



The use of KPCA over subspaces for cross-scale superpixel based hyperspectral image classification

Haoyang Yu, Zhen Xu, Yulei Wang, Tong Jiao & Qiandong Guo

To cite this article: Haoyang Yu, Zhen Xu, Yulei Wang, Tong Jiao & Qiandong Guo (2021) The use of KPCA over subspaces for cross-scale superpixel based hyperspectral image classification, Remote Sensing Letters, 12:5, 470-477, DOI: [10.1080/2150704X.2021.1897180](https://doi.org/10.1080/2150704X.2021.1897180)

To link to this article: <https://doi.org/10.1080/2150704X.2021.1897180>



Published online: 22 Mar 2021.



Submit your article to this journal [↗](#)



Article views: 72



View related articles [↗](#)



View Crossmark data [↗](#)

ARTICLE



The use of KPCA over subspaces for cross-scale superpixel based hyperspectral image classification

Haoyang Yu^{id}^a, Zhen Xu^a, Yulei Wang^{a,b}, Tong Jiao^c and Qiangdong Guo^d

^aCenter of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China; ^bState Key Laboratory of Integrated Services Networks, Xidian University, Xian, China; ^cGraduate School of Geography, Clark University, Worcester, MA, USA; ^dSchool of Geosciences, University of South Florida, Tampa, FL, USA

ABSTRACT

This paper introduces a new object-based spectral-spatial classification method for hyperspectral image. The kernel principal component analysis (KPCA) is firstly performed over subspaces (KPCAsub) derived from the original spectral domain, which incorporates linear information with nonlinear formulation. The obtained image is then processed via a feature-level fusion with superpixel segmentation at different scales. The final classification result is achieved by a cross-scale superpixel based (CSSP) decision fusion framework based on each individual operation of support vector machine. The resulting method, called KPCAsub-CSSP, contributes to better characterization under-limited sample condition, and promotes spectral-spatial integration in terms of echoing the complex distribution of ground objects. The experimental results on two real hyperspectral data sets demonstrate that the proposed method exhibits good performance in comparison to the other related methods.

ARTICLE HISTORY

Received 18 June 2020

Accepted 17 February 2021

1. Introduction

With the advancement of earth observation systems onboard satellites and aircraft (Zalpour, Akbarizadeh, and Alaei-Sheini 2020), multimodal data has been available to characterize land surface, providing rich information for distinguishing objects and monitoring changes (Samadi, Akbarizadeh, and Kaabi 2019). Hyperspectral remote sensing, as one of the advanced environmental monitoring techniques, is featured by acquiring data at a high spectral resolution and thus, allows subtle differences of earth surface to be captured. It has been widely used in crop yield estimation, mineral exploration, and target identification. However, hyperspectral images (HSI) usually contain redundant information for classification because of the high correlation among spectral bands, and data at high spectral resolution tend to result in the Hughes phenomenon, especially given limited training samples. Thus, HSIs are usually pre-processed to reduce the dimension of spectral bands in most of the state-of-the-art

CONTACT Yulei Wang  wangyulei@dlnu.edu.cn  Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian 116026, China

 Supplemental data for this article can be accessed [here](#).

© 2021 Informa UK Limited, trading as Taylor & Francis Group

algorithms for classification. Principal component analysis (PCA) and kernel PCA (KPCA) are two commonly used methods for pre-processing, which could effectively concentrate the information contained in raw data into a few independent components with high loads of information. PCA is powerful in optimal dimensionality reduction for linear data but does a poor work for non-linear datasets (Schölkopf, Smola, and Müller 1997). In contrast, Kernel PCA improves the perform for non-linear datasets through mapping raw data into a linearly separable space with kernel functions and then applying PCA over newly created space (Jolliffe 2002). By incorporating both linear and non-linear information, KPCA enables a better representation of the original datasets. Furthermore, its close association to statistical classifiers like support vector machines facilitate the characterization and classification of high-dimensional data, including HSI. In addition, many other methods for dimensionality reduction have been proposed as well recently. For instance, Shao and Zhang (2014) developed a novel method based on semi-supervised local Fisher discriminant analysis (SELF), which can get an explicit feature mapping. Jiang et al. (2018) proposed a simple but very effective superpixelwise PCA (SuperPCA) approach, which can learn the intrinsic low-dimensional features of HSIs. Luo et al. (2020) proposed a hybrid-graph learning method to reveal the complex high-order relationships of the HSI, termed enhanced hybrid-graph discriminant learning (EHGDL). The above methods have also achieved good effects in dimensionality reduction.

