Nonlinear dimensionality reduction of hyperspectral images based on spectral angles and exploiting the spatial context

E V Myasnikov

1Samara National Research University, Moskovskoe Shosse 34, Samara, Russia, 443086

e-mail: mevg@geosamara.ru

Abstract. I proposed a nonlinear method for the dimensionality reduction of hyperspectral images in this paper. A special feature of the proposed method is the use of spectral angles in the initial hyperspectral space as a dissimilarity measure between pixels of an image, as well as taking into account the spatial context of the hyperspectral image pixels. I used a well-known hyperspectral image dataset in the experiments. The experiments showed the advantage of the developed method over the basic nonlinear dimensionality reduction methods and the linear principal component analysis technique.

1. Introduction

Each pixel of a hyperspectral image is a vector composed of hundreds of spectral components corresponding to a wide range of wavelengths. Due to a fine spectral resolution, hyperspectral images are widely used in many different areas. However, high spectral dimensionality of such images cause difficulties in storing, transmitting, processing and recognition of such images. Due to this reason, the reduction of redundancy becomes an important stage in the processing of hyperspectral images.

The attention of researchers was given both to supervised feature selection techniques and to unsupervised dimensionality reduction techniques. Nevertheless, due to the objective reasons such as valuable computational costs, less stability to changes in an image scene, the necessity to the presence of ground truth information, the feature selection techniques became less popular than the dimensionality reduction techniques [1].

Both linear and nonlinear dimensionality reduction techniques are used with hyperspectral images. The linear techniques are used more often, and the most popular one is the principal component analysis technique (PCA) [2]. This technique searches the projection of data in the lower dimensional linear subspace, which minimizes the variance of data. Other examples of linear techniques are independent component analysis (ICA) [3], projection pursuit and some others.

The examples of the nonlinear dimensionality reduction techniques used in hyperspectral image analysis are Locally-linear embedding (LLE) [4], Laplacian Eigenmaps (LE) [5], Local Tangent Space Alignment (LTSA) [6], isometric embedding (ISOMAP) [7], Curvilinear component analysis (CCA), Curvilinear distance analysis (CDA) [8], and Nonlinear Mapping (NLM) [9]. These techniques are
used in the hyperspectral image processing less often, but it is known [10] that hyperspectral remote sensing images are affected by nonlinear mixing effects due to multipath light scattering and other reasons.

However, in the last years, nonlinear dimensionality reduction techniques become more popular in the field of hyperspectral image analysis. These methods allow to increase the effectiveness of pixel-wise classification, to generate pseudo color image representations with unique properties, and solve other problems. Today we can indicate a number of papers, which describe successful application of the nonlinear dimensionality reduction techniques in multi- or hyperspectral image processing.

Many nonlinear dimensionality reduction techniques require to measure dissimilarity between image pixels in a hyperspectral space. In most cases, the Euclidean distance is used as a dissimilarity measure, but other measures exist, which were successfully applied in hyperspectral image analysis. These measures include the spectral angle measure (SAM), the spectral correlation, the spectral information divergence (SID), and some other measures. The most popular among these measures is the SAM measure [11].

Unfortunately, the SAM measure was rarely used in the nonlinear dimensionality reduction techniques of hyperspectral images. For example, in the paper [12] the SAM measure (together with the Euclidean distance) was used at the first step of the ISOMAP technique to find neighbor pixels in a hyperspectral space. In paper [13] the effectiveness of the same measures was studied for the Laplacian eigenmaps technique. In paper [14] the SAM measure and Laplacian eigenmaps technique were used in the time series analysis of multispectral images. In paper [15] the SAM measure was evaluated with the nonlinear mapping technique in the task of per-pixel classification of hyperspectral images. In particular, several nonlinear mapping techniques were proposed based on the principle of spectral angle preserving.

Other non-Euclidean measures seem to be even less studied in the nonlinear dimensionality reduction of hyperspectral images. In paper [16] the nonlinear mapping technique based on the principle of spectral correlation preserving was proposed. In paper [17], the SID measure was used with the ISOMAP technique for the analysis of hyperspectral images.

