

MORPHOLOGICAL IMAGE DISTANCES FOR HYPERSPECTRAL DIMENSIONALITY EXPLORATION USING KERNEL-PCA AND ISOMAP

S. Velasco-Forero, J. Angulo

J. Chanussot

CMM-Centre de Morphologie Mathématique
MINES Paristech, FRANCE

{santiago.velasco;jesus.angulo}@ensmp.fr

GIPSA-Lab
Grenoble Institute of Technology, FRANCE

jocelyn.chanussot@gipsa-lab.inpg.fr

1. INTRODUCTION: MOTIVATION AND AIM

The application of nonlinear manifold learning for hyperspectral image analysis has been widely studied in last years [1, 4]. One of the main ingredients of these data reduction techniques is the distance used to compare the spectral band images. By means of this distance the pairwise similarity matrix is built and then, the matrix is used to explore the intrinsic dimensionality of the hyperspectral image.

There are two main families of image distances which have been considered in previous works: i) the distance between the pixels using Minkowski metrics, such as the Euclidean distance or the L_1 distance; ii) the distances between the image histograms, such as the Kullback-Leibler Divergence [8] or the chi-squared distance [2]. By comparing pairwise pixels, Minkowski metrics take into account the spatial structure, however they can be very sensitive to acquisition noise of different spectral bands (L_1 is more robust to noise than L_2) or to slight spatial shifts between the different bands. Histogram distances are robust to both intensity and spatial variations but they do not consider the spatial structure. A recent work [12] has considered this lack of spatial coherence in the classical distances for nonlinear data reduction by replacing the pixelwise Euclidean distance by the Euclidean distance between neighbor patches ($n \times n$ surrounding pixels centred at each pixel).

The aim of this paper is to propose two new families of spatial image distances for spectral band comparison. Both are based on notions from mathematical morphology [14], a nonlinear image processing methodology based on the application of lattice theory to spatial structures. The first distance is based on the formulation using morphological dilations of Hausdorff distance for gray-scale images [15]. The second distance is more original and it is founded in the leveling operator [11]. Levelings are geodesic filters which modify, without blurring the contours, one of the images according to the other image. The application of these morphological distances for hyperspectral dimensionality exploration is illustrated with two powerful nonlinear data analysis techniques: Kernel-PCA and ISOMAP. Using standard image examples,

their performance is studied in comparison with other image distances such as Euclidean distance and KL-divergence.

Scalars are denoted by lower case letters (a, b, \dots), vectors by bold lower case letters ($\mathbf{a}, \mathbf{b}, \dots$), images and matrices by bold upper-case letters ($\mathbf{X}, \mathbf{Y}, \dots$), and hyperspectral images by calligraphic upper-case letters ($\mathcal{X}, \mathcal{Y}, \dots$).

2. REMAINDER ON KERNEL-PCA AND ISOMAP

2.1. Kernel PCA

Kernel PCA (KPCA) is a nonlinear generalization of Principal Component Analysis. By choosing nonlinear kernels adapted to data nature, the dimensionally reduction is usually stronger than using linear PCA. In practice, classical and effective kernels are based on a function of a distance between the points, such as the radial basis function kernel. The original theoretical foundation may be found [12] [13]. In our case, the distance corresponds to the distance between the two spectral bands. The classical formulation of KPCA is adapted for a set of images and focus on pairwise distance matrix. Given a collection of images $\{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_k\}$, KPCA solves the following eigenvalue problem:

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

where \mathbf{K} is a square Gram matrix ($k \times k$) containing pairwise kernel distance. Using for instance the gaussian kernel, we have:

$$\mathbf{K}(i, j) = \kappa(\mathbf{I}_i, \mathbf{I}_j) = \exp\left(-\frac{\|\mathbf{I}_i - \mathbf{I}_j\|^2}{2\sigma^2}\right) \quad (1)$$

Note that it is assumed that \mathbf{K} is centered, otherwise it can be center using:

$$\mathbf{K}_c = \mathbf{K} - \mathbf{1}_k \mathbf{K} - \mathbf{K} \mathbf{1}_k + \mathbf{1}_k \mathbf{K} \mathbf{1}_k$$

where $\mathbf{1}_k$ is a constant square matrix with value $1/k$.

2.2. Isomap

Isometric feature mapping (Isomap) [16] is a method for estimating the intrinsic geometry of a data manifold based on a rough estimate of each data point neighbors on the manifold. More precisely, it is a low-dimensional embedding approach based on geodesic distances on a weighted neighborhood graph and multidimensional scaling (MDS). Again a distance between the spectral bands is needed to build the neighborhood graph.

