

Exploiting Spatial Context in Nonlinear Mapping of Hyperspectral Image Data

Evgeny Myasnikov^(✉)

Samara University, 34, Moskovskoye shosse, Samara 443086, Russia
mevg@geosamara.ru
<http://www.ssau.ru>

Abstract. Hyperspectral remote sensing image analysis is a challenging task due to the nature of such images. Therefore, dimensionality reduction techniques are often used as a step prior to image analysis. Although there are approaches, which exploit spatial information in image analysis, there is a lack of papers devoted to the problem of exploiting spatial information in dimensionality reduction methods. This paper is devoted to the problem of exploiting spatial context in nonlinear mapping method, which is one of the oldest and well-known dimensionality reduction techniques. To address this task, we use two possible approaches, based on window functions, and order statistics. We provide experimental results for several tasks of hyperspectral image analysis, namely classification, segmentation, and visualization. All the experiments were conducted using well-known hyperspectral images.

Keywords: Hyperspectral image · Spatial context · Nonlinear mapping · Segmentation · Classification · Clustering · Rand index

1 Introduction

A hyperspectral image is a three-dimensional array having two spatial dimensions, and one spectral dimension. Every pixel of a hyperspectral image is a vector containing hundreds of components corresponding to a wide range of wavelengths. Compared to grayscale and multispectral images, hyperspectral images offer new opportunities allowing to extract information about materials (components) located on images. Thanks to these unique properties, hyperspectral images are used in agriculture, medicine, chemistry and many other fields.

However, the high dimensionality of hyperspectral images often makes it impossible to directly apply traditional image analysis techniques to such images. For this reason, dimensionality reduction techniques are often used as a step prior to image analysis.

To reduce the dimensionality of hyperspectral images both linear and nonlinear dimensionality reduction techniques are used. Linear techniques including principal component analysis [1], independent component analysis [2], and projection pursuit are used more often. Nonlinear dimensionality reduction techniques (based on Isomap [3], locally linear embedding [4], laplacian eigenmaps [5],

nonlinear mapping [6]) are used less often, although it is known [3] that hyperspectral remote sensing images are subject to nonlinear effects due to nonlinear variations in reflectance, multipath light scattering, and the variable presence of water.

All considered dimensionality reduction techniques operate in a spectral space. But as it was outlined earlier, a hyperspectral image contains both spectral and spatial information. And the last type of information remains unused in traditional dimensionality reduction techniques. Although we can point out many papers that exploit spatial information in image analysis, and, particular, in hyperspectral image analysis (a recent review on this topic can be found in [8]), there is a lack of papers devoted to the problem of exploiting image spatial information in dimensionality reduction techniques. The purpose of this paper is to study the effectiveness of exploiting spatial information in the nonlinear mapping method, which is one of the oldest and well-known dimensionality reduction techniques.

The paper is organized as follows. In Sect. 2 we describe the dimensionality reduction approach used in this paper, and describe two schemes of exploiting spatial context. Section 3 contains the results of experiments for several tasks of image analysis, namely classification, segmentation (clustering), and visualization. The paper ends with conclusion.

2 Methods

2.1 Dimensionality Reduction

The nonlinear mapping method, that was adopted in this study as a dimensionality reduction technique, maps pixels of a hyperspectral image from a high dimensional hyperspectral space R^M into a low dimensional space R^L .

Let us denote the number of image pixels as N , the input pixel vectors in R^M as x_i , the output pixel vectors in R^L as y_i . Then the quality of a mapping can be written as a weighted sum of squared differences of distances between corresponding points in R^M and R^L :

$$\varepsilon = \mu \cdot \sum_{i,j=1, i < j}^N (\nu_{ij}(d(x_i, x_j) - d(y_i, y_j))^2). \quad (1)$$

Here $d()$ is a distance function, which is usually the Euclidean, μ and ν_{ij} are constants. In this study, we used $\mu = 1 / \sum_{i < j} d^2(x_i, x_j)$, $\nu_{i,j} = 1$. However, other values of constants can be used (see, for example, the Sammon's mapping [7]).