Besides the spectral information, spatial features retrieved from HSI also provide useful information for distinguishing complex objects on the ground. Object-based image classification (OBIC) is such a framework which employs not only spectral but also spatial features to partition the image into spatially non-overlapping regions, also known as segments (Li et al. 2016). Each segment is defined as a superpixel and regarded as a processing unit in classification. Simple linear iterative clustering algorithm with 0 parameter (SLIC0) is one of the widely used algorithms for image segmentation, which is known to provide a sound trade-off between computational cost and classification accuracy (Achanta et al. 2012). As the accuracy of classification depends on the size of the training samples used, it is critical to find the optimal number that balances accuracy and the time consumed for collecting training data (Tirandaz, Akbarizadeh, and Kaabi 2020).

In this letter, an object-oriented cross-scale superpixel based (CSSP) classification method based on KPCA over subspaces of hyperspectral image (KPCASub-CSSP) was proposed. The method consists of three steps: (1) Extract spectral features with KPCA over subspaces grouped by cross-correlation. (2) Partition original images into superpixels at multiple scales using SLIC0. 3) Apply SVM to spectral images aggregated over superpixels at each scale and generate the final classification map with a cross-scale decision fusion mode. Our proposed method improves the classification of HSI with limited training samples by integrating spectral and spatial features across multiple scales.

2. Proposed Framework

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d]$ denotes the n samples of a HSI, where $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,d}]^T$ is the spectral vector associated with the pixel $\mathbf{x}_i \in \mathbf{X}$ and composed of d spectral bands.

2.1. KPCA over subspaces

PCA projects the dataset in a space that maximizes the variance along each orthogonal basis vectors. As a linear algorithm, it fails to retrieve nonlinear information in HSIs. KPCA solves this problem by introducing the kernel function to PCA, such as the Radial Basis Function (RBF) Kernel. It allows the algorithm to project the data into a higher dimensional space and thus retrieve higher-order statistics. However, it also leads to the reverse problem encountered with PCA. In this context, our method is proposed to perform KPCA over subspaces created based on the similarity of spectral bands.

The linear information shared among all the bands can be measured by computing the normalized cross-correlation matrix $\mathbf{C} \in d \times d$. Its symmetrical element $C_{k,l}$ or $C_{l,k}$ is defined as the following equations, where k and l denote the band indexes.

$$C_{k,l} = \frac{1}{(n-1)\sqrt{\sigma_k\sigma_l}} \sum_{i=1}^n |(x_{i,k} - \mu_k)(x_{i,l} - \mu_l)| \tag{1}$$

$$\sigma_k = \frac{1}{n-1} \sum_{i=1}^n (x_{i,k} - \mu_k)^2 \tag{2}$$

$$\mu_k = \frac{1}{n} \sum_{i=1}^n x_{i,k} \tag{3}$$

where n and d denote the number of pixels and bands, and $x_{i,k}$ represents the value of \mathbf{x}_i in band k . Since \mathbf{C} is a symmetrical greyscale image, the subsequent operations can be implemented in either horizontal or vertical directions. The purpose of follow-up process is to group similar bands and divide them into different subspaces. As shown in Figure 1, a vertical Sobel edge filter together with a morphological filter is performed on \mathbf{C} , such that to estimate the approximation of the position of the subspace boundary. The obtained binary image is denoted as matrix \mathbf{B} , whose rows can be considered as a possible linear decomposition of the spectral space in subspaces. Thus, the vector \mathbf{V} is computed as

$$V_i = \sum_{j=1}^d B_{j,i} \tag{4}$$

here each component $V_i, 1 \leq i \leq d$, represents the length of time to find the boundary of the position of the spectral band i . A mean threshold is applied on \mathbf{V} and the positions of the final boundaries are stored in a binary vector \mathbf{F} as

