Three-Dimensional Spatial-Spectral Filtering Based Feature Extraction for Hyperspectral Image Classification

Hasan Ali AKYÜREK[1], Barış KOÇER[2]
[1]*Department of Management Information Systems, School of Applied Sciences, Necmettin Erbakan University, Konya, Turkey*
[2]*Department of Computer Engineering, Faculty of Engineering, Selcuk University, Konya, Turkey*
hakyurek@konya.edu.tr

***Abstract*—Hyperspectral pixels which have high spectral resolution are used to predict decomposition of material types on area of obtained image. Due to its multidimensional form, hyperspectral image classification is a challenging task. Hyperspectral images are also affected by radiometric noise. In order to improve the classification accuracy, many researchers are focusing on the improvement of filtering, feature extraction and classification methods. In the context of hyperspectral image classification, spatial information is as important as spectral information. In this study, a three-dimensional spatial-spectral filtering based feature extraction method is presented. It consists of three main steps. The first is a pre-processing step, which include spatial-spectral information filtering in three-dimensional space. The second comprises extract functional features of filtered data. The last one is combining extracted features by serial feature fusion strategy and using to classify hyperspectral image pixels. Experiments were conducted on two popular public hyperspectral remote sensing image, 1%, 5%, 10% and 15% of samples of each classes used as training set, the remaining is used as test set. The proposed method compared with well-known methods. Experimental results show that the proposed method achieved outstanding performance than compared methods in hyperspectral image classification task.**

***Index Terms*—adaptive algorithms, feature extraction, gaussian noise, hyperspectral imaging, image classification.**

I. INTRODUCTION

The classification of hyperspectral images is a challenging task, since the acquired hyperspectral images, which provides huge amount of spectral and spatial information, are often affected by radiometric noise such as sensor noise, photon (or pulse) noise, calibration error, atmospheric scattering and absorption. So, that is not only reduce the visual quality of hyperspectral image data, but also reduces the sensitivity of image interpretation and analysis. Thus, denoising is a very important process in hyperspectral image classification. Also, denoising can improve spatial classification accuracy when used with spatial processing techniques.

Therefore, denoising in hyperspectral image classification has attracted attention of the researchers. For example Wei He et al. proposed an iterative regularization framework to remove noise from hyperspectral images named as noise adjusted iterative low-rank matrix approximation (NAILRMA)[1]. Aswathy et al. proposed a low pass sparse banded filter matrix for improved sparsity based hyperspectral image classification[2]. Cui et al. proposed edge-preserving filtering method for classification visible and infrared hyperspectral images[3]. Srivatsa et al. applied least square denoising method to improve Alternating Direction Method of Multipliers based hyperspectral image classification[4]. Huang et al. presents a hybrid spatial-spectral denoising method for infrared hyperspectral image classification using 2-dimensional principal component analysis[5]. Edge-preserving filtering[6] and bilateral filtering[7], were also proposed as denoising methods for hyperspectral image classification. Apart from using denoising methods directly, feature extraction methods and designing effective classifier methods are also proposed as an effective way to improve the hyperspectral classification accuracy.

Consequently, many feature extraction methods such as manifold learning algorithms are proposed for hyperspectral image classification in recent years. Chang et al.[8] proposed nearest feature line embedding approach. Huang et al.[9] utilized double nearest proportion feature extraction method to reduce the dimensionality of hyperspectral images. Tu et al.[10] proposed dimensionality reduction method based laplacian eigenmaps for SAR image classification. Li et al.[11] applied the Local preserving projection (LPP) method for hyperspectral image classification. Some supervised methods are proposed to extract the features by using the label information[12, 13]. Wei et al.[14] implements functional principal component analysis (FPCA) for extracting features of hyperspectral images. According to Wei et al.[14] hyperspectral image with narrow bandwidth values is approximated to a continuous function. The functional properties of hyperspectral images can be explored by this idea.

Recently, support vector machines (SVM)[15] based classification methods are commonly used for hyperspectral image classification due to better classification accuracy than artificial neural network(ANN)[16]. Therefore, researchers are focused on important strategies for designing SVM such relevance kernel function selection, kernel parameter selection, regularization parameter selection and multiclass classification[17]. Demir et al.[18] proposed relevance vector machines(RVM) for hyperspectral image classification. Classification can be carried out much faster by using RVM-based methods. With relatively smaller relevance vector, recognition process can be so fast that it can be possible to realize real time applications. Huang et

Digital Object Identifier 10.4316/AECE.2017.02013

al.[19] proposed method that unifies support vector machine based methods and other regularization based methods to build up a new learning machine named extreme learning machine (ELM). Needing fewer optimization constraints, easy result implementation, faster learning and high generalization performance are some of advantages of ELM over SVM[20]. Pal et al.[21] proposed using kernel extreme learning machines(KELM) which replaces the hidden layer of the ELM with a kernel function for hyperspectral image classification.

In this paper, we proposed spatial-spectral filtering based feature extraction method for hyperspectral image classification. In the proposed method, we combine the advantages of functional data feature extraction method with three-dimensional gaussian filter and three-dimensional adaptive non-local means filter. Firstly, the hyperspectral data cube was filtered by three-dimensional gaussian filter and three-dimensional adaptive non-local means filter separately and output data of the filters represented as functional data. After functional representation, features extracted with functional principal component analysis (FPCA). Consequently, extracted features are combined by serial feature fusion strategy and pixel wise classification map which is created with Support Vector Machine (SVM) classifier.