Usually, gradient descent based approaches are used to minimize the above error (1). But due to the high computational complexity of the basic gradient descent, we use the stochastic gradient descent approach [9]:

$$y_{ik}(t+1) = y_{ik}(t) + 2\alpha\mu \sum_{j=1}^R \nu_{i,r_j} \cdot \frac{d(x_i, x_{r_j}) - d(y_i, y_{r_j})}{d(y_i, y_{r_j})} \cdot (y_{ik}(t) - y_{r_j k}(t)). \quad (2)$$

In this equation r is the random vector of indices generated at each iteration t of the optimization process, r_j is the j -th element of this vector, R is the number of elements in r , α is the coefficient (step size) of the gradient descent.

On the whole, the nonlinear mapping method adopted in this paper consists in the initialization of output coordinates $y_i(0)$ following by the iterative optimization using (2) until coordinates $y_i(t)$ become stable. This simple algorithm allows us to find a suboptimal configuration of points y_i in the output space R^L . The computational complexity is $O(MLRN)$ per one iteration of the optimization algorithm that makes it possible to process hyperspectral images.

2.2 Exploiting Spatial Context in Dimensionality Reduction

The described above nonlinear mapping method (as many other dimensionality reduction techniques) do not provide a direct ability to account for a spatial context. In this paper, to overcome this limitation we embed spatial context information in dissimilarity measures. In particular, we follow the general idea of extending the feature space by the contextual information, extracted from spatial neighborhood of image pixels. The most obvious way is to extend the feature vector of a pixel by concatenating the values of neighbor pixels. The following equation shows such extension for a pixel with spatial coordinates (u, v) in the case of 4 pixel neighborhood (in contrast to the rest of the paper, here we use spatial coordinates of the pixel):

$$x_{u,v}^* = (x_{u,v}, x_{u-1,v}, x_{u,v+1}, x_{u+1,v}, x_{u,v-1}). \quad (3)$$

Window Functions. Contrary to feature selection techniques, which allow one to automatically detect informative and noninformative components of extended feature vectors, for unsupervised dimensionality reduction techniques it is impossible to automatically adjust the impact of particular neighbors. To overcome this problem, we can use some weighting coefficients to control the impact of particular neighbors in a dissimilarity measure. For the Euclidean distance

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^M (x_{im} - x_{jm})^2}, \quad (4)$$

the modification can be written in the form:

$$\rho(x_i^*, x_j^*) = \left(\sum_{k=1}^K w_k \sum_{m=1}^M (x_{i,kM+m}^* - x_{j,kM+m}^*)^2 \right)^{1/2}. \quad (5)$$

Here K is the number of pixels in a spatial neighborhood, w_k are weighting coefficients.

Intuitively, we can suppose that greater values of coefficients should correspond to closer pixels, and smaller values should correspond to further ones. Also we can suppose that weighting coefficients don't depend on a direction.

In this case, such coefficients can be described by two-dimensional radial-symmetric window function. There were proposed a number of window functions that used widely in signal processing. Examples of such functions include Rectangle, Triangle, Gauss, Bartlet, Welch, Blackman, Hann, Nuttall window functions and others. In this paper, we use window functions based on the following equation (Fig. 1):

$$w(r) = \begin{cases} 1 - (r/R)^p, & \text{if } (r \in [0; R]) \text{ and } (p \geq 0); \\ (1 - r/R)^{-p}, & \text{if } (r \in [0; R]) \text{ and } (p < 0); \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

It should be noted that for particular values of p , this equation leads to some special windows: triangle window for $p = 1$, rectangle window for $p = \infty$.

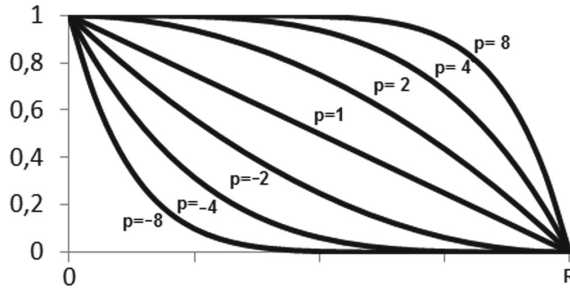


Fig. 1. Window functions.

Order Statistics. Another way of exploiting contextual information consists in using of the order statistics. Let us consider the sample x_1, x_2, \dots, x_K . By sorting the sample and reindexing the values of the sample so that $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(K)}$, we obtain the set of the order statistics $x_{(i)}$. The first and the K -th order statistics are $x_{(1)} = \min\{x_i\}$, $x_{(K)} = \max\{x_i\}$ correspondingly.

