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

ABSTRACT

Kernel Principal Component Analysis (KPCA) is investigated for feature extraction from hyperspectral remote-sensing data. Features extracted using KPCA are used to construct the Extended Morphological Profile (EMP). Classification results, in terms of accuracy, are improved in comparison to original approach which used conventional principal component analysis for constructing the EMP. Experimental results presented in this paper confirm the usefulness of the KPCA for the analysis of hyperspectral data. The overall classification accuracy increases from 79% to 96% with the proposed approach.

Index Terms— Morphological Profile, hyperspectral data, kernel principal component analysis, SVM.

1. INTRODUCTION

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 $\mathbf{x} \in \mathbb{R}^n$, 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:

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

where $\Sigma_{\mathbf{x}} = \mathbb{E}\left[\mathbf{x}_c \mathbf{x}_c^T\right] \approx \frac{1}{\ell-1} \sum_{i=1}^{\ell} \left(\mathbf{x}^i - \boldsymbol{\mu}_x\right) \left(\mathbf{x}^i - \boldsymbol{\mu}_x\right)^T$ is the empirical covariance estimator of $\mathbf{x}$, $\boldsymbol{\mu}_{\mathbf{x}}$ the empirical mean 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 reviewed in Section 2. Spectral features extraction for the EMP construction is addressed in Section 3. Experiments are presented in Section 4 and conclusions are drawn in Section 5

2. KERNEL PCA

2.1. KPCA algorithm

The main idea of KPCA is simply to map the data onto another space $\mathcal{H}$ before applying the PCA:

$$\begin{array}{rcl} \Phi : \mathbb{R}^n & \to & \mathcal{H} \\ \mathbf{x} & \mapsto & \Phi(\mathbf{x}) \end{array} \qquad (2)$$

where Φ is a function that may be non-linear, and the only restriction on $\mathcal{H}$ is that it must have the structure of a reproducing kernel Hilbert space (RKHS). PCA in $\mathcal{H}$ can be performed as in the input space, but thanks to the kernel trick [9], it can be performed directly in the input space. The KPCA solves the following eigenvalue problem:

$$\lambda \boldsymbol{\alpha} = \mathbf{K} \boldsymbol{\alpha}, \text{ subject to } \|\boldsymbol{\alpha}\|_2 = \frac{1}{\lambda} \qquad (3)$$

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

$$\mathbf{K} = \begin{pmatrix} k(\mathbf{x}^1, \mathbf{x}^1) & \dots & k(\mathbf{x}^1, \mathbf{x}^\ell) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}^\ell, \mathbf{x}^1) & \dots & k(\mathbf{x}^\ell, \mathbf{x}^\ell) \end{pmatrix}. \qquad (4)$$

Table 1. Information classes and training/test samples.

ROSIS-03				HYDICE			
Class		Samples		Class		Samples	
No	Name	Train	Test	No.	Name	Train	Test
1	Asphalt	548	6641	1	Roof	40	3794
2	Meadow	540	18649	2	Road	40	376
3	Gravel	392	2099	3	Trail	40	135
4	Tree	524	3064	4	Grass	40	1888
5	Metal Sheet	265	1345	5	Tree	40	365
6	Bare Soil	532	5029	6	Water	40	1184
7	Bitumen	375	1330	7	Shadow	40	57
8	Brick	514	3682	-	-	-	-
9	Shadow	231	947	-	-	-	-
	Total	3921	42776		Total	280	6929

Table 2. Eigenvalues and cumulative variance in percentage.