The remainder of this paper is structured as follows. The related works are described in Section 2. The proposed spatial-spectral hyperspectral image classification method is introduced in Section 3. The effectiveness of the proposed method is demonstrated by experimental results on several popular hyperspectral images in Section 4. Finally, Section 5 presents the conclusions and future lines of research.

II. Related Work

Many image-denoising methods adjust intensity of each pixel in image by averaging spatial neighbors' intensity with some ways. Basic and simplified idea is changing pixel intensity value with an average of neighbor pixels intensities[22]. The biggest advantage of such filters is they can be used iteratively until enough amount of smoothing is acquired. So, they can be implemented very efficiently.

Everything is not perfect about that type of data independent filters which works by eliminating the different neighbor pixels from current pixel. Because they may blur edges, small structures and textures, which are important interests in the image. Therefore, basic statistical operators like median filter can be used for this purpose.

Sophisticated method like adaptive smoothing [23] which aims to enable edge detection after few filter iterations and anisotropic diffusion [24] which aims to keep region boundaries sharp are also filter samples. [25], [26] and variants [27, 28] are neighborhood filters which calculates average value of spatially close pixels that have similar intensity.

Michael Elad[29] studied the additive noise removal and developed locally adaptive recovery paradigm by bridging bilateral filter and adaptive methods. The paradigm advocates that new pixel value should only depend on the pixels in spatial neighborhood.

A. Non-Local Means Filter

According to Buades et al.[30] recently proposed algorithms are not applicable enough if image is not suitable for algorithms assumptions. In [30] methods reviewed the mathematical aspects of comparing image denoising algorithms and proposed a new one (non-local means (NOLM)). The main idea of NOLM based on any natural image has redundancy and that any pixel of the image has similar pixels that are not necessarily located in a spatial neighborhood.

NOLM calculates the weight of the pixels based on the intensity similarity of their neighborhoods as opposite to classical methods, which based on spatial proximity. So, the NOLM filter can be described as neighborhood filters with infinite spatial kernel. The similarity of the neighborhood intensities is substituted to the point-wise similarity of intensity value as in commonly used bilateral filtering. All the effort in NOLM is about to assure saving edges while removing noise.

B. Adaptive Non-Local Means Filter

The main drawback of the NOLM, especially for three-dimensional images, is the computational complexity. Let N^3 denote the size of three-dimensional image, then the complexity of the filter is in the order of $O((N(2M+1)(2d+1))^3)$. For a hyperspectral image of size 145 x 145 x 200 with the smallest values for d=1 and M=5, the computational time reaches up to 10 hours on 2.4GHZ CPU. For this reason, Manjon et al.[31] proposed Adaptive Non-Local Means (ANOLM) filter with spatially varying noise levels.

Furthermore, to speed up the filtering process a block wise approach which proposed in [32] can be used to decrease the computational complexity. NOLM like adjustment performed to these blocks and then pixels' values adjusted with blocks they belong to. This is different from the pixel wise version where only the central pixel of the neighborhood is averaged.

Avoiding the use of pixels/blocks with small weights speeds up the filter and significantly improves the denoising results. For the ensuring this Coupe et al. [32] proposed a preselection approach based on the local mean and variance of three-dimensional volumes.

According to [31] preselection based on the local mean is intensity sensitive: high and low intensity pixels are treated differently. In order to minimize such differences, the preselection was performed using the original means and the inverted means.

C. Three-Dimensional Gaussian Filter

Gaussian Filter is a denoising technique which uses the Gaussian function called as a normal distribution function in statistics, and is also called Gaussian average operator. In this filter, neighbor pixels averaged with weighted by function value. N-Dimensional Gaussian function can be defined as (1). Sigma is the standard deviation of distribution and acts as smoothing parameter. Mean of distribution is assumed as zero[33].

$$G_{ND}(\vec{x},\sigma)=\frac{1}{\left(\sqrt{2\pi}\sigma\right)^N}e^{-\frac{\left|\vec{x}^2\right|}{2\sigma^2}} \quad (1)$$

Gaussian filter can be performed One-dimensional version for each dimension in N-Dimensional data. Furthermore, Gaussian filter can be performed with different sigma values for each dimension.

Gaussian filter can be performed in the spatial domain and frequency domain. In the frequency domain, Fourier transform performed on image and filter matrix. After that, convolution of image and filter are performed. Finally, filtered image can be obtained from the inverse Fourier transform. Fig. 1 shows the workflow of filtering in the frequency domain.

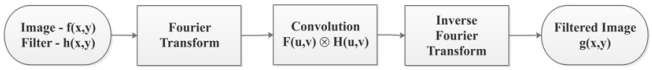

Figure 1. Workflow of filtering in the frequency domain.