Given a collection of images $\{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_k\}$, Isomap algorithm consists of three steps:

1. *Neighborhood graph*: Fix either an integer m and given a distance matrix $\mathbf{D}_{k \times k}$. Determine which images are "neighbors" by connecting each image either to its m nearest neighbors. This gives us a weighted neighborhood graph $\mathcal{G} = \mathcal{G}(\mathcal{V}, \mathcal{E})$, where the set of vertices $\mathcal{V} = \{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_k\}$ are the input image; and the set of edges $\mathcal{E} = \{e_{ij}\} = \{d_{ij}^m\}$ indicate the neighborhood relationships between the images. The value d_{ij}^m contained in row i and column j of the distance matrix \mathbf{D} , if \mathbf{I}_j is in the m -nearest neighbors of \mathbf{I}_i and 0 otherwise.
2. *Compute graph distances*: Compute the shortest path between every pair of vertices (i, j) denoted as $d^{\mathcal{G}}(i, j)$ in the graph \mathcal{G} .
3. *Embedding via multidimensional scaling*: Let $\mathbf{D}^{\mathcal{G}} = (d_{ij}^{\mathcal{G}})$ be the symmetric matrix of geodesic distances. Apply classical multidimensional scaling to reconstructed points in a k' -dimensional feature space, such that the geodesic distances between images is preserved as much as possible.

3. CLASSICAL DISTANCES

In KPCA and Isomap presented in section 2, it is necessary to obtain the distance matrix $\mathbf{D}_{k \times k}$ to produce the low-dimensional representation. Given a hyperspectral image $\mathcal{I}_{n_1 \times n_2 \times k}$ with n_1 rows, n_2 columns, and k bands our approach consider \mathcal{I} as a collection of images $\mathcal{I} = [\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_k]$, where each \mathbf{I}_j is a grey scale image, for all $j = 1, \dots, k$. Classical KPCA uses Euclidean distance to calculate the distance matrix \mathbf{D} , defined as $D_E(\mathbf{I}_i, \mathbf{I}_j)(i, j) = \|\sum (\mathbf{I}_i - \mathbf{I}_j)^2\|$ where summation is over the $n_1 \times n_2$ elements of the matrix.

The histogram operator denoted by $HIST_{\mathbf{b}}(\mathbf{I})$, calculate a classical histogram based on a bins vector \mathbf{b} and normalized for the total number of pixel producing a probability vector. Thus, $HIST_{\mathbf{b}}(\mathbf{I})$ is a sum-one vector of probability with the same dimension that \mathbf{b} . In this paper, two important distance for probability distribution are considered: Bhattacharyya distance and Kullback-Leibler divergence (KLD) distance. They had been used in pattern recognition, computer vision

and remote sensing scenarios for histogram matching [5] [7]. Introducing the following notation allows more succinct expression for a pair of grey scale image \mathbf{I}_1 and \mathbf{I}_2 : $\mathbf{p} = HIST_{\mathbf{b}}(\mathbf{I}_1)$, and $\mathbf{q} = HIST_{\mathbf{b}}(\mathbf{I}_2)$. The Bhattacharyya measure (or coefficient)[3] was formulated as the square of the angle between two position vectors \mathbf{p} and \mathbf{q} .

$$DB(\mathbf{I}_1, \mathbf{I}_2) = -\ln(BC(\mathbf{p}, \mathbf{q})) \quad (2)$$

where $BC(\mathbf{p}, \mathbf{q}) = \sum \sqrt{\mathbf{p}\mathbf{q}}$. The Kullback-Leibler divergence (KLD) is defined as:

$$KL(\mathbf{I}_1|\mathbf{I}_2) = \sum \mathbf{p}_j \log \left(\frac{\mathbf{p}_j}{\mathbf{q}_j} \right),$$

when $\mathbf{p}_j, \mathbf{q}_j \neq 0$. Then the Kullback-Leibler divergence distance is defined by

$$KLD(\mathbf{p}, \mathbf{q}) = \frac{KL(\mathbf{p}|\mathbf{q}) + KL(\mathbf{q}|\mathbf{p})}{2} \quad (3)$$

4. MULTI-SCALE HAUSDORFF DISTANCES

The well known Hausdorff distance is a natural metric for comparing sets (i.e., binary images). Its extension to scalar functions (i.e., gray level images) allows comparing also the image structures of spectral bands. The value of distance represents the "size" of the dilation in such a way that one image covers totally the other. In addition, before computing the Hausdorff distance, the images can be simplified by removing the structures of a certain scale or size and then compute the Hausdorff distance. In such a case, a value of Hausdorff distance is obtained for each scale and the final global distance can be obtained as the sum or the max of the different scale distance values. The formulation for Hausdorff distance in binary image is given by [15]:

$$H(\mathbf{B}_i, \mathbf{B}_j) = \inf\{\lambda : \mathbf{B}_i \subseteq \delta_{\lambda}(\mathbf{B}_j); \mathbf{B}_j \subseteq \delta_{\lambda}(\mathbf{B}_i)\} \quad (4)$$

where δ_{λ} is the dilation by the compact ball with radius λ .