Component		1	2	3	4	5
ROSIS-03						
PCA	%	72.8	21.0	01.2	00.8	00.3
	Cum. %	72.8	93.9	98.1	99.0	99.3
KPCA	%	43.9	21.0	15.5	05.2	03.9
	Cum. %	43.9	64.9	80.4	85.6	89.5
HYDICE						
PCA	%	53.4	18.7	03.8	02.0	00.7
	Cum. %	53.4	72.0	75.9.	77.9	78.0
KPCA	%	41.0	20.2	13.7	06.0	05.2
	Cum. %	41.0	61.2	74.9	80.9	86.16

The function k is the core of the KPCA. It is a real-valued positive semi-definite function on $\mathbb{R}^n \times \mathbb{R}^n$ that introduces non-linearity into the processing. It is called a *kernel* and has the following property (kernel trick):

$$k(\mathbf{x}^i, \mathbf{x}^j) = \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}} \quad (5)$$

As with conventional PCA, once (3) has been solved, projection onto the mth component is then performed:

$$\Phi_{kpc}^m(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i^m k(\mathbf{x}^i, \mathbf{x}). \quad (6)$$

Note it is assumed that $\mathbf{K}$ is centered, otherwise it can be centered as [10] ($\mathbf{1}_\ell$ is a square matrix such as $(\mathbf{1}_\ell)_{ij} = \frac{1}{\ell}$):

$$\mathbf{K}_c = \mathbf{K} - \mathbf{1}_\ell \mathbf{K} - \mathbf{K} \mathbf{1}_\ell + \mathbf{1}_\ell \mathbf{K} \mathbf{1}_\ell. \quad (7)$$

2.2. KPCA versus PCA

To better understand the link and the difference between PCA and KPCA, one must note that the eigenvectors of $\Sigma_{\mathbf{x}}$ can be obtained from those of $\mathbf{X}\mathbf{X}^T$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\ell]^T$ [11]. Consider the eigenvalue problem:

$$\gamma \mathbf{u} = \mathbf{X}\mathbf{X}^T \mathbf{u}, \text{ subject to } \|\mathbf{u}\|_2 = 1. \quad (8)$$

The left part is multiplied by $\mathbf{X}^T$ giving

$$\begin{aligned} \gamma \mathbf{X}^T \mathbf{u} &= \mathbf{X}^T \mathbf{X}\mathbf{X}^T \mathbf{u} \\ \gamma \mathbf{X}^T \mathbf{u} &= (\ell-1)\Sigma_{\mathbf{x}} \mathbf{X}^T \mathbf{u} \\ \gamma' \mathbf{X}^T \mathbf{u} &= \Sigma_{\mathbf{x}} \mathbf{X}^T \mathbf{u} \end{aligned} \quad (9)$$

which is the eigenvalue problem (1): $\mathbf{v} = \mathbf{X}^T \mathbf{u}$. But $\|\mathbf{v}\|^2 = \mathbf{u}^T \mathbf{X}\mathbf{X}^T \mathbf{u} = \gamma \mathbf{u}^T \mathbf{u} = \gamma \neq 1$. Therefore, the eigenvectors of $\Sigma_{\mathbf{x}}$ can be computed from eigenvectors of $\mathbf{X}\mathbf{X}^T$ as $\mathbf{v} = \gamma^{-0.5} \mathbf{X}^T \mathbf{u}$. The matrix $\mathbf{X}\mathbf{X}^T$ is equal to:

$$\begin{pmatrix} \langle \mathbf{x}^1, \mathbf{x}^1 \rangle & \ldots & \langle \mathbf{x}^1, \mathbf{x}^\ell \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{x}^\ell, \mathbf{x}^1 \rangle & \ldots & \langle \mathbf{x}^\ell, \mathbf{x}^\ell \rangle \end{pmatrix} \quad (10)$$

which is the kernel matrix with a linear kernel: $k(\mathbf{x}^i, \mathbf{x}^j) = \langle \mathbf{x}^i, \mathbf{x}^j \rangle_{\mathbb{R}^n}$. Using the kernel trick, $\mathbf{K}$ can be rewritten in a similar form as (10):

$$\begin{pmatrix} \langle \Phi(\mathbf{x}^1), \Phi(\mathbf{x}^1) \rangle_{\mathcal{H}} & \ldots & \langle \Phi(\mathbf{x}^1), \Phi(\mathbf{x}^\ell) \rangle_{\mathcal{H}} \\ \vdots & \ddots & \vdots \\ \langle \Phi(\mathbf{x}^\ell), \Phi(\mathbf{x}^1) \rangle_{\mathcal{H}} & \ldots & \langle \Phi(\mathbf{x}^\ell), \Phi(\mathbf{x}^\ell) \rangle_{\mathcal{H}} \end{pmatrix}. \quad (11)$$

From (10) and (11), the advantage of using KPCA comes from an appropriate projection Φ of $\mathbb{R}^n$ onto $\mathcal{H}$: In this space, the data should better match the PCA model. It is clear that the KPCA shares the same properties as the PCA, but in different space.