D. Functional Data Analysis

Functional data analysis (FDA) is an alternative way to handle with huge dimensional data problems. Essentially functional data has infinite dimension. This situation presents challenges for both the theory and the computation; These challenges vary with how functional data is sampled. On the other hand, the infinite dimension of the data is a rich source of information that offers many opportunities for research and data analysis. The FDA has come very popular recently[34, 35]. Functional data basically represents discrete observation values by a continuous function, which can be calculated for any input value. Functional principal components analysis (FPCA) is proposed for handling with the curse of dimensionality problem in functional data. FPCA is an important dimension reduction method for functional data.

1) Functional Data Representations

Computational complexity is an important measurement in computer science. Therefore, analyzed data should be expressed in the finite size. Commonly finite basis function expansion is used in the functional data analysis. Firstly, data is transformed to a continuous function, which can be calculated for any input value by using basis function expansion. If the data is considered to have no error, interpolation method is used for this process. Else if data contains some errors due to measurement, smoothing method is used for this process [36].

2) Functional Principal Components Analysis

Purpose of functional principal components analysis as well as classical principal component analysis is resolving the curse of dimensionality problem. Functional Principal Components Analysis (FPCA) is a powerful method to identify the important parts of difference between curves. FPCA is exploring the relationships that were expected to be in the system and also not already noticed [36].

FPCA is getting a few orthogonal functions describing the changes in data efficiently instead orthogonal vectors in PCA applied to multivariate data. In this way, effective features of functional data may be obtained.

III. Proposed Method

In the context of hyperspectral image classification, spatial information is as important as spectral information. In order to include spatial information in classification process a three-dimensional filtering based feature extraction approach is proposed. Three-dimensional filters which in pixel wise filtering context, uses information on three-dimensional space for filtering selected pixel. In this context, we obtain new information which contains neighborhood effects.

The main idea behind the three-dimensional Gaussian filter is nearest neighborhoods more effective then farthest ones. In the three-dimensional gaussian filter, each pixel value updated with neighbor pixel values in three-dimensional space by using gaussian function. This provides local neighborhood information on classification context.

As we know any natural image has redundancy structures and that any pixel of the image has similar pixels that are not necessarily located in local neighborhood. In the adaptive non-local means filter, each block (set of pixels) values updated with intensity based neighbor block values in three-dimensional space. The Adaptive Non-Local Means filter is used to obtain nonlocal neighborhood information to use in classification.

As we know from our background, functional data representation is giving us a rich information source for analyzing high dimensional data. To handle with infinite dimensionality of functional data, Functional Principal Component Analysis is used to obtain orthogonal function which describe the changes in data efficiently for extracting effective features of functional data. We used FPCA for extracting features of data obtained from three-dimensional filters.

The general strategy of serial feature fusion is two feature vectors of F_1 and F_2 are concatenated together[37]. If m and n are the weightage parameters of F_1 and F_2, respectively, then according to serial feature fusion strategy the combined feature [m x F_1; n x F_2]. Due to creating final feature vector for classifying data we use serial feature fusion strategy. Obtained feature vectors describes hyperspectral data as well as effective for classification task.

Fig. 2 shows the workflow of proposed method.

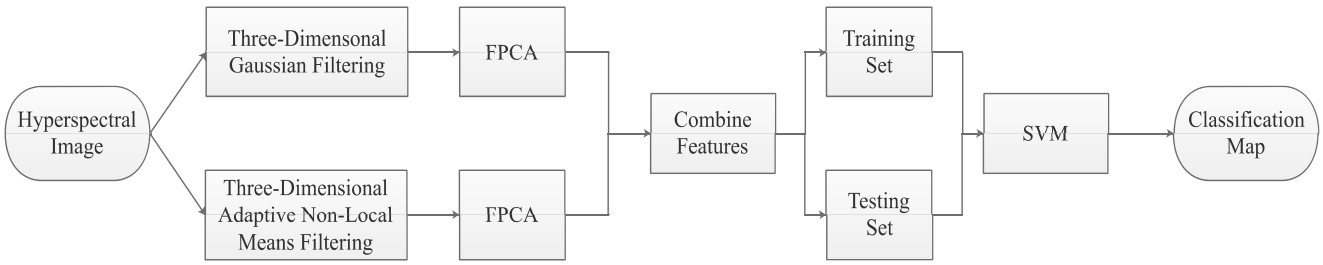

Figure 2. Workflow of proposed methods.

According to the above descriptions, our proposed method can be summarized in 5 main steps. These steps are Filtering, Functional Representation, Feature Extraction, Feature Fusion and Classification. In this method, we focused on combining the advantages of local neighborhood and non-local neighborhood. Pseudocodes of proposed method were listed in Table I.

TABLE I. PSEUDOCODES OF PROPOSED METHOD

Algorithm	**Three-Dimensional Spatial-Spectral Filtering based Feature Extraction**
Inputs	***HSI***: MxNxD sized Hyperspectral Image, σ: Gaussian filtering parameter, ***W***: Window size for gaussian filter, ***v***: Radius of search area for ANONLM filter, ***p1***: Radius of similarity patch (small) for ANONLM filter, ***p2***: Radius of similarity patch (big) for ANONLM filter, ***Rdim***: Size of reduced dimension, ***bType***: Basis type for functional representation, ***m,n***: Weight parameters for serial feature fusion
Output	Classification Map
Step 1	Filter input image by three-dimensional Gaussian filter and three-dimensional adaptive non-local means filter separately.
Step 2	Apply B-Spline curve smoothing method to both filter output data for representing spectral data as functional data.
Step 3	Apply Functional Principal Component Analysis method to obtained functional data for extracting functional features.
Step 4	Combine extracted features by serial feature fusion strategy for obtaining final feature vectors.
Step 5	Classify feature vectors by using RBF kernel based Support Vector Machine classifier.