Our proposal to extend this version to grey scale image starts applying a level set decomposition denoted by $LS(\mathbf{I})$ defined as:

$$LS(\mathbf{I}, \mathbf{s}) : \mathbf{I} \rightarrow [\mathbf{B}_{\mathbf{I}}^{(s_1)}, \mathbf{B}_{\mathbf{I}}^{(s_2)}, \dots, \mathbf{B}_{\mathbf{I}}^{(s_S)}] \quad (5)$$

where $\mathbf{B}_{\mathbf{I}}^{s_i}$ is a binary matrix produced by thresholding in the original grey scale image \mathbf{I} at level s_i , defined by:

$$\mathbf{B}_{\mathbf{I}}^{(s_i)}(x, y) = \begin{cases} 1, & I(x, y) \geq s_i; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Using 4 and 5, the Hausdorff distance for grey scale image can be defined as:

$$H(\mathbf{I}_i, \mathbf{I}_j) = \max_k \{H(LS(\mathbf{I}_i, s_k), LS(\mathbf{I}_j, s_k))\} \quad (7)$$

where distance for LS meaning apply Hausdorff distance between correspondent binary images.

5. LEVELING-BASED MORPHOLOGICAL DISTANCES

A leveling filter has two input images: the reference image and the marker image (which is generally a rough simplification of the reference image), and it simplifies textures and eliminates small details of the reference image according to the marked structures, but preserving the contours of remaining objects. In fact, the leveling is obtained by iteration of geodesic dilations and erosions until idempotence [11]. Consequently, besides the final leveled image, a series of images (successive modifications of marker to approach the reference) is obtained.

Given two spectral images, two associated leveling images (and their corresponding intermediate image series) are obtained, depending on which is used as reference and which as marker. Instead of computing the distance between original spectral images, the basic idea of the new distance is to calculate a spatial distance (using for instance the Euclidean distance) between both mutually leveled images.

In comparison with standard Minkowski metrics, the leveling-based distances take into account the structural difference of both images and it is more robust against effects owing to noise or spatial shifts.

5.1. Morphological Levelings

Let $\mathbf{I} = f(x, y)$ be a reference image and a marker image $\mathbf{M} = m(x, y)$ both are grey-scale images, where the marker determines the structures to be reconstructed and those to be simplified during the geodesic erosion/dilation procedure. The leveling $\Lambda(\mathbf{I}, \mathbf{M})$ can be obtained by the following iterative algorithm:

$$\Lambda(\mathbf{I}, \mathbf{M}) = \lambda^\infty(\mathbf{I}, \mathbf{M}) = [\mathbf{I} \wedge \delta_{\mathbf{I}}^i(\mathbf{M})] \vee \varepsilon_{\mathbf{I}}^i(\mathbf{M})$$

where $\lambda^\infty(\mathbf{I}, \mathbf{M}) = \lambda^i(\mathbf{I}, \mathbf{M})$ such that $\lambda^i(\mathbf{I}, \mathbf{M}) = \lambda^{i+1}(\mathbf{I}, \mathbf{M})$ (convergence until idempotency).