3. SPECTRAL FEATURE EXTRACTION

3.1. Data sets

Airborne data from the ROSIS-03 (Reflective Optics System Imaging Spectrometer) optical sensor are used for the first experiments. According to specifications, the ROSIS-03 sensor provides 115 bands with a spectral coverage ranging from 0.43 to 0.86μm. The spatial resolution is 1.3 m per pixel. The test set is around the Engineering School at the University of Pavia. It is 610×340 pixels. Twelve channels have been removed due to noise. The remaining 103 spectral channels are processed. All variables have been stretched between 0 and 1. Nine classes of interest are considered: Tree, asphalt, bitumen, gravel, metal sheet, shadow, bricks, meadow, and soil. Available training and test sets are given in Table 1. These are selected pixels from the data by an expert, corresponding to predefined species/classes. Pixels from the training set are excluded from the test set in each case and vice-versa.

Airborne data from the HYDICE sensor (Hyperspectral Digital Imagery Collection Experiment) were used for the second experiments. The HYDICE sensor was used to collect data from flightline over the Washington DC Mall. Hyperspectral HYDICE data originally contained 210 bands in the 0.4-2.4μm region. Channels from near-infrared and infrared wavelengths are known to contain more noise than in channels from visible wavelengths. Noisy channels due to water absorption have been removed and the set consists of 191 spectral channels. The data were collected in August 1995 and each channel has 1280 lines with 307 pixels each. Seven information classes were defined, namely: Roof, road, grass, tree, trail, water and shadow.

3.2. Feature extraction

Solving the eigenvalue problems (1) and (3) yields the results reported in Table 2. A Gaussian kernel was used for

Table 3. Classification results for the University Area data set.

Features	Nb. of Features	OA	AA	κ	1	2	3	4	5	6	7	8	9
Raw	103	79.5	88.1	74.5	84.4	66.2	72.0	98.0	99.5	93.1	91.2	92.2	96.6
EMP$_{\text{PCA}}$	27	92.0	93.2	89.6	94.6	88.8	73.1	99.2	99.5	95.2	98.87	99.1	90.0
EMP$_{\text{KPCA}}$	108	96.5	96.2	95.4	96.23	97.6	83.6	99.4	99.5	92.9	99.1	99.5	98.3

Table 4. Classification results for the HYDICE data set.

Features	Nb. of Features	OA	AA	κ	1	2	3	4	5	6	7
Raw	191	98.1	96.9	97.4	97.05	98.1	100	100	98.0	99.6	85.6
EMP$_{\text{PCA}}$	360	98.6	98.0	98.0	97.5	95.5	100	100	99.51	99.92	89.7
EMP$_{\text{KPCA}}$	99	98.7	99.4	98.1	97.5	98.8	100	100	99.5	99.5	100

(a) PCA (b) KPCA
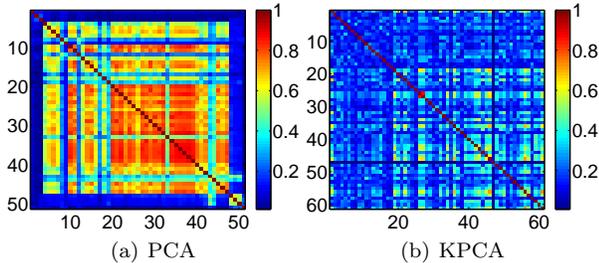
Fig. 1. Mutual Information matrices for the HYDICE data.

the KPCA, its width has been set to 4 and 5000 randomly-selected samples were used to compute the kernel matrix.