It is worth noting that the listed above nonlinear dimensionality reduction techniques are general – purpose techniques. When applied to hyperspectral images, these techniques act in a spectral space and do not take into account the spatial information contained in hyperspectral images. While there are a lot of papers devoted to the use of spatial information in hyperspectral image analysis (see [18] for a recent review on this topic), there are very few papers devoted to the use of spatial information in the dimensionality reduction of hyperspectral images (for example, see [19-21]).

Thus, the field of using non-Euclidean measures and exploiting the spatial information in the dimensionality reduction of hyperspectral image data is insufficiently investigated. In this paper we study one of the possible approaches to exploit both the spectral angle information and the spatial context in the nonlinear mapping technique, which is one of the oldest and well-known dimensionality reduction techniques.

This paper is organized as follows. The next section describes the proposed nonlinear dimensionality reduction technique, based on the spectral angles and exploiting the spatial context of image pixels. The description starts with the introduction of a pixel dissimilarity measure. Then the measure of the dimensionality reduction quality is introduced, and the numerical optimization procedure based on the stochastic gradient descent is derived. An experimental study is described in Section 3. We describe the dataset used in the experiments, study the proposed method for various values of the parameters, and compare it to the principal component analysis and base nonlinear mapping techniques in the term of the classification accuracy. The conclusion is given in Section 4.

2. Methods

The dimensionality reduction method proposed in this paper is based on the idea of preserving the pairwise dissimilarities between pixels of a hyperspectral image. In contrast to earlier methods, which
were based on preserving of the Euclidean distances [22], spectral angles [15] or exploited spatial context [21], the proposed method combines both spectral angles and spatial context.

The motivation behind the combination of the above techniques is the better classification performance of the above techniques compared to the linear PCA technique and earlier nonlinear mapping methods [15, 21]. To combine the spectral angles and spatial context, we use new dissimilarity measure for the pixels of a hyperspectral image. The proposed measure takes into account not only the spectral dissimilarity of two pixels of an image, but also the dissimilarity of the neighborhoods of the pixels.

2.1. Dissimilarity measure

To describe this measure, let’s first introduce the hyperspectral image \( I \), which contains \( N=W \times H \) image pixels \( x_i, i=1..N \) (here we assume that the image pixels are numbered, for example, according to the progressive scanning). Each pixel \( x_i \) is the vector \( x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,M}) \), where \( M \) is the number of spectral components (bands) in the hyperspectral image. Typically, \( M \) is of the order of a few hundred.

The spectral angle mapper measure (SAM) [11] is defined by the following equation:

\[
\theta(x_i, x_j) = \arccos \left( \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|} \right). \tag{1}
\]

It measures the spectral dissimilarity between two image pixels \( x_i \) and \( x_j \), but it does not take into account the spatial context of the pixels.

To exploit the spatial context, let’s introduce the neighborhood of an image pixel. The neighborhood of the pixel \( x_i \) is the set of pixels \( H_i = \{ x_j \mid r(x_i, x_j) \leq R \} \), where \( r() \) is the Euclidean distance between pixels \( x_i \) and \( x_j \) in the spatial domain of the image. An example of the neighborhood for the neighborhood radius \( R=2 \) is shown in figure 1.

To take into account the spatial context, we exploit the idea of using the order statistics. This idea was implemented earlier in [21] for the Euclidean distances. In our case, for each pixel \( x_i \) of the image, we obtain the ordered set \( x_i^* = (x_{i(1)}, x_{i(2)}, \ldots, x_{i(K)}) \), \( K = |H_i| \) by sorting neighbor pixels \( x_j \in H_i \) according to the SAM measure between pixels \( x_i \) and \( x_j \):

\[
\theta(x_i, x_{i(1)}) \leq \theta(x_i, x_{i(2)}) \leq \ldots \leq \theta(x_i, x_{i(K)}). \tag{2}
\]

Figure 1. The neighborhood of a pixel.