To embed the contextual information, for each pixel x_i we consider a spatial neighborhood of radius R , and order pixels of the neighborhood by Euclidean distances to the pixel x_i in a spectral space. Then the feature space is extended by the first S order statistics (x_i is included in the neighborhood as $x_{(1)}$):

$$x_i^* = (x_{(1)}, x_{(2)}, \dots, x_{(S)}). \quad (7)$$

The modified dissimilarity measure takes the form:

$$\rho(x_i^*, x_j^*) = \left(\sum_{s=1}^S w_s \sum_{m=1}^M (x_{i,sM+m}^* - x_{j,sM+m}^*)^2 \right)^{1/2}. \quad (8)$$

Here we use inverse weighting $w_s = 1/s$, and S is the number of the order statistics used as a spatial context.

3 Experimental Study

In our experiments, we used open and well-known hyperspectral remote sensing scenes [10]. Here we provide experimental results for Indian Pines scene (Fig. 2(a)), 145×145 pixels, 224 spectral bands acquired using AVIRIS sensor. Only 200 bands were selected by removing bands with the high level of noise and water absorption. This hyperspectral scene is provided with the groundtruth segmentation mask (Fig. 2(b)) that was used to evaluate the quality of classification and segmentation.

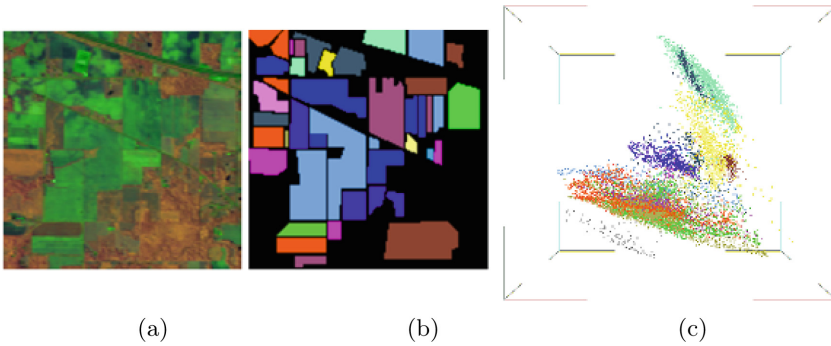


Fig. 2. Indian Pines scene: (a) false color representation of the image produced using nonlinear mapping; (b) ground truth image (classified pixels are shown with colors); (c) image pixels mapped into 3D space using nonlinear mapping and spatial context (Color figure online).

3.1 Classification of Hyperspectral Image Data

In this section we present the evaluation of described approaches to exploiting spatial context in terms of a classification quality. We use two well-known classifiers for features obtained using the described nonlinear mapping algorithm, and exploiting spatial context. In the evaluation we use the following classifiers: k-nearest neighbors classifier (NN), and support vector machine (SVM). To perform experiments we divide the whole set of ground truth samples into a training subset, containing 60% of the sample, and a test subset, containing 40% of the sample.

Window Functions. Some results devoted to the study of window functions are shown in the Fig. 3. In all present cases the leftmost value of p corresponds to the window function

$$\delta(r) = \begin{cases} 1, & \text{if } (r = 0), \\ 0, & \text{else,} \end{cases} \quad (9)$$

which means that no spatial context is exploited in the nonlinear mapping. The rightmost value corresponds to the rectangular window function (*Rect*), which