Looking at the cumulative eigenvalues, three principal components (PCs) reach 95% of total variance for the ROSIS-03 data, while 12 KPCs are needed to achieve 95%. On the contrary, for the HYDICE data, 40 PCs are needed to reach 95% of total variance and only 11 KPCs. Acquired from a higher range of wavelengths, more noise is contained in the data and more bands were removed by comparison to the ROSIS data. That explains why more PCs than KPCs are needed to reach 95% of the cumulative variance. That may be also an indication that more information is extracted and the KPCA is more robust to non-Gaussian noise, since a reasonable number of features are extracted from the HYDICE data set.

To test this assumption, the Mutual Information (MI) between each (K)PC has been computed. The classical correlation coefficient was not used since the PCA is optimal for that criterion. For comparison, the normalized MI was computed ($\text{MI}_n(\mathbf{x}, \mathbf{y}) = \frac{\text{MI}(\mathbf{x}, \mathbf{y})}{\sqrt{\text{MI}(\mathbf{x}, \mathbf{x})}\sqrt{I(\mathbf{y}, \mathbf{y})}}$). The MI is used to test independence between two variables and intuitively the 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. Fig. 1 presents the MI matrices, which represent the MI for each pair of extracted features with both PCA and KPCA, for the HYDICE data set. From Fig. 1(a), PCs number 4 to 40 contain more or less the same information since they correspond to a high MI. Although uncorrelated, these features are still dependent. This phenomenon is due to the noise contained in the data which is not Gaussian [7] and is distributed over several PCs. From Fig. 1, KPCA is less sensitive to the noise, *i.e.*, in the feature space the data match better the PCA model and the noise tends to be Gaussian. Note that with KPCA, only the first 11 KPCs are retained against 40 with conventional PCA.

4. EXPERIMENTS

For the experiments, the EMP was constructed using a number of (K)PCs corresponding to 95% of the total variance. The EMP constructed with the PCA (respectively KPCA) is noted EMP$_{\text{PCA}}$ (respectively EMP$_{\text{KPCA}}$). A circular structuring element with step size increment of 2 was used. Four openings and closings were computed for each (K)PC, resulting in an EMP of dimension $9 \times p$ (p being the number of retained (K)PCs).

The classification was done by an *one-versus-one* multi-class SVM with a Gaussian kernel. Optimal parameters were selected using 5-fold cross-validation. The classification accuracy was assessed with:

- An overall accuracy (OA) which is the number of well classified samples divided by the number of test samples.
- An average accuracy (AA) which represents the average of class classification accuracy.
- A kappa coefficient of agreement (κ) which is the percentage of agreement corrected by the amount of agreement that could be expected due to chance alone [12].
- A class accuracy which is the percentage of correctly classified samples for a given class.

These criteria were used to compare classification results and were computed using a confusion matrix. Furthermore, the statistical significance of differences was computed using Mc-Nemar's test, which is based upon the standardized normal test statistic [13]:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \qquad (12)$$

where f_{12} indicates the number of samples classified correctly by classifier 1 and incorrectly by classifier 2. The difference in accuracy between classifiers 1 and 2 is said to be statistically significant if $|Z| > 1.96$. The sign of Z indicates whether classifier 1 is more accurate than classifier 2 ($Z > 0$) or vice-versa ($Z < 0$). This test assumes that the training and the test samples are related and is thus adapted to the analysis since the training and test sets were the same for each experiment for a given data set.