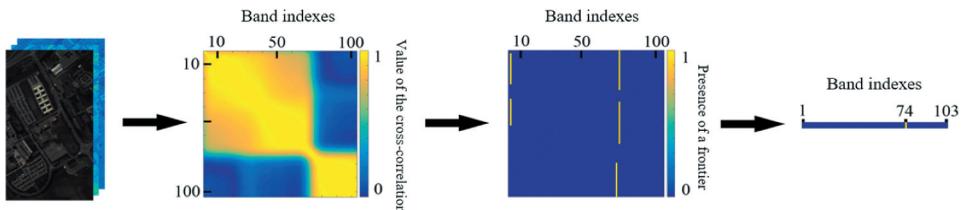


Figure 1. The procedures of KPCA over subspaces. (a) is hyperspectral image of ROSIS Pavia University. (b) is the normalized cross-correlation matrix \mathbf{C} . (c) is the matrix \mathbf{B} . (d) is the vector \mathbf{F} .

$$F_i = \begin{cases} 1 & \text{if } V_i \mu_V \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$\mu_V = \frac{1}{N} \sum_{i=1}^d V_i \quad (6)$$

$$N = \sum_{i=1}^d \frac{V_i}{V_i} \quad (7)$$

where $F_i = 1$ denotes that the boundary of a subspace lies in the position of band i .

2.2. Cross-scale decision fusion

The spatial distribution of various materials in real images is usually different, which leads to the fusion and integration in a fixed scale is not representative and may cause the imbalance among categories. In this case, cross-scale decision fusion is an effective framework, which is a two-step implementation that requires at least two different over-segmentation maps: one is with fine resolution and the other one is with coarse resolution. As illustrated in Figure 2, they are named as scale j and scale $j + 1$, which are obtained by SLIC0 algorithm with two different sets of superpixels. The superpixels with the finest scale j need to be classified beforehand. The first step is a spatial voting where the superpixels with scale $j + 1$ are used as reference for the superpixels with scale j . For each superpixel S_i in scale $j + 1$, the label of its contained pixels is identified as follows:

$$\text{class}(S_i) = \arg \max_{c \in [1, C]} \sum_{\mathbf{x} \in S_i} \mathbf{1}(\text{class}(\mathbf{x}), c) \quad (8)$$

where $\mathbf{1}(x)$ is an indicator function. It equals to 1 when the classification result of sample \mathbf{x} in scale j equals to class c , where $c \in [1, C]$ and C indexes the total number of classes.

The second step is required only if three scales or more are used: it is a typical pixel-wise voting among all the spatial-scale results (Yu et al. 2017).

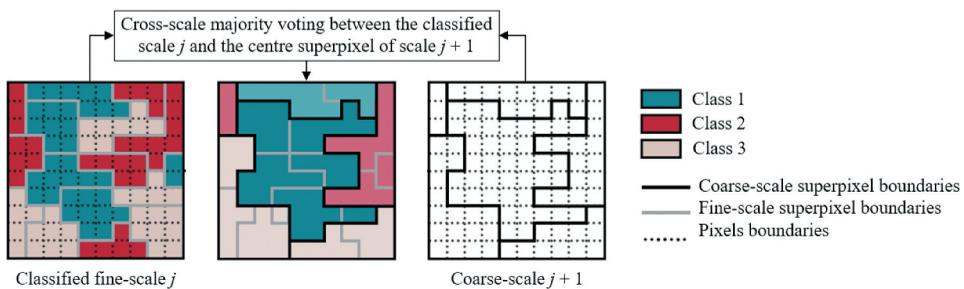


Figure 2. Cross-scale decision fusion process.

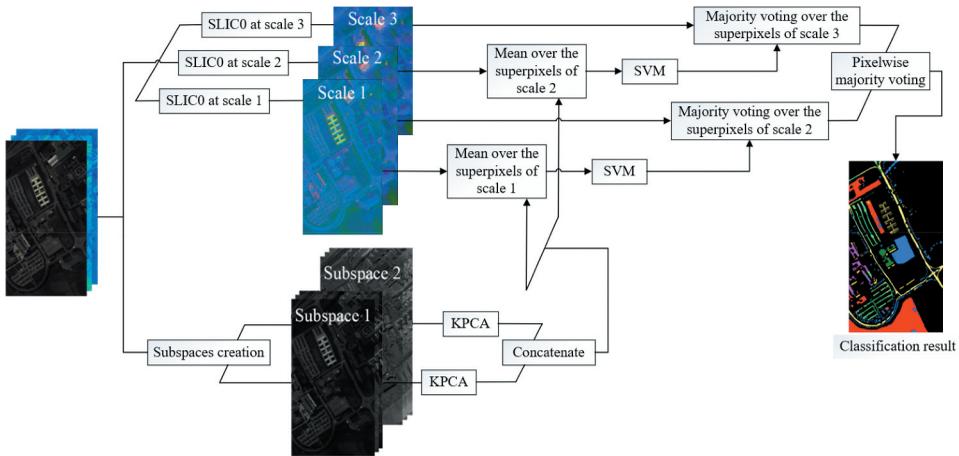


Figure 3. The general framework of the proposed KPCAsub-CSSP. (The settings follow the former definition and procedure in Figures 1 and 2).

