

PAPER • OPEN ACCESS

Nonlinear dimensionality reduction of hyperspectral data based on spectral information divergence preserving principle

To cite this article: E V Myasnikov 2019 *J. Phys.: Conf. Ser.* **1368** 032030

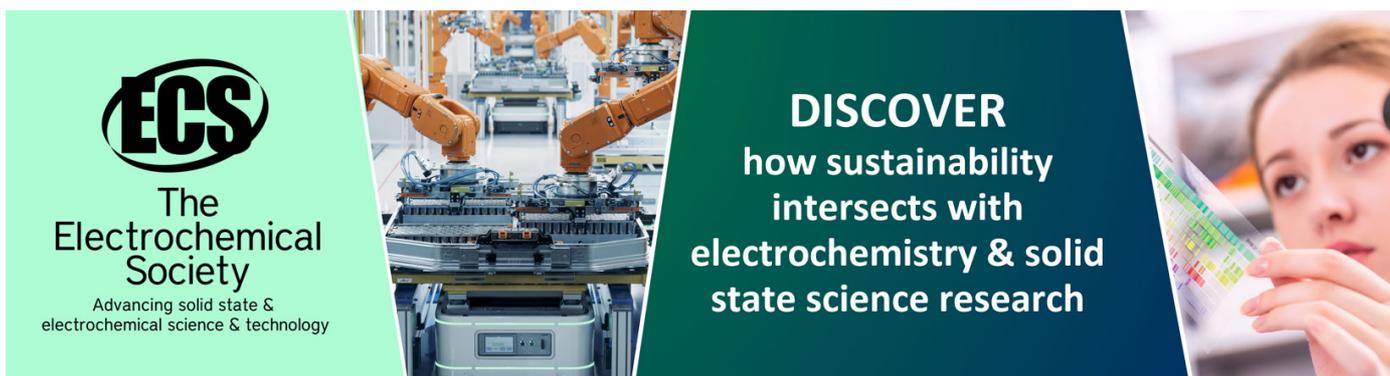
View the [article online](#) for updates and enhancements.

You may also like

- [Adaptive and automatic red blood cell counting method based on microscopic hyperspectral imaging technology](#)
Xi Liu, Mei Zhou, Song Qiu et al.

- [Intraoperative hyperspectral label-free imaging: from system design to first-in-patient translation](#)
Michael Ebner, Eli Nabavi, Jonathan Shapey et al.

- [Hyperspectral Image Classification Using Deep Learning Models: A Review](#)
Deepak Kumar and Dharmender Kumar



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Nonlinear dimensionality reduction of hyperspectral data based on spectral information divergence preserving principle

E V Myasnikov^{1,2}

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

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: mevg@geosamara.ru

Abstract. The paper proposes a nonlinear dimensionality reduction technique for hyperspectral data based on the principle of preserving the spectral information divergence (SID). In this technique, the spectral information divergence is used as a spectral dissimilarity measure in both the original and reduced space. In the experimental studies performed using open hyperspectral images, we solve the problem of per-pixel classification of hyperspectral images. The developed method extends the existing hyperspectral data analysis toolkit, which allows reducing the dimensionality of hyperspectral data based on the principle of preserving specified spectral dissimilarity measures.

1. Introduction

A hyperspectral image is a three-dimensional array with two spatial and one spectral dimension. Hyperspectral images are recorded using high spectral resolution sensors. Hence, each sample of such images is a vector that includes up to several hundred spectral components.

Hyperspectral images are widely used in various fields. Their use, on the one hand, provides new opportunities, allowing you to extract information about the materials (components) presented in an image, and on the other hand, imposes difficulties caused by the high spectral dimension. As a result, the most important task is to eliminate the redundancy of such images while maintaining important spectral information.

Linear dimensionality reduction methods are most often used to eliminate this redundancy. The most popular linear method is the principal component analysis technique (PCA) [1], which searches for a linear projection into a subspace of lower dimension that maximizes the variance of data. In recent years, in light of the considered problem, the popularity of nonlinear dimensionality reduction methods grows. These nonlinear methods are Locally-linear embedding (LLE) [2], Laplacian Eigenmaps (LE) [3], Local Tangent Space Alignment (LTSA) [4], isometric embedding (ISOMAP) [5], Curvilinear Component analysis (CCA), Curvilinear Distance Analysis - CDA) [6] and Nonlinear Mapping (NLM) [7].