In the step 1, input image is filtered by three-dimensional Gaussian filter and three-dimensional adaptive non-local means filter separately. At the step 2, B-Spline curve smoothing method applied to both filter output data for representing spectral data as functional data. In the step 3, FPCA is applied to obtained functional data for extracting functional features. Extracted features combined by serial feature fusion strategy at step 4 for obtaining final feature vectors. Finally, Feature vectors classified by using RBF kernel based Support Vector Machine classifier in step 5. The SVM was used due to its capability to deal with high-dimensional data. Its resilience, because of the kernel function, allows alternative strategies for including spatial features in the classification process such as feature fusion or composite kernels.

The proposed method combines the advantages of three-dimensional Gaussian filter and three-dimensional adaptive non-local means filter and functional principal components analysis method in SVM classifier.

IV. EXPERIMENTAL RESULTS

In this section, we provide the results obtained from applying the proposed method to two popular hyperspectral datasets which are Indian Pines[38] and Salinas Valley[39] described. The experiments were carried out on a computer with 16GB RAM and 2.40GHz i7-3630QM processor and the code was implemented in MATLAB. During the experiments on Indian Pines dataset, 1%, 5%, 10% and 15% of the samples of each class were randomly selected to create a training set and the remaining samples were taken as test samples.

In our experiments on Salinas dataset with 10% training samples is obtained nearly 100% accuracy. Therefore, 15% training set is not used in Salinas dataset. In the experiments, spectral reflectance values normalized to [0,1] and all parameters obtained by 10-fold cross validation. All reported classification accuracies are average of 10 experiments. Furthermore, classification accuracy of extracted features from different filters and different combinations is reported for showing the advantages of feature combination.

A. Experiments on Indian Pines Image

Indian Pines hyperspectral image obtained from Indian Pines test scene located at North-western Indiana by AVIRIS sensor. Image has 145x145 pixel spatial resolution and 224 spectral bands in the range of 0.4-2.5μm bandwidth. 24 spectral band removed for water adsorption and image contains 10249 number of samples.

Experiments on Indian Pines image were compared with well-known algorithm such as Shapelet-Based Sparse Representation (Shape-DL) [40], Joint Robust Sparse Representation based Classification (JRSRC) [41], Kernel Collaborative Representation with Tikhonov Regularization with Composite Kernel (KCRT-CK) [42] and Superpixel-based Classification via Multiple Kernels (SC-MK) [43]. Ground truth with 16 classes and number of samples in each class which used in experiments are described in Table II.

TABLE II. GROUND TRUTH CLASSES FOR THE INDIAN PINES SCENE AND THEIR RESPECTIVE SAMPLES NUMBER

Class	**Class Name**	**Samples**	**%1**		**%5**		**%10**		**%15**	
			Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	Alfalfa	46	1	45	3	43	5	41	7	39
2	Corn-no till	1428	15	1413	72	1356	143	1285	215	1213
3	Corn-mintill	830	9	821	42	788	83	747	125	705
4	Corn	237	3	234	12	225	24	213	36	201
5	Grass-pasture	483	5	478	25	458	49	434	73	410
6	Grass-trees	730	8	722	37	693	73	657	110	620
7	Grass-pasture-mowed	28	1	27	2	26	3	25	5	23
8	Hay-windrowed	478	5	473	24	454	48	430	72	406
9	Oats	20	1	19	1	19	2	18	3	17
10	Soybean-notill	972	10	962	49	923	98	874	146	826
11	Soybean-mintill	2455	25	2430	123	2332	246	2209	369	2086
12	Soybean-clean	593	6	587	30	563	60	533	89	504
13	Wheat	205	3	202	11	194	21	184	31	174
14	Woods	1265	13	1252	64	1201	127	1138	190	1075
15	Buildings-Grass-Trees-Drives	386	4	382	20	366	39	347	58	328
16	Stone-Steel-Towers	93	1	92	5	88	10	83	14	79
	Total	10249	110	10139	520	9729	1031	9218	1543	8706

Table III. shows comparison of class based classification accuracy on 10 different algorithms which 4 of them were presented in literature and remains were produced in this study. In Table III. classification accuracy for each class represented in rows, applied methods represented in columns and Average Accuracy(AA), Overall Accuracy(OA) and Kappa value is given below classes. As shown in Table III. proposed GAN-SVM method accounts improved class based classification accuracy and overall classification accuracy.