2.3. KPCAsub-CSSP

The framework of the KPCAsub-CSSP is shown in Figure 3. At first, superpixels are computed by SLIC0 at different segmentation scales that are automatically chosen. Then, the size of superpixels is doubled from one scale to another. At last, the coarsest scale only contains the superpixels with about 10% of the size of the original HSI, to prevent the under-segmentation. Afterwards, for each scale, each superpixel is integrated with a feature-extracted image of KPCA over subspaces by means of feature averaging in the corresponding coverage region. All the scales except the coarsest are classified by SVM. After that, the cross-scale decision fusion process is implemented to integrate the results from different scale to generate the final classification map. From the structure of Figure 3, if the upper part only considers the processing of a single scale, it can be regarded as a single-scale version. Also, if the lower part does not have the processing of subspace projection, it can be regarded as a version without feature optimization. The whole framework is a comprehensive processing of spectral and spatial information.

3. Experimental Results

There are two well-known datasets used in the experiments: the Pavia University dataset and the Indian Pines dataset. The Pavia University dataset was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) in 2001. It consists of 103 spectral bands, ranging from 430 nm to 860 nm, with 12 noisy bands removed. Its spatial resolution is 1.3 m and it contains 610×340 pixels, with 42,776 labelled samples divided into 9 classes. 0.42–2.1% of all the samples were randomly drawn to construct the training set.

The second real dataset is the Indian Pines dataset with 145×145 pixels captured in 1992 by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The spatial resolution is 20 m and the covered wavelength ranges between 400 nm and 2500 nm with 220 spectral bands. Besides, 16 classes were derived from a total of 10,249 samples. Here, 1.6–7.8% of the samples were randomly selected to construct the training set.

For comparison purposes, several other related algorithms are presented: (1) A regular SVM classifier with an RBF kernel is used as a baseline for evaluation. (2) The KPCA-SSSP. The basic version of the framework without the processes of subspaces and cross-scale fusion. (3) KPCAsub-SSSP/PCAsub-SSSP, a single-scale version of the framework without the process of cross-scale fusion. (4) KPCA-CSSP. The cross-scale framework without the process of subspaces. (5) KPCAsub-CSSP/PCAsub-CSSP. The proposed framework with KPCA/PCA for dimensionality reduction. Considering the variation of different single-scale versions, the result of the best scale is presented.

Overall accuracy (OA) is used for the accuracy assessment. Table 1 shows that PCAsub-SSSP always obtains the highest accuracy among the algorithms only using single-scale for 12–18 features. However, the performance of KPCAsub-SSSP is very close to that of PCAsub-SSSP when using 24 or 30 features. In addition, KPCAsub-SSSP constantly outperforms KPCA-SSSP, which indicates the usefulness of the subspaces for the classification. We selected 12–30 features in the experiments, because less than 12 features are not enough for these algorithms, and with more than 30 features, the OA statistics started to decrease. No matter how many training samples were selected, with 12 to 18 features, PCAsub-CSSP is always the best classifier; whereas with 24–30 features KPCAsub-CSSP becomes the leader. Figure 4 shows the classification maps obtained by different tested methods with 900 training samples (100 samples per class). Based on the results of Figure 4 and Table 1, comparing the single-scale methods, the cross-scale methods can better incorporate the spatial information, which leads to increases in both the OA statistics and classification performances.

Table 2 shows the classification results using the Indian Pines dataset. Similar findings can be drawn from it, but the gap between the single-scale results and the cross-scale results is smaller than that of the first dataset due to the low spatial resolution of the Indian Pines dataset. Moreover, the regular SVM is always the worst classifier for this

Table 1. Overall Accuracies (in percent) for the ROSIS Pavia University scene. The table presents the mean results of 10 tests (The best results are highlighted in bold).