The proposed dissimilarity measure is expressed as a weighted sum of spectral angles calculated over the first \( S \) corresponding pixels from ordered neighborhoods:

\[
\rho(x_i, x_j) = \eta \sum_{s=1}^{S} w_s \theta(x_{i,s}^*, x_{j,s}^*) \tag{3}
\]

Here \( \eta = \left( \sum_{s=1}^{S} w_s \right)^{-1} \) is the normalizing coefficient, \( w_s = 1/s \) are inverse weighting coefficients, \( S \) is the number of pixels in the ordered neighborhood (order statistics) used as a spatial context.
2.2. Dimensionality reduction

The dimensionality reduction method maps hyperspectral image pixels (vectors) \( x_i \) from the multidimensional hyperspectral space \( R^M \) into the lower-dimensional embedding space \( R^m \). Let \( y_i \) be the coordinates of the corresponding vectors in the embedding space. Then the hyperspectral data mapping error can be expressed as the following [21]:

\[
\varepsilon = \mu \sum_{i=1}^{N} \sum_{j \neq i+1}^{N} \rho(x_i, x_j) - d(y_i, y_j) \]

where \( d(\cdot) \) is the Euclidean distance between vectors in the embedding space

\[
d(y_i, y_j) = \| y_i - y_j \| = \sqrt{\sum_{k=1}^{m} (y_{ik} - y_{jk})^2}
\]

\( \mu \) is a normalizing coefficient

\[
\mu = \left( \sum_{i=1}^{N} \sum_{j \neq i+1}^{N} \rho^2(x_i, x_j) \right)^{-1}
\]

To minimize the above data mapping error, we use the stochastic gradient descent approach. This approach was used in previous works [15, 22, 21] and provided acceptable results.

To perform the minimization, we introduce the extended vector of parameters \( Y = (y_1, y_2, ..., y_N) \), which consists of low-dimensional coordinates of the vectors in the embedding space. The numerical optimization procedure is based on the following recurrence equation:

\[
Y(t+1) = Y(t) - \alpha \nabla \varepsilon(t)
\]

where \( t \) is the number of an iteration, \( \nabla \varepsilon \) is the estimation of a gradient. This estimation in the stochastic gradient descent technique can be done using a random subsample (mini-batch):

\[
\nabla \varepsilon(t) = \sum_{j=1}^{B} \nabla \varrho_{y_j}(t)
\]

where \( b_j \) is the \( j \)-th sample of this subsample, and \( B \) is the cardinality of the subsample \( b \).

After finding the partial derivatives, the optimization procedure can be represented in the form:

\[
y_i(t+1) = y_i(t) + 2\alpha \mu \sum_{j=1}^{B} \frac{\rho(x_i, x_{b_j}) - d(y_i, y_{b_j})}{d(y_i, y_{b_j})}(y_i(t) - y_{b_j}(t))
\]

The computational complexity of the optimization procedure is \( O(NMmSB) \) per one iteration, where \( N \) is the number of pixels, \( M \) is the dimensionality of the hyperspectral space (the number of bands), \( m \) is the dimensionality of the embedding space, \( S \) is the number of order statistics, and \( B \) is the number of samples used to estimate the gradient value (the cardinality of mini-batch). The application of the stochastic gradient descent technique allows us to significantly speed-up the optimization procedure, as one iteration for the base gradient descent technique is \( N/B \) times slower, and usually \( N > > R \).

Thus, the proposed method consists of the following steps:

- the initialization of the coordinates \( y_i, i=1..N \) in the embedding space by projecting the source vectors \( x_i, i=1..N \) onto the first \( m \) principal components;
- the iterative optimization in according to (9) until the coordinates \( y_i \) become stable.

This method allows to find a suboptimal solution to the \( \varepsilon \rightarrow \varepsilon_{\text{min}} \) problem.

It is worth noting that the described method attempts to approximate the spectral dissimilarities \( \rho(x_i, x_j) \) calculated in the hyperspectral space by the Euclidean distances \( d(y_i, y_j) \) in the embedding space.
3. Experiments
In the experimental study, we used open access test hyperspectral image scenes [23]. In this paper, we provide the results of the experiments for the Indian Pines hyperspectral scene. This test image (figure 1 (a)) was acquired using the AVIRIS sensor. It contains 145 x 145 pixels. For the experiments, we used the version with 200 spectral bands, as some bands were discarded due to a high noise and water absorption.

![Figure 2. Indian Pines test hyperspectral image: (a) pseudocolor representation generated using the nonlinear mapping technique; (b) groundtruth classification (classified pixels are shown in color).](image)

During the preliminary experiments, it was found that the error (4) is significantly reduced during the optimization process. For the presented data set, the error (4) decreased by several orders of magnitude after the first few dozens of iterations.