TABLE III. CLASS BASED CLASSIFICATION ACCURACY COMPARISONS ON INDIAN PINES IMAGE

Class	SVM	G-SVM	N-SVM	AN-SVM	GN-SVM	Shape-DL [40]	JRSRC [41]	KCRT-CK [42]	SC-MK [43]	Proposed GAN-SVM
1	58.97	94.87	56.41	76.92	92.31	94.38	93.23	98.15	100.00	**100.00**
2	83.61	97.94	84.02	85.26	98.35	95.84	86.39	98.12	97.11	**99.26**
3	70.92	96.60	74.75	83.26	97.87	97.12	92.08	**99.76**	97.65	98.16
4	63.18	93.03	57.71	58.21	98.01	97.52	99.52	95.30	97.82	**100.00**
5	92.70	**99.03**	94.89	94.40	98.78	98.21	90.00	96.98	96.38	98.54
6	96.61	96.45	98.55	99.03	99.68	99.76	99.75	99.46	**100.00**	99.84
7	95.65	100.00	82.61	91.30	100.00	96.09	93.85	96.15	100.00	**100.00**
8	98.28	99.75	99.51	100.00	100.00	99.80	99.91	100.00	100.00	**100.00**
9	100.00	93.33	60.00	66.67	86.67	94.44	80.00	100.00	100.00	**100.00**
10	68.64	96.85	77.24	82.69	97.58	97.06	95.78	96.28	93.35	**98.91**
11	84.62	**99.28**	78.68	88.93	98.66	99.18	92.67	98.10	99.02	99.14
12	83.53	97.62	82.14	85.91	98.21	94.51	87.86	**99.02**	97.80	98.41
13	99.43	90.23	98.28	98.85	97.70	99.89	**100.00**	99.53	99.60	98.85
14	95.81	98.42	93.86	96.00	99.72	99.88	98.38	99.92	**99.98**	99.93
15	64.02	96.95	63.11	79.27	99.39	97.05	86.90	**99.47**	97.56	99.09
16	91.14	98.73	86.08	78.48	98.73	93.06	**100.00**	67.37	97.15	98.73
AA	84.20	96.82	80.49	85.32	97.60	97.11	93.52	96.47	98.34	**99.30**
OA	84.07	97.81	83.46	88.50	98.64	97.99	93.23	98.22	98.06	**99.18**
Kappa	0.82	0.97	0.81	0.87	0.98	0.97	0.92	0.98	0.98	**0.99**

Table IV. shows overall classification accuracy on 6 different algorithms which were produced in this study with 4 different ratios of training sample.

TABLE IV. CLASSIFICATION ACCURACY COMPARISONS ON INDIAN PINES

Percentage	1%	5%	10%	15%
SVM	64.28	78.60	83.03	84.07
GAUSS-SVM (G-SVM)	76.02	92.06	96.50	97.81
NONLM-SVM (N-SVM)	57.79	76.14	82.04	83.46
ANONLM-SVM (AN-SVM)	66.05	80.81	86.12	88.50
GAUSS-NONLM-SVM (GN-SVM)	85.04	95.87	97.54	98.64
GAUSS-ANONLM-SVM (GAN-SVM)	**87.38**	**97.87**	**99.00**	**99.18**

Fig. 3 shows classification maps of Indian Pines image. In Figure 3. a-g represents SVM, Shape-DL, JRSRC, KCRT-CK, SC-MK, Proposed and Groundtruth respectively. As shown in Table III. and Fig. 3, the classification accuracy of Indian Pines data set is improved by the proposed method and the resulting classification map is more homogeneous and spatially consistent.