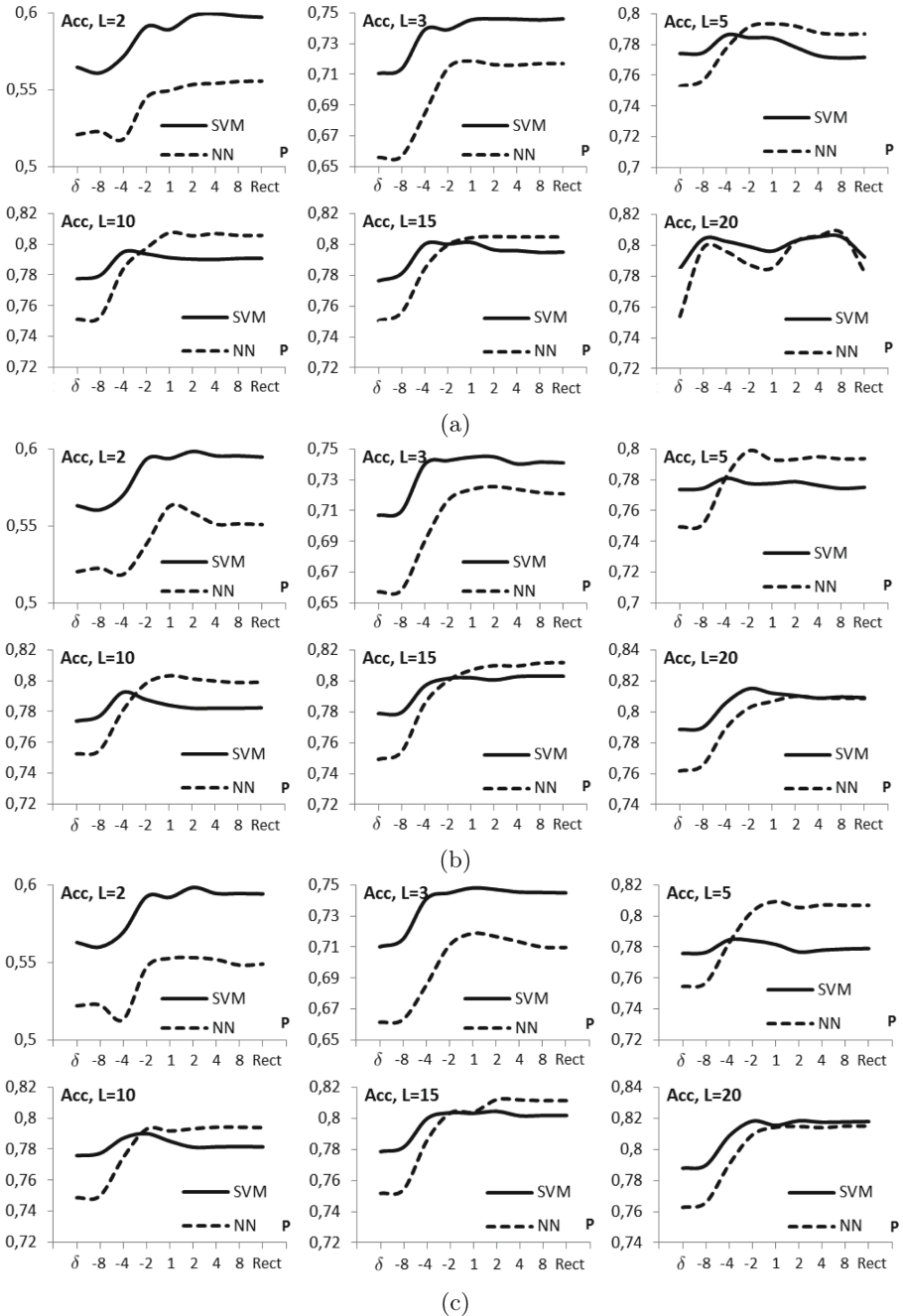


Fig. 3. The dependency of the classification accuracy (Acc) on the parameter p of the window function (6) for the specified dimensionality L of the reduced space R^L , and different neighborhood radius R : radius $R = 1$ (a), neighborhood radius $R = 2$ (b), neighborhood radius $R = 3$ (c). Parameter values δ and Rect correspond to special cases of window function: delta function (9) and rectangle window function.

is equal to 1 for all pixels inside a neighborhood. The intermediate value $p = 1$ corresponds to triangle window function.

As it can be seen, in almost all considered cases the classification accuracy was significantly improved by exploiting the spatial context with $p \geq 1$ over the accuracy at δ (without spatial context). The triangle window ($p = 1$) can be considered as a good choice for all considered examples.

Order Statistics. Results devoted to the study of the approach based on the order statistics are shown in the Fig. 4. Here the leftmost value $S = 1$ corresponds to the case when only central pixel of a spatial neighborhood was used, and no spatial context was exploited. The value $S = 2$ means that the first nearest (in a spectral space) neighbor from spatial neighborhood was used as a spatial context, and so on. The rightmost values of S correspond to cases, when the whole spatial neighborhood was used in calculations.

As it can be seen from the figure, in all considered cases the classification accuracy was significantly improved by exploiting the spacial context again. In almost all considered examples, the more order statistics was used, the better was the classification quality.