To evaluate the proposed technique we used two well-known and often used classifiers, namely the nearest neighbor classifier (NN), and the support vector machine (SVM). To measure the classification quality, we used the classification accuracy (CA), which is defined as the proportion of properly classified pixels. To perform the experiments the whole set of groundtruth pixels (the groundtruth image is provided with the hyperspectral scene, see figure 1(b)) were divided into the learning (60 percent) and test (40 percent) sets.

In the first experiment, we studied the dependence of the classification accuracy on the number of order statistics. We performed experiments for two values of the neighborhood radius, which is used to define the spatial context of an image pixel: $R=1$, and $R=2$. Some results of the experiments are shown in figures 3, 4.

As it can be seen from the figures, the more order statistics $S$ were used (for the fixed neighborhood radius $R$), the better was the classification quality (CA). Reciprocally, the results for the neighborhood radius $R=2$ were better than for $R=1$. We did not consider greater values of $R$ due to an unacceptably long operating time.

![Figure 3. The dependency of the NN classification accuracy (CA) on the dimensionality of the embedding space ($m$) for the neighborhood radius $R=1$ (left), and $R=2$ (right).](image)
In the next experiment we evaluated three different dimensionality reduction techniques, namely the principal component analysis technique (PCA), the nonlinear mapping method based on the approximation of the SAM measures by the Euclidean distances (SAED) [15], the nonlinear mapping technique using the Euclidean distances and spatial context (ED+SC) [21], and the proposed dimensionality reduction technique using the spectral angles and spatial context (SAM+SC).

For two nonlinear techniques using spatial context (ED+SC and SAM+SC), we used identical spatial context parameters: neighborhood radius $R=1$, $R=2$, and the number of order statistics $S=5$, $S=10$.

Some results of the experiments are shown in figures 5, 6.

As it can be seen from the figures, both nonlinear techniques (ED+SC and SAM+SC) exploiting the spatial context significantly outperform the linear PCA technique in almost all the considered cases. The proposed nonlinear method based on spectral angles and spatial context (SAM+SC) provides slightly better results than the method based on Euclidean distances. It is worth noting that the difference is more evident for the NN classifier. This observation can be explained by the principle that is adopted in the dimensionality reduction method. According to this principle, the pairwise dissimilarities are approximated by the Euclidean distances in the embedding space.

As it was mentioned in Section 2, the computational complexity linearly depends on the number of order statistics. This is confirmed by the figure 7. Unfortunately, the whole optimization process can take considerable time depending on the size of the image, the mini-batch cardinality, the initial and output dimensionality, the number of order statistics and the total number of iterations.

**Figure 5.** The dependence of the NN classification accuracy for the principal components analysis technique (PCA), nonlinear mapping method based on the approximation of the SAM measures by the Euclidean distances (SAED), nonlinear mapping technique using Euclidean distances and spatial context (ED+SC) and the technique using spectral angles and spatial context (SAM+SC): neighborhood radius $R=1$ (left), and $R=2$ (right).
The dependence of the SVM classification accuracy for the principal components analysis technique (PCA), nonlinear mapping method based on the approximation of the SAM measures by the Euclidean distances (SAED), nonlinear mapping technique using Euclidean distances and spatial context (ED+SC) and the technique using spectral angles and spatial context (SAM+SC): neighborhood radius $R=1$ (left), and $R=2$ (right).

Dependence of the average time $\tau$ (sec.) of one iteration on the number of order statistics $S$ used as a spatial context.

4. Conclusion
A new technique for the nonlinear dimensionality reduction of hyperspectral image data is proposed in this paper. This technique is based on the approximation of the spectral angles by the Euclidean distances and exploiting the spatial context of hyperspectral image pixels. The experimental study showed that the proposed technique allows to significantly improve the accuracy of per-pixel classification.

As the optimization process in the proposed technique can take considerable time depending on the input hyperspectral image and parameters, the parallel implementation of the proposed technique is of particular interest. Another promising line of research is the use of the proposed technique in the unsupervised clustering and segmentation scenarios [24].