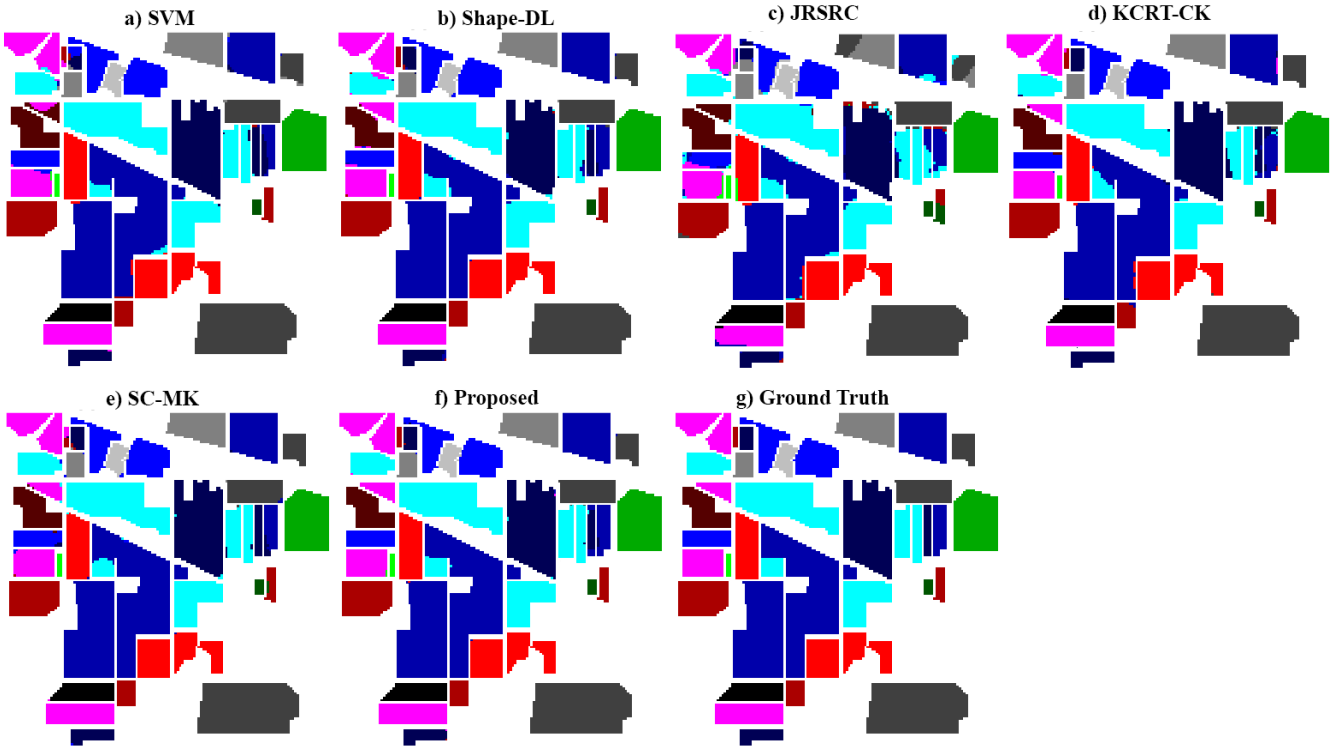

Figure 3. a)-f) Classification maps and corresponding overall accuracies obtained by SVM (84.07%), Shape-DL (97.99%), JRSRC (93.23%), KCRT-CK (98.22%), SC-MK (98.06%), Proposed GAN-SVM (99.18%). g) ground truth of the Indian Pines data. The ratio of training samples used here is 15% per class.

B. Experiment on Salinas Image

Salinas hyperspectral image obtained from Salinas Valley located at California by AVIRIS sensor.

Image has 512x217 pixel spatial resolution and 224 spectral band in the range of 0.4-2.5 μm bandwidth. 20 spectral bands removed for water adsorption and image contains 54129 numbers of samples.

Experiments on Salinas image were compared with well-known algorithm such as Superpixel-based Classification via Multiple Kernels (SC-MK) [43], Extended Random Walker-based Classification (ERW) [44] and Discontinuity Preserving Relaxation Scheme (ppMLRpr) [45]. Ground truth with 16 classes and number of samples in each class described in Table V.

TABLE V. GROUND TRUTH CLASSES FOR THE SALINAS SCENE AND THEIR RESPECTIVE SAMPLES NUMBER

Class	Class Name	Samples	%1		%5		%10	
			Training	Testing	Training	Testing	Training	Testing
1	Brocoli green weeds 1	2009	21	1988	101	1908	201	1808
2	Brocoli green weeds 2	3726	38	3688	187	3539	373	3353
3	Fallow	1976	20	1956	99	1877	198	1778
4	Fallow rough plow	1394	14	1380	70	1324	140	1254
5	Fallow smooth	2678	27	2651	134	2544	268	2410
6	Stubble	3959	40	3919	198	3761	396	3563
7	Celery	3579	36	3543	179	3400	358	3221
8	Grapes untrained	11271	113	11158	564	10707	1128	10143
9	Soil vineyard develop	6203	63	6140	311	5892	621	5582
10	Corn senesced green weeds	3278	33	3245	164	3114	328	2950
11	Lettuce romaine 4wk	1068	11	1057	54	1014	107	961
12	Lettuce romaine 5wk	1927	20	1907	97	1830	193	1734
13	Lettuce romaine 6wk	916	10	906	46	870	92	824
14	Lettuce romaine 7wk	1070	11	1059	54	1016	107	963
15	Vinyard untrained	7268	73	7195	364	6904	727	6541
16	Vinyard vertical trellis	1807	19	1788	91	1716	181	1626
	Total	54129	549	53580	2713	51416	5418	48711

Table VI. shows comparison of class based classification accuracy on 9 different algorithms which 3 of them were presented in literature and remains were produced in this study. In Table VI. classification accuracy for each class represented in rows, applied methods represented in columns and Average Accuracy(AA), Overall Accuracy(OA) and Kappa value is given below classes. As shown in Table VI. proposed GAN-SVM method accounts improved class based classification accuracy and overall classification accuracy.

