably used. Further research are needed to better identify which kernel is the most appropriated for the hyperspectral images.

# References


[1] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, February 2001.

[2] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, September 2003.

[3] P. Soille, *Morphological Image Analysis, Principles and Applications- 2nd edition*, Springer, 2003.

[4] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 466–479, 2005.

[5] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[6] M. Fauvel J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profile," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[7] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, New Jersey, 2003.

[8] B. Schölkopf, A.J. Smola, and K.-R Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.

[9] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for feature reduction in hyperspectral images analysis," in *IEEE $7^{th}$ Nordic Signal Processing Symposium*, 2006.