5. References
[1] Lunga D, Prasad S, Crawford M and Ersoy O 2014 Manifold-Learning-Based Feature Extraction for Classification of Hyperspectral Data IEEE Signal Processing Magazine 31(1) 55-66
[2] Richards J A, Jia X, Ricken D E and Gessner W 1999 Remote Sensing Digital Image Analysis: An Introduction (New York: Springer-Verlag)
[3] Wang J and Chang C-I 2006 Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis IEEE Trans. Geosci. Remote Sens 44(6) 1586-1600
[4] Roweis S T and Saul L K 2000 Nonlinear Dimensionality Reduction by Locally Linear Embedding Science 290 2323-2326
[5] Belkin M and Niyogi P 2001 Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering Advances in Neural Information Processing Systems 14 586-691
[6] Zhang Zh and Hongyuan Zha 2005 Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment SIAM Journal on Scientific Computing 26(1) 313-338
[7] Tenenbaum J B, de Silva V and Langford J C 2000 A Global Geometric Framework for Nonlinear Dimensionality Reduction Science 290 2319-2323
[8] Demartines P and Hérault J 1997 Curvilinear Component Analysis: A Self-Organizing Neural Network for Nonlinear Mapping of Data Sets IEEE Transactions on Neural Networks 8(1) 148-154
[9] Sammon J W Jr. 1969 A nonlinear mapping for data structure analysis IEEE Trans. Comput. C 18(5) 401-409
[10] Bachmann C M, Ainsworth T L and Fusina R A 2006 Improved Manifold Coordinate Representations of Large-Scale Hyperspectral Scenes IEEE Trans. Geosci. Remote Sens 44(10) 2786-2803
[11] Kruse F A, Boardman J W, Lefkoff A B, Heidebrecht K B, Shapiro A T, Barloon P J and Goetz A F H 1993 The Spectral Image Processing System (SIPS) interactive visualization and analysis of imaging spectrometer data Remote Sens. Environ. 44 145-163
[12] Bachmann C M, Ainsworth T L and Fusina R A 2005 Exploiting manifold geometry in hyperspectral imagery IEEE Trans. Geosci. Remote Sens. 43(3) 441-454
[13] Yan L and Niu X 2014 Spectral-angle-based Laplacian Eigenmaps for nonlinear dimensionality reduction of hyperspectral imagery Photogramm. Eng. Remote Sens. 80(9) 849-861
[14] Yan L and Roy D P 2015 Improved time series land cover classification by missing observation-adaptive nonlinear dimensionality reduction Remote Sens. Environ. 158 478-491
[15] Myasnikov E 2017 Nonlinear Mapping Based on Spectral Angle Preserving Principle for Hyperspectral Image Analysis Lecture Notes in Computer Science 10425 416-427
[16] Myasnikov E 2018 Nonlinear dimensionality reduction of hyperspectral data using spectral correlation as a similarity measure Lecture Notes in Computer Science 10716 237-244
[17] Du P, Wang X, Tan K and Xia J 2011 Dimensionality reduction and feature extraction from hyperspectral remote sensing imagery based on manifold learning Geomatics and Information Science of Wuhan University 36(2) 148-152
[18] Wang L, Shi C, Diao C, Ji W and Yin D 2016 A survey of methods incorporating spatial information in image classification and spectral unmixing International Journal of Remote Sensing 37(16) 3870-3910
[19] Borhani M, Ghassemian H 2015 Kernel Multivariate Spectral-Spatial Analysis of Hyperspectral Data IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(6) 2418-2426
[20] Sun W, Liu C and Li W 2015 Hyperspectral imagery classification using the combination of improved laplacian eigenmaps and improved k-nearest neighbor classifier Geomatics and Information Science of Wuhan University 40(9) 1151-1156
[21] Myasnikov E V 2017 Exploiting spatial context in nonlinear mapping of hyperspectral image data Lecture Notes in Computer Science 10485 180-190
[22] Myasnikov E V 2016 Evaluation of stochastic gradient descent methods for nonlinear mapping of hyperspectral data Lecture Notes in Computer Science 9730 276-283
[23] Hyperspectral Remote Sensing Scenes (Access mode: http://www.ehu.eus/ccwintco/index.php?title=HyperspectralRemoteSensingScenes)
[24] Myasnikov E V 2017 Hyperspectral image segmentation using dimensionality reduction and classical segmentation approaches Computer Optics 41(4) 564-572 DOI: 10.18287/2412-6179-2017-41-4-564-572

Acknowledgments
The reported study was funded by RFBR according to the research project no. 18-07-01312-a.