



HAL
open science

Randomized Nonlinear Component Analysis for Dimensionality Reduction of Hyperspectral Images

Bharath Bhushan Damodaran, Nicolas Courty, Romain Tavenard

► **To cite this version:**

Bharath Bhushan Damodaran, Nicolas Courty, Romain Tavenard. Randomized Nonlinear Component Analysis for Dimensionality Reduction of Hyperspectral Images. IGARSS 2017 - IEEE International Geoscience and Remote Sensing Symposium, Jul 2017, Houston, United States. pp.1-4. hal-01620604

HAL Id: hal-01620604

<https://hal.science/hal-01620604v1>

Submitted on 20 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RANDOMIZED NONLINEAR COMPONENT ANALYSIS FOR DIMENSIONALITY REDUCTION OF HYPERSPECTRAL IMAGES

Bharath Bhushan Damodaran¹, Nicolas Courty¹, Romain Tavenard²

¹Univ. Bretagne-Sud, UMR 6074, IRISA, F-56000 Vannes, France

²LETG-Rennes COSTEL/IRISA-Univ.Rennes-2

bharath-bhushan.damodaran@irisa.fr

ABSTRACT

Kernel based feature extraction method overcomes the curse of dimensionality and captures the non-linearities present in the data. However, these methods are not scalable with large number of pixels found with hyperspectral images. Thus, a small subset of pixels are randomly selected to make the solution of kernel based methods tractable. In this paper, we propose scalable nonlinear component analysis for dimensionality reduction of hyperspectral images. The proposed method relies on the randomized feature maps to capture the non-linearities between the variables in the hyperspectral data. Experiments conducted with three hyperspectral datasets show that our proposed method has provided better quality components and outperformed the state-of-the-art in terms of classification performance.

Index Terms— Hyperspectral image classification, Kernel PCA, Kernel MNF, Nonlinear component analysis, Random Fourier feature, Kernel approximation

1. INTRODUCTION

The automatic analysis of hyperspectral images is a challenging task mainly due to the high dimensionality of the data [1]. Dimensionality reduction methods can overcome the Hughes phenomena by projecting the high dimensional data into a lower dimensional subspace without losing significant amount of information. Principal component analysis (PCA) and minimum

noise fraction (MNF) are widely used to reduce the dimension of the hyperspectral data due to its simplicity and its demonstrated performance in real applications. These methods capture only the linear relationship between the variables in the data and it does not take into the account of non-linearities present in the data. In order to overcome these limitations their non-linear extensions have been proposed using kernel methods.

For PCA, the non-linear extension is Kernel principal component analysis (KPCA), in which mapping function is used to map the data in reproducing kernel Hilbert space (RKHS) space and PCA is applied in RKHS space [2]. For MNF [3], the non-linear extension is kernel minimum noise fraction (KMNF) [4] and it is also computed similar to KPCA. However these extensions tend to have high computational complexity and it is often cubic in the sample size. It is worth to note that in remote sensing images the available number of pixels are very large, so often it is impossible to directly apply the KPCA or KMNF for hyperspectral data. Thus in order to make tractable, a small subset of samples is selected to construct the kernel matrix [2, 4]. However obtained solutions by these techniques are sub-optimal, since it only includes small amount of samples and it might fail to incorporate information of the objects which have less-spatial extend, which are crucial for target detection applications.

Recently, randomized feature maps have been introduced to construct features that can help to reveal the non-linear relationships in the data [5]. These methods find an explicit low dimensional feature map to approximate the exact kernel matrix with little or no loss in approximation error. The advantage of these techniques is it achieves drastic saving in the computational complexity from cubic to linear in the sample size, which

This work was supported by the French Agence Nationale de la Recherche (ANR) under the reference ANR-13-JS02-0005-01 (As-tex project) and by the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n. PCOFUND-GA-2013-609102, through the PRESTIGE programme coordinated by Campus France.

is highly favored for high dimensional remote sensing images. The approximation to the KPCA using randomized feature maps is recently proposed in machine learning [6]. However their potential on the hyperspectral images is not studied yet. Thus, the first objective of this paper is to investigate the potential of randomized PCA for the hyperspectral data. The second objective is to propose the randomized MNF for dimension reduction of hyperspectral data. We demonstrate the effectiveness of the proposed method based on visual analysis and classification experiments on the several real world hyperspectral datasets.