Num. of Training Samples	SVM on 103 bands	Num. of Features	KPCA-SSSP	PCAsub-SSSP	KPCAsub-SSSP	KPCA-CSSP	PCAsub-CSSP	KPCAsub-CSSP
180	79.3%	12	69.8%	79.4%	73.5%	78.6%	86.3%	82.3%
		18	74.7%	76.3%	72.7%	81.3%	86.1%	82.7%
		24	75.5%	80.0%	79.3%	83.6%	85.4%	87.0%
360	82.3%	30	79.6%	78.7%	80.1%	86.4%	85.8%	87.5%
		12	74.1%	86.2%	79.2%	86.7%	92.1%	86.7%
		18	81.0%	87.1%	82.1%	88.8%	93.4%	89.8%
540	85.8%	24	82.5%	88.5%	86.9%	90.5%	93.8%	94.3%
		30	85.0%	88.8%	89.5%	91.7%	93.5%	94.5%
		12	79.1%	88.9%	82.3%	87.9%	94.3%	89.1%
720	86.8%	18	82.8%	90.8%	84.6%	90.2%	95.4%	91.4%
		24	85.2%	90.8%	89.7%	92.3%	95.2%	95.8%
		30	86.7%	90.1%	90.6%	93.2%	95.2%	96.4%
900	88.6%	12	78.9%	91.5%	83.9%	86.7%	95.5%	90.3%
		18	84.6%	91.3%	86.0%	92.2%	95.6%	93.0%
		24	87.4%	91.3%	91.1%	94.1%	95.9%	96.9%
		30	88.9%	92.1%	92.6%	94.7%	96.1%	97.4%
		12	79.6%	92.1%	84.8%	87.5%	95.9%	91.0%
		18	84.7%	92.6%	87.3%	92.5%	96.4%	94.1%
		24	86.8%	92.4%	92.5%	92.9%	96.5%	97.4%
		30	89.2%	93.1%	93.2%	95.0%	96.6%	97.8%

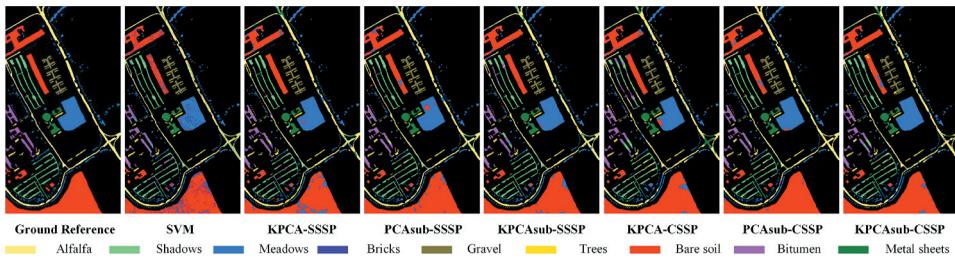


Figure 4. Classification maps obtained by different tested methods for the ROSIS University of Pavia data set. In all cases, 900 training samples in total (100 samples per class) were used.

Table 2. Overall Accuracies (in percent) for the AVIRIS Indian Pines Scene, the results correspond to the mean over 10 tests (The best results are highlighted in bold).

Num. of Training Samples	SVM on 200 bands	Num. of Features	KPCA-SSSP	PCAsub-SSSP	KPCAsub-SSSP	KPCA-CSSP	PCAsub-CSSP	KPCAsub-CSSP
160	48.7%	12	61.7%	65.2%	64.5%	64.2%	73.4%	72.3%
		18	65.0%	61.3%	64.4%	67.6%	71.5%	72.1%
		24	64.8%	66.1%	66.0%	68.6%	72.2%	72.7%
		30	65.2%	65.2%	67.0%	68.7%	71.6%	71.3%
320	63.6%	12	73.2%	75.5%	76.3%	77.7%	82.9%	82.3%
		18	76.0%	76.8%	79.0%	79.6%	84.3%	84.2%
		24	76.7%	77.1%	78.3%	82.2%	84.7%	84.4%
		30	76.4%	77.0%	78.2%	82.5%	85.3%	84.8%
480	68.8%	12	78.5%	80.6%	81.5%	83.4%	88.4%	87.2%
		18	82.7%	82.9%	83.9%	86.8%	89.8%	90.0%
		24	80.7%	83.4%	84.2%	87.5%	89.4%	89.2%
		30	81.3%	83.3%	84.3%	88.2%	90.7%	90.8%
640	71.1%	12	81.4%	83.9%	84.3%	86.3%	90.9%	89.9%
		18	84.4%	85.5%	86.1%	89.0%	91.6%	91.6%
		24	83.5%	86.2%	87.2%	90.2%	91.6%	91.6%
		30	85.0%	85.4%	85.8%	91.5%	91.9%	91.8%
800	75.0%	12	84.1%	85.4%	86.2%	88.6%	92.3%	91.4%
		18	87.0%	87.9%	88.9%	91.3%	93.6%	93.5%
		24	86.2%	87.2%	87.8%	92.7%	94.3%	93.9%
		30	86.9%	88.1%	88.2%	92.9%	95.4%	95.1%