It is worth noting that in both experiments (window functions and order statistics) a greater gain is observed for a nearest neighbor classifier compared to SVM. In some cases NN classifier outperformed SVM in a reduced space. We can explain this by the fact that the reduced space is formed by the nonlinear mapping, which operates on the principle of preserving pairwise distances between pixels (in hyperspectral and reduced spaces).

Comparison to Principal Component Analysis. In this subsection we compare two approaches to incorporating of the spatial context, and the widely used linear Principal component analysis technique. Principal component analysis (PCA) technique finds a linear projection to a lower dimensional subspace maximizing the variation of data. PCA is often thought of as a linear dimensionality reduction technique minimizing the information loss. The results of the comparison are shown in the Fig. 5. As it can be seen, both approaches outperform the standard PCA technique, and the nonlinear mapping with the order statistics was better than the nonlinear mapping with window function (maximum number of order statistics was used in this experiment, and $P = 1$ was used as a parameter for window function).

3.2 Clustering and Segmentation of Hyperspectral Image Data

In this section we present an evaluation of the described approach for clustering and segmentation of hyperspectral image.

A segmentation method adopted in this paper is based on a clustering technique, and is quite straightforward. It consists of two steps. At first, the clustering of image pixels is performed in the reduced space. At this stage, a clustering algorithm partitions a set of image pixels into some number of subsets, according to pixels features. At second, an image markup procedure extracts connected