2. KERNEL APPROXIMATION AND NONLINEAR COMPONENT ANALYSIS

2.1. Random Fourier Features

The objective of the kernel approximation methods is to find a low dimensional feature map to approximate the exact kernel [5]. More specifically, given the kernel $\mathbf{K}(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, kernel approximation seeks to find a nonlinear transformation ϕ such that, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) \simeq \phi(\mathbf{x})^T \phi(\mathbf{y}) \quad (1)$$

Random Fourier Features (RFF) is used to approximate the shift invariant kernels. By Bochner's theorem, the Fourier transform of the shift invariant kernels is a non-negative measure [5], then

$$\begin{aligned} \mathbf{K}(\mathbf{x} - \mathbf{y}) &= \int p(\mathbf{w}) e^{-j\mathbf{w}^T(\mathbf{x}-\mathbf{y})} d\mathbf{w} \\ &\simeq \frac{2}{D} \sum_{i=1}^B \cos(\mathbf{w}_i^T \mathbf{x} + b_i) \cos(\mathbf{w}_i^T \mathbf{y} + b_i) = \mathbf{z}^T \mathbf{z} \end{aligned}$$

where

$$\mathbf{z} = \sqrt{\frac{2}{D}} [\cos(\mathbf{w}_1^T \mathbf{x} + b_1), \dots, \cos(\mathbf{w}_D^T \mathbf{x} + b_D)]. \quad (2)$$

The number of random Fourier features used in the experiment is $2d$, where d is the number bands in the hyperspectral image. The Gaussian RBF kernel can be approximated by sampling \mathbf{w} from $N(0, \sigma^{-2})$ ¹ and b from uniform distribution. Thus \mathbf{z} can be used as explicit

¹The σ^2 is estimated based on the mean of the pairwise distance of between the samples.

lower dimensional feature map. This explicit feature map is used to build scalable kernel component analysis in this paper.

2.2. Scalable nonlinear component analysis

In this subsection, we propose the scalable nonlinear component analysis using random Fourier features.

2.2.1. Randomized PCA

Let $X \in \mathbb{R}^{d \times N}$, be the input hyperspectral image, where N is the number of pixels in the image and $W \in \mathbb{R}^{d \times 2d}$ be the RFF coefficients, and the randomized Fourier feature maps $Z(X)$ can be obtained by (2). The Randomized PCA (RPCA) can be computed in two steps: (1) compute randomized non-linear mapping of the data by $Z(X)$, (2) perform the linear PCA on the step 1.

$$\begin{aligned} \text{RPCA}(\mathbf{X}) &= \mathbf{F}^T \mathbf{Z} \\ &= \text{PCA}(\mathbf{Z}(\mathbf{X})) \simeq \text{KPCA}(\mathbf{X}) \end{aligned} \quad (3)$$

where \mathbf{F} is the principal component coefficients computed on randomized feature maps. The RPCA can be viewed as the low rank approximation of the KPCA.

2.2.2. Randomized MNF

Randomized MNF (RMNF) is also the low rank approximation of the KMNF. The computation of RMNF is similar to RPCA. Firstly, transform the data using RFF and then perform linear MNF on the randomized features. The computation of RMNF is given as

$$\begin{aligned} \text{RMNF}(\mathbf{X}) &= \mathbf{Y}^T \mathbf{Z} \\ &= \text{MNF}(\mathbf{Z}(\mathbf{X})) \simeq \text{KMNF}(\mathbf{X}) \end{aligned} \quad (4)$$

where \mathbf{Y} is the eigen vectors of the product of noise covariance matrix and inverse of co-variance matrix of \mathbf{Z} . For more details of MNF transform, please see [3]

3. EXPERIMENTAL RESULTS

3.1. Experimental Datasets

In order to evaluate our proposed method, we have considered three hyperspectral datasets namely, Pavia University, Pavia Centre and Houston datasets. The first two datasets is captured by the ROSIS sensor over the Pavia University and City centre, and it has 103 and 102 bands respectively. The last dataset was acquired by the NSF-funded Center for Airborne Laser Mapping (NCALM)

over the University of Houston campus and the neighboring urban area. The Pavia University, Pavia Centre dataset consist of nine land cover classes and Houston consists of fifteen land cover classes, and the available labeled reference samples are 42776, 148152, and 13997 respectively.