TABLE VI. CLASS BASED CLASSIFICATION ACCURACY COMPARISONS ON SALINAS IMAGE

Class	SVM	G-SVM	N-SVM	AN-SVM	GN-SVM	SC-MK [43]	ERW [44]	ppMLRpr [45]	Proposed GAN-SVM
1	99.72	99.78	99.78	99.83	100.00	100.00	88.6	99.90	**100.00**
2	99.97	99.97	99.82	99.70	100.00	100.00	100.00	96.67	**100.00**
3	99.72	99.83	99.66	99.94	100.00	100.00	99.95	96.08	**100.00**
4	99.44	**99.68**	99.28	99.36	98.80	98.62	97.51	98.90	99.60
5	99.50	99.75	99.17	99.29	99.67	98.74	88.54	97.42	**99.79**
6	99.89	100.00	99.92	99.92	100.00	99.74	100.00	98.49	**100.00**
7	99.57	100.00	99.66	99.88	100.00	99.92	99.97	98.25	**100.00**
8	89.63	96.41	88.31	88.10	99.42	99.81	**99.96**	84.68	99.90
9	99.98	100.00	99.95	99.79	100.00	99.95	100.00	97.46	**100.00**
10	96.71	98.78	96.37	97.49	99.97	97.65	99.76	92.17	**100.00**
11	98.75	98.44	99.17	97.09	100.00	95.77	100.00	99.58	**100.00**
12	99.94	100.00	99.71	99.88	100.00	100.00	100.00	98.94	**100.00**
13	98.42	100.00	98.67	99.27	99.27	98.15	94.68	98.73	**100.00**
14	95.12	99.27	97.20	98.23	99.58	91.31	72.12	97.13	**100.00**
15	74.84	92.74	70.86	77.10	99.16	99.78	**100.00**	90.55	99.92
16	99.02	99.69	99.14	99.94	100.00	100.00	100.00	96.93	**100.00**
AA	96.89	99.02	96.67	97.18	99.74	98.72	96.32	96.37	**99.95**
OA	93.98	98.11	93.17	94.07	99.70	99.38	98.00	93.79	**99.95**
Kappa	0.93	0.98	0.92	0.93	1.00	0.99	0.97	0.93	**1.0**

Table VII. shows overall classification accuracy on 6 different algorithms which were produced in this study with 3 different ratios of training sample.

TABLE VII. OVERALL CLASSIFICATION ACCURACY COMPARISONS ON SALINAS IMAGE

Percentage	1%	5%	10%
SVM	89.50	93.08	93.98
GAUSS-SVM (G-SVM)	96.11	97.49	98.11
NONLM-SVM (N-SVM)	89.15	92.07	93.17
ANONLM-SVM (AN-SVM)	89.89	93.40	94.07
GAUSS-NONLM-SVM (GN-SVM)	97.45	99.40	99.70
GAUSS-ANONLM-SVM (GAN-SVM)	**98.84**	**99.84**	**99.95**

Fig. 4 shows classification maps of Salinas Valley image. In Figure 3. a-f represents SVM, SC-MK, ERW, ppMLRpr, Proposed and Groundtruth respectively. As shown in Table VI. and Fig. 4, the classification accuracy of Salinas Valley data set is improved by the proposed method and the resulting classification map is more homogeneous and spatially consistent.

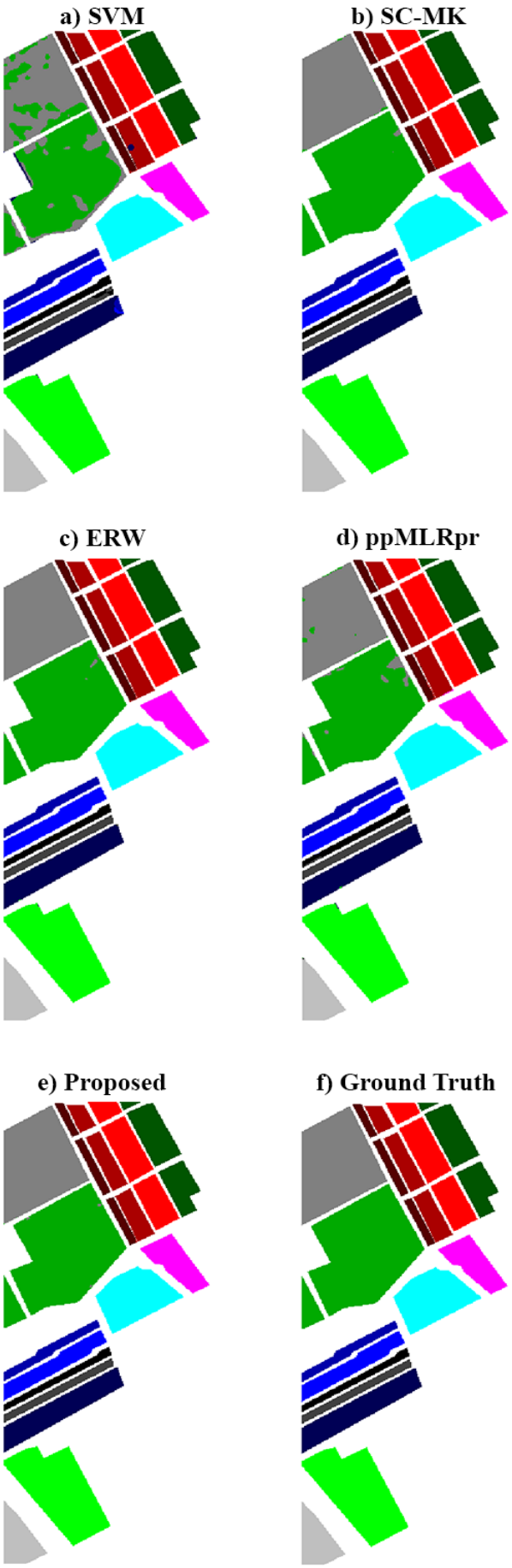

Figure 4. a)-e) Classification maps and corresponding overall accuracies obtained by SVM (93.98%), SC-MK (99.38%), ERW (98.00%), ppMLRpr (93.79%), Proposed GAN-SVM (99.95%). f) ground truth of the Salinas Valley data. The ratio of training samples used here is 10% per class.