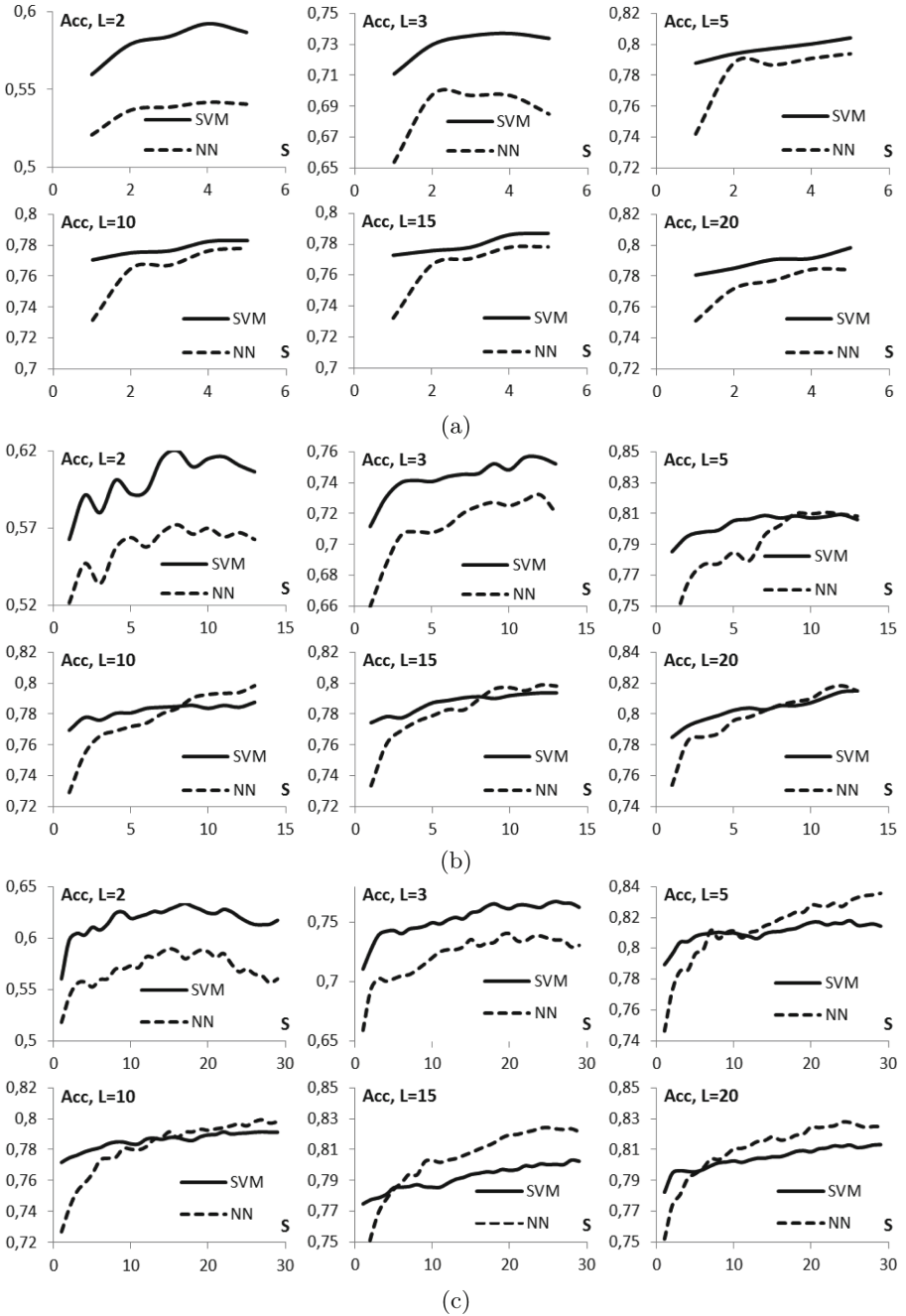


Fig. 4. The dependency of the classification accuracy (Acc) on the number S of the order statistics for the specified dimensionality L of the reduced space R^L , and different neighborhood radius R : radius $R = 1$ (a), neighborhood radius $R = 2$ (b), neighborhood radius $R = 3$ (c).

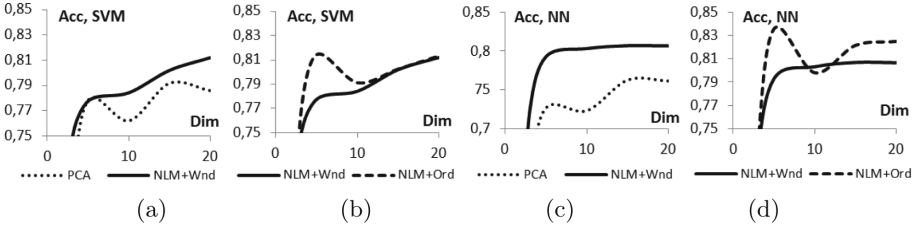


Fig. 5. The dependency of the classification accuracy (Acc) on the dimensionality Dim of the reduced space (R^L) for different techniques: SVM classifier with PCA and NLM + Window based spatial context (a), SVM classifier with NLM + window based spatial context and NLM + order statistics (b), NN classifier with PCA and NLM + Window based spatial context (c), NN classifier with NLM + window based spatial context and NLM + order statistics (d).

regions of an image containing pixels of corresponding clusters. There is a number of clustering algorithms belonging to the following classes [11]: hierarchical clustering, density-based clustering, spectral clustering, etc. While a number of clustering algorithms have been proposed, the well-known k-means algorithm [12] remains the most frequently mentioned approach. In this paper, we used this algorithm with the squared Euclidean distance measure. To initialize cluster centers we used the k-means++ algorithm [13]. It was shown that k-means++ achieves faster convergence to a lower local minimum than the base algorithm. To obtain a satisfactory solution, we varied the number of clusters from 10 to 50. For each specified number of clusters we initialized and ran clustering for 5 times to get the best arrangement out of initializations.

To measure the quality of segmentation and clustering, we used one of the most commonly used measures - Rand index [14]. This measure is intended for the evaluation of clustering methods, and is defined as follows:

$$RI(S_1, S_2) = \frac{1}{\binom{N}{2}} \sum_{i,j,i \neq j} (I(l_i^1 = l_j^1 \wedge l_i^2 = l_j^2) + I(l_i^1 \neq l_j^1 \wedge l_i^2 \neq l_j^2)) \quad (10)$$

Here $I()$ is the indicator function, l_i^k is the label (segment) of the i -th pixel on the k -th segmentation. The denominator is the number of all possible unique pairs of N pixels.

The results of the study are shown in the Fig. 6. As it can be seen, exploiting spatial context in nonlinear mapping provide higher values of Rand index compared to linear PCA technique in all cases. Higher values of RI mean better clustering and segmentation results. Besides that, using the first approach provided about 25% less number of segments on average (not shown on the figure), that means less oversegmentation.