The possibility of taking into account desirable spectral dissimilarity measures is of particular interest when working with nonlinear dimensionality reduction techniques. Although many of the above nonlinear methods allow us to take into account the chosen spectral dissimilarity measure in the hyperspectral input space (see, for example, [8, 9, 10]), only a few works [11, 12] are known, in which



an attempt is made to take into account the chosen dissimilarity measure in the output reduced space. In both of these works, the methods being developed are based on the principle of preserving selected measures of spectral mismatch. In the paper [11], the authors proposed several nonlinear mapping methods based on the principle of preserving the spectral angle mapper (SAM) measures [13]. In the paper [12], the method based on the principle of preserving the spectral correlation measure [13] was introduced.

The method proposed in this paper is based on the same ideas as the above papers [11, 12]. However, unlike previous works, this method is based on the principle of preserving the divergence of spectral information (SID) [13]. This measure, along with the spectral angle and spectral correlation measures, is among the three most well-known non-Euclidean measures of spectral dissimilarity in the hyperspectral data analysis. This measure shows good results in solving hyperspectral data classification problems.

The work is organized as follows. Section 2 describes the basics of the developed dimensionality reduction method. In particular, we introduce the quality measure for the dimensionality reduction and derive a numerical optimization procedure based on the stochastic gradient descent method. Section 3 describes the experimental studies. In this section, using publicly available hyperspectral images, the proposed method is compared with alternative nonlinear dimensionality reduction methods based on other measures of spectral dissimilarity. The effectiveness of the proposed approach is studied in terms of classification accuracy. The work ends with a conclusion in section 4 and a list of used literature in section 5.

2. Method

Consider a hyperspectral image X , of size $W \times H$, containing M spectral components. Each sample x_i , $i = 1..N$, $N = WH$ of such an image can be considered as a vector in the multidimensional hyperspectral space R^M of dimension M .

As mentioned in the introduction, spectral information divergence (SID) is a fairly popular measure in hyperspectral image analysis. For a pair of vectors x_i and x_j , this measure is given as follows [13]:

$$SID(x_i, x_j) = D(x_i \| x_j) + D(x_j \| x_i),$$

$$D(x_i \| x_j) = \sum_{k=1}^M p_k(x_i) \log(p_k(x_i) / p_k(x_j)),$$

$$p_k(x_i) = \frac{x_{ik}}{\sum_{l=1}^M x_{il}}.$$
(1)

Here, $p(x) = (p_1(x), \dots, p_M(x))^T$ is considered as the probability distribution function of a discrete random variable generated by the vector x , and $D(x_i \| x_j)$ is the difference of the self-information for the vector x_i with respect to the information for the vector x_j averaged over all spectral ranges (relative entropy). Note that the spectral information divergence (SID) is a symmetric measure by construction.

The nonlinear method proposed in this paper for the dimensionality reduction of hyperspectral data is based on the principle of preserving the spectral information divergence between samples of a hyperspectral image.

First, to each sample x_i , $i = 1..N$ of the original image, we associate a vector y_i in the reduced space R^L of smaller dimension $L < M$. Next, we introduce the following measure of the preserving of the spectral information divergence:

$$\varepsilon = \mu \sum_{i,j=1(i < j)}^N (SID(x_i, x_j) - SID(y_i, y_j))^2$$
(2)

this measure shows how well the pairwise spectral dissimilarity measures between the samples of the hyperspectral image are preserved in the space of lower dimension.

The factor μ is assumed to be constant in the above expression (2) for a given image X . It is calculated in the present work as

$$\mu = \left(\sum_{i,j=1(i<j)}^N SID^2(x_i, x_j) \right)^{-1}$$

Formally, the task of dimensionality reduction can be posed as the problem of minimizing the error functional (2) for all possible values of the parameters, which are the coordinates of vectors (image samples) in the reduced space $Y = \{y_i, i = 1..N\}$.

The minimization problem can be solved by analogy with the works [11, 12, 14] using the methods of stochastic gradient descent. With this approach, the true value of the gradient $\nabla \varepsilon$ in the gradient descent procedure

$$Y(t) = Y(t-1) - \alpha \nabla \varepsilon \quad (3)$$

is approximated at each iteration of the procedure using a subset consisting of R randomly selected vectors $\{y_{r_j}, j = 1..R\}$, where r_j are random indices.

The implementation of such an optimization scheme requires finding partial derivatives of the error function (2):

$$\frac{\partial \varepsilon}{\partial y_{ik}} = -2\mu \sum_{i,j=1(i<j)}^N (SID(x_i, x_j) - SID(y_i, y_j)) \frac{\partial}{\partial y_{ik}} SID(y_i, y_j)$$

and dissimilarity measure (1):

$$\frac{\partial}{\partial y_{ik}} SID(y_i, y_j) = \frac{1}{\sum_{l=1}^L y_{il}} \left(1 - s_{ij} - \frac{p_k(y_j)}{p_k(y_i)} + \log \frac{p_k(y_i)}{p_k(y_j)} \right),$$

where

$$s_{ij} = \sum_{l=1}^L \left(p_l(y_i) + p_l(y_i) \log \frac{p_l(y_i)}{p_l(y_j)} - p_l(y_j) \right)$$