The geodesic dilation of size i is given by $\delta_{\mathbf{I}}^i(\mathbf{M}) = \delta_{\mathbf{I}}^1 \delta_{\mathbf{I}}^{i-1}(\mathbf{M})$ where the unitary conditional dilation is $\delta_{\mathbf{I}}^1(\mathbf{M}) = \delta_B(\mathbf{M}) \wedge \mathbf{I}$ with $\delta_B(\mathbf{M})$ an unitary dilation. The geodesic erosion of size i , $\varepsilon_{\mathbf{I}}^i(\mathbf{M})$ can be obtained by the property of duality by complement $\varepsilon_{\mathbf{I}}^i(\mathbf{M}) = [\delta_{\mathbf{I}^c}^i(\mathbf{M}^c)]^c$, where \mathbf{I}^c is the negative of image \mathbf{I} . Thus leveling operator Λ for an original image \mathbf{I}_2 and a destination image \mathbf{I}_1 produce a sequence of grey scale images $\{\lambda^1(\mathbf{I}_1, \mathbf{I}_2), \lambda^2(\mathbf{I}_1, \mathbf{I}_2), \dots, \lambda^\infty(\mathbf{I}_1, \mathbf{I}_2)\}$ with the property that $\|\lambda^i(\mathbf{I}_1, \mathbf{I}_2) - \mathbf{I}_2\|^2 \leq \|\lambda^{i+1}(\mathbf{I}_1, \mathbf{I}_2) - \mathbf{I}_2\|^2$ and $\|\lambda^i(\mathbf{I}_1, \mathbf{I}_2) - \mathbf{I}_1\|^2 \geq \|\lambda^{i+1}(\mathbf{I}_1, \mathbf{I}_2) - \mathbf{I}_1\|^2$ for all i . Note that ∞ is denoting convergence in leveling operator. Figure 1 sketches the sequences produced per morphological leveling. $\Lambda(\mathbf{I}, \mathbf{M})$ is the closest image to $\mathbf{I} \in \text{Inter}(\mathbf{I}, \mathbf{M})$ using geodesic propagations from \mathbf{M} . We call $\text{Inter}(g, f)$ the class of functions h verifying $g \wedge f \leq h \leq g \vee f$ [9]. Given two grey scale images \mathbf{I}_1 and \mathbf{I}_2 , the global separation

along the leveling path is defined as follows:

$$L(\mathbf{I}_1, \mathbf{I}_2) = \sum_i \|\lambda^{i+1}(\mathbf{I}_1, \mathbf{I}_2) - \lambda^i(\mathbf{I}_1, \mathbf{I}_2)\|^2 \quad (8)$$

where the original image is denoted as $\lambda^0 = \mathbf{I}_2$ and the target image as $\lambda^{\infty+1} = \mathbf{I}_1$. From 8, we define the *total leveling distance* by:

$$D_L(\mathbf{I}_1, \mathbf{I}_2) = \max \{L(\mathbf{I}_1, \mathbf{I}_2), L(\mathbf{I}_2, \mathbf{I}_1)\} \quad (9)$$

Additionally, principal leveling separation is defined by:

$$L_2(\mathbf{I}_1, \mathbf{I}_2) = \|\mathbf{I}_2 - \Lambda(\mathbf{I}_1, \mathbf{I}_2)\|^2 + \|\Lambda(\mathbf{I}_1, \mathbf{I}_2) - \mathbf{I}_1\|^2 \quad (10)$$

and the *principal leveling distance* is defined by:

$$D_{L_2}(\mathbf{I}_1, \mathbf{I}_2) = \max \{L_2(\mathbf{I}_1, \mathbf{I}_2), L_2(\mathbf{I}_2, \mathbf{I}_1)\} \quad (11)$$

Figure 1 illustrates that the sequence of images produced by leveling operator is in a subspace where the Euclidean path may not be included.

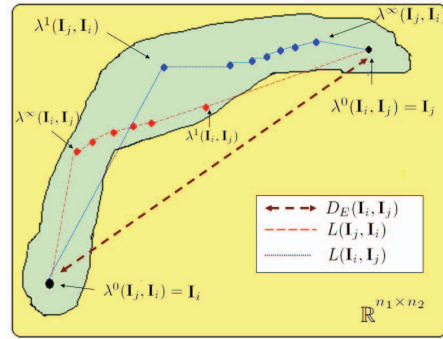


Fig. 1. Leveling-based morphological distance scheme

6. EXPERIMENTS AND RESULTS

In the first scenario, performance of morphological distances is studied in comparison with other standard distances for Indian Pines hyperspectral image, obtained by the AVIRIS sensor. Figures 2(a) and 2(c) presents the dimension of the reduced feature space in the x-axis and overall classification rate in the y-axis for different distances used in KPCA and Isomap. The classification was obtained with nonlinear SVMs [10] in 16 classes and 20 training samples per class in 25 repetitions. The different distances do not present significant differences in very low-dimensional space but Hausdorff distance is lightly better when the number of feature increase. To illustrate the advantages in our approach, the second scenario includes noise in a random subset of 20 contiguous bands. Figures 2(b) and 2(d) shows the average classification rate versus increasing values of the noise standard deviation for 20 repetitions. In this experiment, total

variance of the reduction had been set at 99% and the largest axes are retained in the KPCA and Isomap projection. The ranking of the distance (based on robustness) from best to worst are principal leveling, KLD, Bhattacharyya, Hausdorff, total leveling and Euclidean using KPCA. Differences in performance using Isomap reduction are notorious in figure 2(d). Scenario 2 highlights the disadvantages of Euclidean distance in images affected by noise. Another experiments incorporating horizontal and vertical shifts in random subset of bands present similar results but they are not discussed because of space limitations.