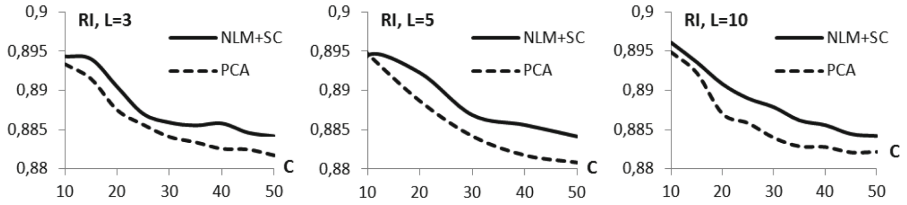


Fig. 6. The dependency of the Rand index (RI) on the number of clusters C for the specified dimensionality L of the reduced space (R^L), and dimensionality reduction techniques (nonlinear mapping with spatial context $NLM + SC$ and principal component analysis PCA).

3.3 Visualization of Hyperspectral Image Data

Described above approach to the dimensionality reduction can be used as a means of visual data analysis. By specifying the dimensionality $L = 2$ or $L = 3$ of the output space R^L , we can obtain 2D or 3D mappings of hyperspectral data. An example of 3D mapping of the test hyperspectral image is shown in the Fig. 2(c). We do not provide an evaluation of the quality of such mapping as this is out of the scope of the paper.

Besides that, dimensionality reduction techniques operating on the principle of preserving the pairwise distances between image pixels, can be applied to make hyperspectral image representation with special properties. An example of such representation is shown in the Fig. 2(a). Its distinguishing feature is that the distances between image pixels of the rendered image in color space approximate the distances between pixels in the source hyperspectral space. Similarly, the technique proposed in this paper can be used to produce false color representations of hyperspectral images using nonlinear mapping and spatial context.

4 Conclusion

In this work we proposed and evaluated the unsupervised nonlinear mapping technique, which exploit the spatial context of hyperspectral images. We evaluated two approaches to incorporating of the spatial context in the mapping technique. To evaluate the proposed technique, we considered three different tasks of hyperspectral image analysis, namely classification, segmentation, and visualization of hyperspectral images. Experimental study showed that proposed approaches can be successfully applied in hyperspectral image analysis.

Acknowledgements. The reported study was funded by RFBR according to the research project no. 16 – 29 – 09494 ofi.m, 16 – 37 – 00202–mol.a.

References

1. Richards, J.A., Jia, X., Ricken, D.E., Gessner, W.: Remote Sensing Digital Image Analysis: An Introduction. Springer, New York Inc. (1999)
2. Wang, J., Chang, C.-I.: Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.* **44**(6), 1586–1600 (2006)
3. Bachmann, C.M., Ainsworth, T.L., Fusina, R.A.: Improved manifold coordinate representations of large-scale hyperspectral scenes. *IEEE Trans. Geosci. Remote Sens.* **44**(10), 2786–2803 (2006)
4. Kim, D.H., Finkel, L.H.: Hyperspectral image processing using locally linear embedding. In: First International IEEE EMBS Conference on Neural Engineering, pp. 316–319 (2003)
5. Doster, T., Olson, C.C.: Building robust neighborhoods for manifold learning-based image classification and anomaly detection. In: Proceedings SPIE, vol. 9840, p. 984015 (2016)
6. Myasnikov, E.V.: Nonlinear mapping methods with adjustable computational complexity for hyperspectral image analysis. In: Proceedings of the SPIE, vol. 9875, p. 987508 (2015)
7. Sammon, J.W.: A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.* **18**(5), 401–409 (1969)
8. Wang, L., Shi, C., Diao, C., Ji, W., Yin, D.: A survey of methods incorporating spatial information in image classification and spectral unmixing. *Int. J. Remote Sens.* **37**(16), 3870–3910 (2016)
9. Myasnikov, E.: Evaluation of stochastic gradient descent methods for nonlinear mapping of hyperspectral data. In: Campilho, A., Karray, F. (eds.) ICIAR 2016. LNCS, vol. 9730, pp. 276–283. Springer, Cham (2016). doi:[10.1007/978-3-319-41501-7_31](https://doi.org/10.1007/978-3-319-41501-7_31)
10. Hyperspectral Remote Sensing Scenes. http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes
11. Cariou, C., Chehdi, K.: Unsupervised nearest neighbors clustering with application to hyperspectral images. *IEEE J. Select. Top. Sig. Process.* **9**(6), 1105–1116 (2015)
12. Lloyd, S.P.: Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982)
13. David, A., Vassilvitskii, S.: K-means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035 (2007)
14. Rand, W.M.: Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* **66**(336), 846–850 (1971)