The final recurrence relation, which allows refining the coordinates of the vectors $y_i, i=1..N$ in the reduced spectral space, is as follows:

$$y_{ik}(t) = y_{ik}(t-1) + 2\alpha\mu \sum_{j=1}^R \frac{SID(x_i, x_{r_j}) - SID(y_i, y_{r_j})}{\sum_{l=1}^L y_{il}} \left(1 - s_{ir_j} - \frac{p_k(y_{r_j})}{p_k(y_i)} + \log \frac{p_k(y_i)}{p_k(y_{r_j})} \right) \quad (4)$$

3. Experiments

This section describes the results of the experimental comparison of the developed method with alternative techniques. For the experiments, we used two well-known hyperspectral scenes [14], which have pixel-by-pixel true classification data: Indian pines and Kennedy space center. Both hyperspectral images were obtained using the AVIRIS sensor.

The first scene contains 145×145 samples in 224 spectral ranges. In the experiments, we used a derived image containing 204 spectral channels, in which some spectral channels were excluded due to a high noise level. The pixels of the classified image are divided into 16 classes.

The Kennedy space center image contains 512×614 pixels. In the experiments, we used a version containing 176 spectral ranges. The image with the true classification contains information only about a relatively small portion of pixels, which are divided into 13 classes.

In our experiments, we performed a pixel-by-pixel classification of the above images using a two-step approach. At the first stage of the approach, we reduced the dimensionality. At the second stage, we performed the classification itself using the descriptions (features) obtained at the first stage. The experimental studies described below reflect the results obtained using the developed dimensionality reduction method as well as using similar methods based on various spectral dissimilarity measures. In particular, we used the following nonlinear dimensionality reduction techniques: nonlinear mapping

method based on the Euclidean distance (NLM), the method based on the approximation of spectral angles by the Euclidean distances (SAED), the method based on the spectral correlation preserving (SCPM), and the developed method based on the preserving of the spectral information divergence (SDPM). At the classification stage, we used the nearest neighbor (1-NN) classifiers corresponding to the methods used for dimensionality reduction. For the NLM and SAED methods, in which the dissimilarity between the output vectors was calculated as the Euclidean distance, the 1-NN classifier based on the Euclidean distance was used. For the SCPM method, 1-NN based on the Pearson correlation coefficient was used. For the developed SDPM method, the spectral information divergence (SID) was used as a measure in the 1-NN classifier.

At first, we applied the true classification mask in our experiments. After that, we reduced the dimensionality of the selected set of samples with a known pixel-by-pixel classification. We changed the output dimensionality in the range $m = 3, 5, 10, \dots, 30$, and measured the classification quality for each value of output dimensionality. As a measure of the per-pixel classification quality, classification accuracy (CA) was used, defined as the proportion of correctly classified samples from the total number of classified samples. The results of the experiments are presented below.

As it can be seen from the results of the experiments, two methods showed the best quality of classification, namely, the method based on spectral angles (SAED) and the developed method based on the spectral divergence preserving (SDPM). The proposed method showed a comparable or better quality of classification in the majority of presented cases, starting from output dimensionality $m = 10$. The worst results were shown by the SCPM method for small dimensions of the output space.

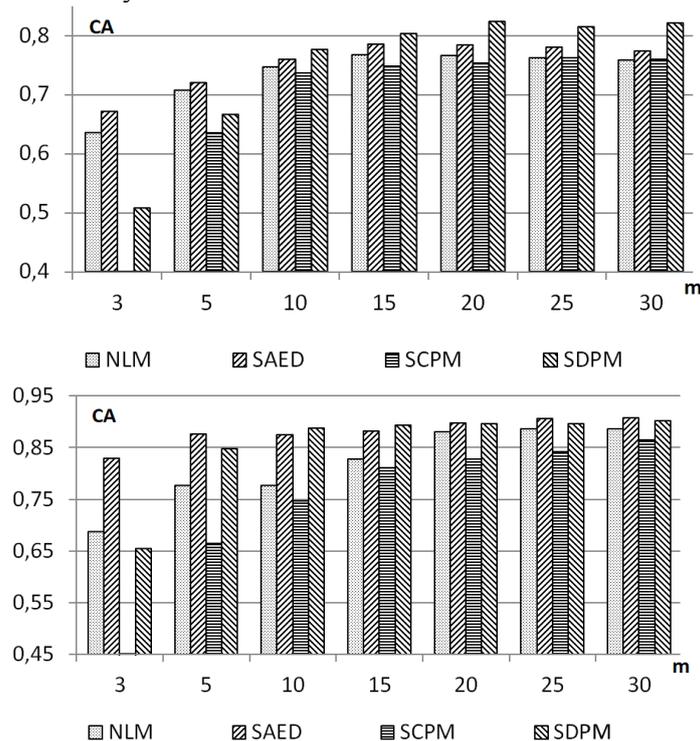


Figure 1. The dependence of the classification accuracy (CA) on the dimensionality of the reduced space (m) for various dimensionality reduction techniques: Indian Pines (top) and Kennedy Space Center (bottom) hyperspectral images.