3.2. Visual analysis

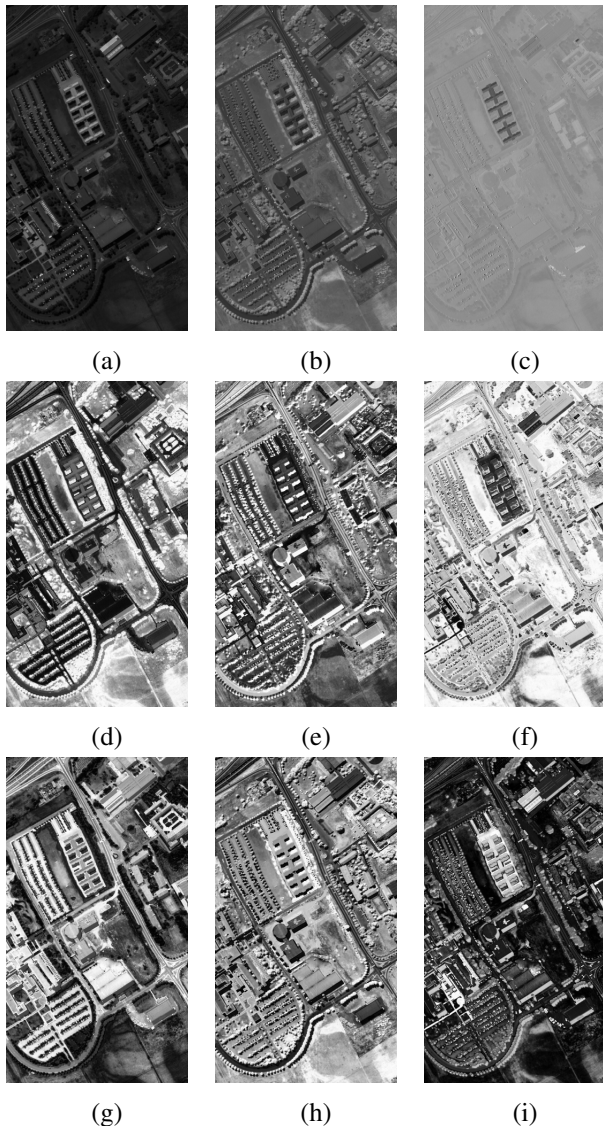


Fig. 1: First three components on Pavia University dataset: (a-c) Linear PCA, (d-f) RPCA, (g-i) KPCA.

Figure 1 shows the first three components of linear PCA, RPCA and KPCA ² of Pavia University. From the

²For the KPCA and KMNF, the kernel matrix are constructed

visual inspection, the components produced by RPCA has better quality than KPCA and PCA. The first three components of PCA accounts for 98.4% of variance where as KPCA and RPCA requires 20 and 30 components to accounts for the same variance. This shows that the lower order components from RPCA contains more useful information. Furthermore, the first three components ³, of the proposed RMNF and KMNF are shown in Figure 2. As observed with Figure 1 our proposed RMNF has more informative content and better quality components. In the next subsection, we evaluate the randomized non-linear components in the classification framework.

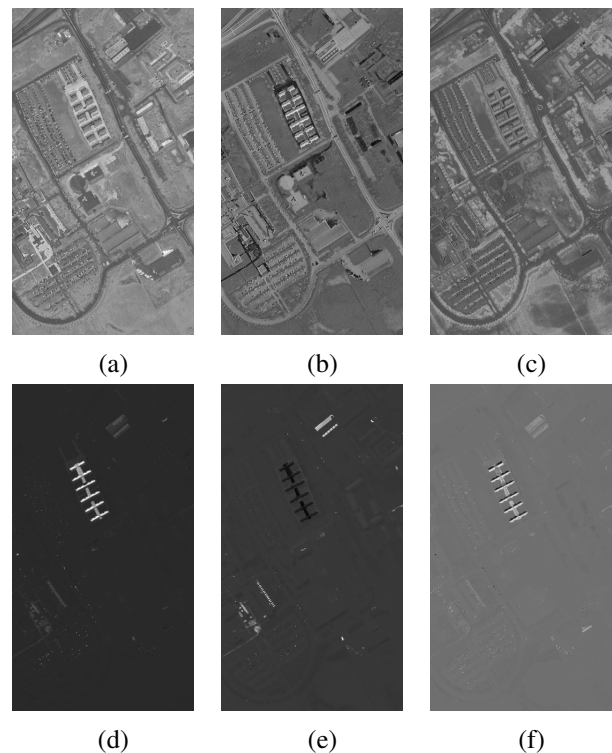


Fig. 2: First three components on Pavia University dataset: (a-c) RMNF, (d-f) KMNF.