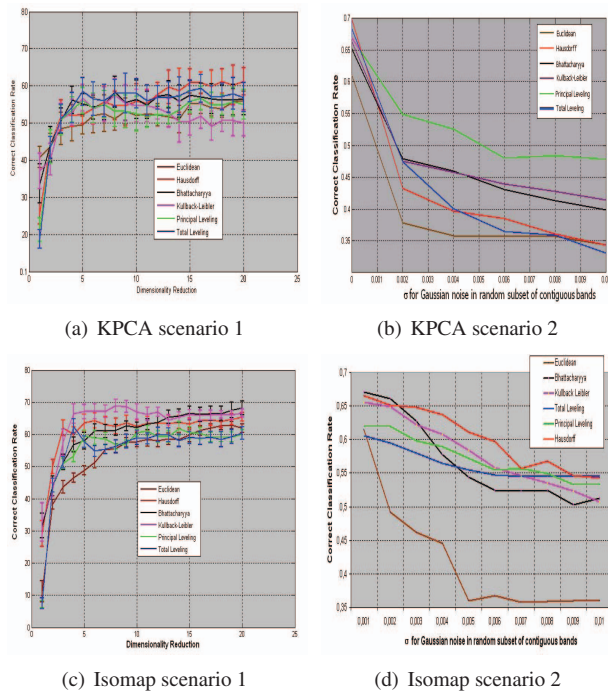


Fig. 2. Classification by SVMs in 16 classes and 20 training samples per class in 25 repetitions. Scenario 1: Different dimensions to reduce. Scenario 2: Including Gaussian Noise in 20 random contiguous bands. See the text for more details.

7. REFERENCES

- [1] C.M. Bachmann, T.L. Ainsworth, R.A. Fusina. Exploiting manifold geometry in hyperspectral imagery. *IEEE Trans. Geosci. And Remote Sens.*, 43(3): 441–454, 2005.
- [2] J.P. Benzécri. *L'Analyse Des Données. L'Analyse des Correspondances II.* Paris, Dunod, 1973.
- [3] Bhattacharyya, A. On a Measure of Divergence Between Two Statistical Populations Defined by their Probability Distributions. *Bull. Calcutta Mathematical Society*, 35: 99-110, 1943.
- [4] Y. Chen, M.M. Crawford, J. Ghosh. Applying Nonlinear Manifold Learning to Hyperspectral Data for Land Cover Classification. In *IEEE Proc.of IGARSS'05*, 6: 4311–4314, 2005.
- [5] Shilpa Inamdar, Francesca Bovolo, Lorenzo Bruzzone, and Subhasis Chaudhuri, Multidimensional Probability Density Function Matching for Preprocessing of Multitemporal Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.*, 46(4): 1243-1252, 2008.
- [6] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using the Hausdorff distance. *IEEE Trans. on PAMI*, 15(9): 850-863, 1993.
- [7] Mark P. Kolba and Leslie M. Collins, Analysis of an information-based sensor manager applicable to landmine detection. *IEEE Geoscience and Remote Sensing Symposium*, 2:521-524, 2008.
- [8] S. Kullback. *Information theory and Statistics*, New York: John Wiley and Sons., 1959.
- [9] G.Matheron. Les nivellements. Technical Report, Centre de Morphologie Mathématique, 1997.
- [10] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.*, 42(8):1778-1796, 2004.
- [11] F. Meyer. Levelings, Image Simplification Filters for Segmentation. *Journal of Mathematical Imaging and Vision*, 20: 59–72, 2004.
- [12] A. Mohan, G. Sapiro, E. Bosch. Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images. *IEEE. Geosci. And Remote Sens Letters*, 4(2): 206–210, 2007.
- [13] B. Schölkopf, A. Smola, K.-R. Müller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Max-Planck-Institut für biologische Kybernetik*, Technical Report No. 44, 1996.
- [14] J. Serra. *Image Analysis and Mathematical Morphology. Vol I, and Image Analysis and Mathematical Morphology. Vol II: Theoretical Advances.* London: Academic Press, 1982,1988.
- [15] J. Serra. Hausdorff distance and interpolations. In (*H. Heijmans and J. Roerdink Eds.*) *Mathematical Morphology and its Applications to Image and Signal Processing*, Kluwer, 1998.
- [16] J.B. Tenenbaum, V. de Silva, J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500): 2319–2323, 2000.