For the ROSIS-03 data, the results are reported in Table 3 and the Z tests state that the differences of classification accuracies are statistically significant. EMP constructed with either PCs or KPCs outperformed the standard approach in classification. The κ is increased by 15 % with EMP$_{\text{PCA}}$ and by 20% with EMP$_{\text{KPCA}}$. The statistical difference of accuracy $Z_{\text{KPCA\_PCA}} = 35.33$ clearly demonstrates the benefit of
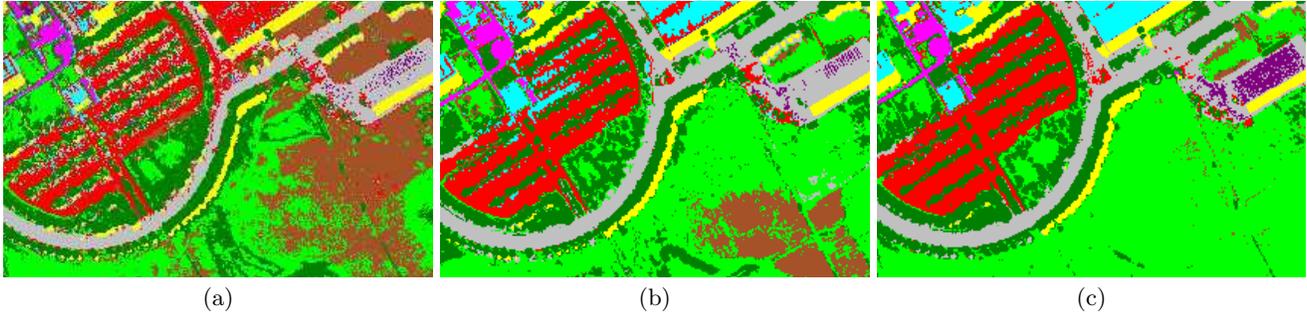
Fig. 2. Thematic map obtained with the University Area: (a) Raw data, (b) EMP$_{PCA}$, (c) EMP$_{KPCA}$. The classification was done by SVM with a Gaussian kernel. The color-map is as follows asphalt, meadow, gravel, tree, metal sheet, bare soil, bitumen, brick and shadow.

using the KPCA rather than the PCA.

Regarding the class accuracy, the highest improvements were obtained for class 1 (*Asphalt*), class 2 (*Meadow*) and class 3 (*Gravel*). For these classes, the original spectral information was not sufficient and the morphological processing provided additional useful information. However, using KPCA for feature extraction helps in classifying better these classes.

Thematic maps obtained with the SVM applied to the Raw data, EMP$_{PCA}$ and EMP$_{KPCA}$ are reported in Fig. 2. For instance, it can be seen that the region corresponding to class 2, *meadow*, are more homogeneous in the image Fig.2.(c) than in the two other images.

For the HYDICE data, the results are reported in Table 4. From the global accuracies, all the different approaches perform similarly. It is confirm with the Z tests, which indicate that the differences in term of classification are not statistically significant. Regarding the class accuracies, the class 7, *shadow*, is perfectly classified only by EMP$_{KPCA}$.

5. CONCLUSIONS

This paper presents a KPCA-based method with application to the analysis of hyperspectral remote sensing data: The construction of the EMP with KPCs is investigated. Comparison to PCA in terms of classification accuracies demonstrates the usefulness of this approach.

From this study we conclude that KPCA should be preferred to standard PCA whenever possible. However one limitation of the KPCA is its computational complexity, related to the size of the kernel matrix, which can limit the number of samples used.

Our current investigations are oriented to non-linear independent component analysis, such as kernel ICA for the construction of the EMP and to a sparse KPCA in order to reduce the complexity.

6. REFERENCES

[1] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, 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 Trans. Geosci. Remote Sens.*, 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 Trans. Geosci. Remote Sens.*, 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 Trans. Geosci. Remote Sens.*, 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 Trans. Geosci. Remote Sens.*, 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] N. Aronszajn, "Theory of reprodusing kernel," Tech. Rep. 11, Harvard University, Division of engineering sciences, 1950.

[10] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support vector machines, regularization, optimization, and beyond*, MIT Press, 2002.

[11] J. A. Lee and M. Verleysen, *Nonlinear dimensionality reduction*, Springer, 2007.

[12] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*, Springer, 1999.

[13] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Engineering & Remote Sensing*, vol. 70, no. 5, pp. 627–633, May 2004.