scene, struggling with the Hughes phenomenon and the noise. These two problems can be eliminated by the proposed algorithms.

4. Conclusion

In this letter, a new object-oriented spectral-spatial method, called KPCA-CSSP, was proposed for hyperspectral image classification. Its major advantages include applying KPCA to subspaces for feature extraction and employing a cross-scale superpixel-level fusion framework for comprehensive information integration. The proposed method has been compared with RBF kernel-based SVM and other related methods for the classification of two real datasets. Experimental results show that our method significantly improves the classification performance of HSI compared to other methods.

Funding

This work was supported by the Chinese Postdoctoral Science Foundation (No. 2020M680925), the Open Research Fund of Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (No. 2020LDE001), and the National Nature Science Foundation of China (No. 61801075).

ORCID

Haoyang Yu  <http://orcid.org/0000-0002-4026-7450>

References

- Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. 2012. "SLIC Superpixels Compared to State-of-the-art Superpixel Methods." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (11): 2274–2282. doi:10.1109/TPAMI.2012.120.
- Jiang, J., J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang. 2018. "SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 1–13. doi:10.1109/TGRS.2018.2828029.
- Jolliffe, I. T. 2002. "Principal Component Analysis for Special Types of Data." *Principal Component Analysis* 339–372. doi:10.1007/978-1-4757-1904-8_11.
- Li, S., L. Ni, X. Jia, L. Gao, B. Zhang, and M. Peng. 2016. "Multi-scale Superpixel Spectral–spatial Classification of Hyperspectral Images." *International Journal of Remote Sensing* 37(19-20) (20): 4905–4922. doi:10.1080/01431161.2016.1225175.
- Luo, F., L. Zhang, B. Du, and L. Zhang. 2020. "Dimensionality Reduction with Enhanced Hybrid-graph Discriminant Learning for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 99: 1–18. doi:10.1109/TGRS.2020.2963848.
- Samadi, F., G. Akbarizadeh, and H. Kaabi. 2019. "Change Detection in SAR Images Using Deep Belief Network: A New Training Approach Based on Morphological Images." *IET Image Processing* 13 (12): 2255–2264. doi:10.1049/iet-ipr.2018.6248.
- Schölkopf, B., A. Smola, and K. R. Müller. 1997. "Kernel Principal Component Analysis." *ICANN* 583–588. doi:10.1007/BFb0020217.
- Shao, Z., and L. Zhang. 2014. "Sparse Dimensionality Reduction of Hyperspectral Image Based on Semi-supervised Local Fisher Discriminant Analysis." *International Journal of Applied Earth Observation and Geoinformation* 31: 122–129. doi:10.1016/j.jag.2014.03.015.
- Trandaz, Z., G. Akbarizadeh, and H. Kaabi. 2020. "PolSAR Image Segmentation Based on Feature Extraction and Data Compression Using Weighted Neighborhood Filter Bank and Hidden Markov Random Field-expectation Maximization." *Measurement* 153. doi:10.1016/j.measurement.2019.107432.
- Yu, H., L. Gao, W. Liao, B. Zhang, A. Pizurica, and W. Philips. 2017. "Multiscale Superpixel-level Subspace-based Support Vector Machines for Hyperspectral Image Classification." *IEEE Geoscience and Remote Sensing Letters* 14 (11): 2142–2146. doi:10.1109/LGRS.2017.2755061.
- Zalpour, M., G. Akbarizadeh, and N. Alaei-Sheini. 2020. "A New Approach for Oil Tank Detection Using Deep Learning Features with Control False Alarm Rate in High-resolution Satellite Imagery." *International Journal of Remote Sensing* 41 (6): 2239–2262. doi:10.1080/01431161.2019.1685720.