4. Conclusion

In this paper, we proposed and investigated a nonlinear method for the dimensionality reduction of hyperspectral data (SDPM), which is based on the preserving of the spectral information divergence (SID). Experiments have shown that, compared to other studied methods, the developed method allows to obtain a comparable or better quality of pixel-by-pixel classification for the dimensionality of the output space, at least, $m = 10$.

In general, the developed method expands the existing hyperspectral data analysis toolkit, which allows reducing the dimensionality of hyperspectral data based on the principle of preserving the specified spectral dissimilarity measures. In the future, such a toolkit can be used in solving problems of analysis and monitoring of the environment [16], recognition of vegetation types in hyperspectral remote sensing data [17], etc.

5. References

- [1] Richards J A, Jia X, Ricken D E and Gessner W 1999 *Remote Sensing Digital Image Analysis: An Introduction* (Springer-Verlag New York, Inc.)
- [2] Roweis S T and Saul L K 2000 Nonlinear Dimensionality Reduction by Locally Linear Embedding *Science* **290** 2323-2326
- [3] Belkin M and Niyogi P 2001 Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering *Advances in Neural Information Processing Systems* **14** 586-691
- [4] Zhenyue Zh and Hongyuan Zh 2005 Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment *SIAM Journal on Scientific Computing* **26(1)** 313-338
- [5] Tenenbaum J B, de Silva V, Langford J C 2000 A Global Geometric Framework for Nonlinear Dimensionality Reduction *Science* **290** 2319-2323
- [6] 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
- [7] Sammon Jr J W 1969 A nonlinear mapping for data structure analysis *IEEE Trans. Comput.* **C-18(5)** 401-409
- [8] Bachmann C M, Ainsworth L and Fusina R A 2005 Exploiting manifold geometry in hyperspectral imagery *IEEE Trans. Geosci. Remote Sens.* **43(3)** 441-454
- [9] Ding L, Tang P and Li H 2013 Dimensionality reduction and classification for hyperspectral remote sensing data using ISOMAP *Hongwai yu Jiguang Gongcheng / Infrared and Laser Engineering* **6(10)** 2707-2711
- [10] Du P, Wang X, Tan K and Xia J 2011 Dimensionality reduction and feature extraction from hyperspectral remote sensing imagery based on manifold learning *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University* **36(2)** 148-152
- [11] Myasnikov E 2017 Nonlinear Mapping Based on Spectral Angle Preserving Principle for Hyperspectral Image Analysis *Lecture Notes in Computer Science* **10425** 416-427
- [12] Myasnikov E 2018 Nonlinear Dimensionality Reduction of Hyperspectral Data Using Spectral Correlation as a Similarity Measure *Lecture Notes in Computer Science* **10716** 237-244
- [13] Chang C-I 2013 *Hyperspectral Data Processing: Algorithm Design and Analysis* (John Wiley & Sons) p 1164
- [14] Myasnikov E 2016 Evaluation of Stochastic Gradient Descent Methods for Nonlinear Mapping of Hyperspectral Data *Lecture Notes in Computer Science* **9730** 276-283
- [15] Grupo de Inteligencia Computacional *Hyperspectral Remote Sensing Scenes* URL: <http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral Remote Sensing Scene>
- [16] Lebedev L I, Yasakov Yu V, Utesheva T Sh, Gromov V P, Borusjak A V and Turlapov V E 2019 Complex analysis and monitoring of the environment based on Earth sensing data *Computer Optics* **43(2)** 282-295 DOI: 10.18287/2412-6179-2019-43-2-282-295
- [17] Bibikov S A, Kazanskiy N L and Fursov V A 2018 Vegetation type recognition in hyperspectral images using a conjugacy indicator *Computer Optics* **42(5)** 846-854 DOI: 10.18287/2412-6179-2018-42-5-846-854

Acknowledgements

The work was partly funded by RFBR according to the research project 18-07-01312-a in parts of «2. Method» – «3. Experiments» and by the Russian Federation Ministry of Science and Higher Education within a state contract with the "Crystallography and Photonics" Research Center of the RAS under agreement 007-Г3/Ч3363/26 in part of «1. Introduction».