3.3. Classification experiments

3.3.1. Randomized PCA and Kernel PCA

Table 1 shows the classification results of randomized PCA and kernel PCA over different number of components using 1000 randomly selected samples, The bandwidth of the RBF kernel is estimated similar to RFF

³Due to the space restrictions, the first three components of linear MNF is not shown here

nents with three hyperspectral datasets. The classification experiments are performed using SVM classifier using RBF kernel with 100 training samples per class. The hyperparameters of the SVM classifier are automatically tuned using grid search method with five-fold cross validation. From the table 1, it is obvious that accuracies of randomized PCA is better than or comparable to the conventional kernel PCA. The significance of performance depends on the datasets and also on the number of components used.

Table 1: Classification accuracies (OA in %) of RPCA and KPCA with different number of components for three hyperspectral datasets. The reported accuracies are averaged over five runs.

Dataset	Method	No of components							
		3	5	10	15	20	25	30	35
Pavia University	RPCA	64.9	68.8	75.0	78.6	78.7	82.2	81.9	83.7
	KPCA	64.9	68.4	75.5	77.3	77.2	79.1	80.2	81.0
Pavia Centre	RPCA	91.8	96.0	96.6	96.9	97.0	97.2	97.2	97.5
	KPCA	90.3	96.1	96.7	96.9	96.8	97.1	97.2	97.3
Houston	RPCA	80.0	86.7	88.4	91.0	92.0	92.6	92.9	93.4
	KPCA	79.8	84.1	87.8	89.3	91.2	92.6	93.3	93.1

3.3.2. Randomized MNF and Kernel MNF

Table 2 shows the classification results of the proposed randomized MNF and kernel MNF with three hyperspectral datasets and the values shows the proposed RMNF has outperformed the conventional KMNF. More specifically, the RMNF has much better performance than the KMNF when upto 10 components are considered, and the improvement is about 5% on average with all the datasets.

Table 2: Classification accuracies (OA in %) of RMNF and KMNF for three hyperspectral datasets. The reported accuracies are averaged over five runs.

Dataset	Method	No of components							
		3	5	10	15	20	25	30	35
Pavia University	RMNF	61.0	70.1	77.1	78.5	81.6	82.9	82.7	83.0
	KMNF	58.0	70.4	74.7	78.1	79.7	80.8	81.9	82.8
Pavia Centre	RMNF	84.3	95.0	96.6	96.8	96.9	97.0	97.1	97.0
	KMNF	82.6	86.7	92.9	94.2	96.5	96.6	96.7	97.0
Houston	RMNF	71.6	83.3	88.8	90.2	91.4	91.8	92.2	92.3
	KMNF	57.8	71.9	85.6	90.6	92.6	93.2	94.0	94.4

4. CONCLUSION

In this paper, we proposed scalable nonlinear component analysis for dimensionality reduction of hyperspectral data using randomized feature maps. More specifically, we have investigated randomized PCA and proposed randomized MNF to capture the non-linear dependence between the variables in the data and to overcome the limitations of the conventional kernel component analysis for hyperspectral image classification. Experiments conducted with three hyperspectral datasets show that the proposed method provided informative components and outperformed the classification results of the conventional methods. In the future work, we would like to study the impact of the number of randomized feature maps and the Gaussian bandwidth parameter on the proposed methods.

Acknowledgment

The authors would like to thank Prof. P. Gamba and 2013 IEEE GRSS Data Fusion contest for providing ROSIS and Houston images respectively.

5. REFERENCES

- [1] B.B. Damodaran, R.R. Nidamanuri, and Y. Tarabalka, "Dynamic Ensemble Selection Approach for Hyperspectral Image Classification With Joint Spectral and Spatial Information," *IEEE JSTARS*, vol. 8, no. 6, pp. 2405–2417, jun 2015.
- [2] Mathieu Fauvel, Jocelyn Chanussot, and Jon Atli Benediktsson, "Kernel Principal Component Analysis for the Classification of Hyperspectral Remote Sensing Data over Urban Areas," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 783194, 2009.
- [3] 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 TGRS*, vol. 26, no. 1, pp. 65–74, 1988.
- [4] Allan Aasbjerg Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations.," *IEEE tran. on image proc.*, vol. 20, no. 3, pp. 612–24, mar 2011.
- [5] Ali Rahimi and Benjamin Recht, "Random features for largescale kernel machines," in *Advances in neural information processing systems*, 2007, pp. 1177–1184.
- [6] David Lopez-Paz, Suvrit Sra, Alexander J. Smola, Zoubin Ghahramani, and Bernhard Scholkopf, "Randomized Nonlinear Component Analysis," in *31st International Conference on Machine Learning*, 2014, vol. 32, p. JMLR: W&CP volume 32.