V. Conclusions and Future Works

In this paper, we proposed three-dimensional filtering based spatial-spectral feature extraction method for hyperspectral image classification. Firstly, local and non-local neighborhood based features extracted by using FPCA. After that extracted features are combined for improving SVM classification accuracy. As shown in experimental results, proposed method presenting superior classification performance on two popular public hyperspectral image data. Future work is planned to investigate more effective ways to finding of non-local neighborhood.

References

[1] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral Image Denoising via Noise-Adjusted Iterative Low-Rank Matrix Approximation," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 3050-3061, 2015. doi:10.1109/jstars.2015.2398433

[2] C. Aswathy, V. Sowmya, and K. P. Soman, "Hyperspectral Image Denoising Using Low Pass Sparse Banded Filter Matrix for Improved Sparsity Based Classification," Procedia Computer Science, vol. 58, pp. 26-33, // 2015. doi:10.1016/j.procs.2015.08.005

[3] B. Cui, X. Ma, X. Xie, G. Ren, and Y. Ma, "Classification of visible and infrared hyperspectral images based on image segmentation and edge-preserving filtering," Infrared Physics & Technology, vol. 81, pp. 79-88, 3// 2017. doi:10.1016/j.infrared.2016.12.010

[4] S. Srivatsa, A. Ajay, C. K. Chandni, V. Sowmya, and K. P. Soman, "Application of Least Square Denoising to Improve ADMM Based Hyperspectral Image Classification," Procedia Computer Science, vol. 93, pp. 416-423, // 2016. doi:10.1016/j.procs.2016.07.228

[5] J. Huang, Y. Ma, X. Mei, and F. Fan, "A hybrid spatial-spectral denoising method for infrared hyperspectral images using 2DPCA," Infrared Physics & Technology, vol. 79, pp. 68-73, 11// 2016. doi:10.1016/j.infrared.2016.09.009

[6] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," IEEE transactions on geoscience and remote sensing, vol. 52, no. 5, pp. 2666-2677, 2014. doi:10.1109/TGRS.2013.2264508

[7] X. Li, J. Pan, Y. He, and C. Liu, "Bilateral filtering inspired locality preserving projections for hyperspectral images," Neurocomputing, vol. 164, pp. 300-306, 9/21/ 2015. doi:10.1016/j.neucom.2015.01.021

[8] Y.-L. Chang, J.-N. Liu, C.-C. Han, and Y.-N. Chen, "Hyperspectral Image Classification Using Nearest Feature Line Embedding Approach," IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 1, pp. 278-287, 2014. doi:10.1109/tgrs.2013.2238635

[9] H.-Y. Huang and B.-C. Kuo, "Double Nearest Proportion Feature Extraction for Hyperspectral-Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 11, pp. 4034–4046, 2010. doi:10.1109/tgrs.2010.2058580

[10] S. T. Tu, J. Y. Chen, W. Yang, and H. Sun, "Laplacian Eigenmaps-Based Polarimetric Dimensionality Reduction for SAR Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 1, pp. 170-179, 2012. doi:10.1109/tgrs.2011.2168532

[11] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-Preserving Discriminant Analysis in Kernel-Induced Feature Spaces for Hyperspectral Image Classification," IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 5, pp. 894-898, 2011. doi:10.1109/lgrs.2011.2128854

[12] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-Preserving Dimensionality Reduction and Classification for Hyperspectral Image Analysis," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 4, pp. 1185-1198, 2012. doi:10.1109/tgrs.2011.2165957

[13] R. Luo, W. Liao, and Y. Pi, "Discriminative Supervised Neighborhood Preserving Embedding Feature Extraction for Hyperspectral-image Classification," TELKOMNIKA Indonesian Journal of Electrical Engineering, vol. 10, no. 5, pp. 1051–1056, 2012. doi:10.11591/telkomnika.v10i5.1346

[14] Y. Wei et al., "Hyperspectral image classification using FPCA-based kernel extreme learning machine," Optik - International Journal for Light and Electron Optics, vol. 126, no. 23, pp. 3942-3948, 2015. doi:10.1016/j.ijleo.2015.07.184

[15] C.-C. Chang and C.-J. Lin, "Libsvm," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 1-27, 2011. doi:10.1145/1961189.1961199

[16] Y. Bazi and F. Melgani, "Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 11, pp. 3374-3385, 2006. doi:10.1109/tgrs.2006.880628

[17] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 3, pp. 247-259, 2011. doi:10.1016/j.isprsjprs.2010.11.001

[18] B. Demir and S. Erturk, "Hyperspectral Image Classification Using Relevance Vector Machines," IEEE Geoscience and Remote Sensing Letters, vol. 4, no. 4, pp. 586-590, 2007. doi:10.1109/lgrs.2007.903069

[19] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," IEEE Trans Syst Man Cybern B Cybern, vol. 42, no. 2, pp. 513-29, Apr 2012. doi:10.1109/TSMCB.2011.2168604

[20] M. Han and B. Liu, "A Remote Sensing Image Classification Method Based on Extreme Learning Machine Ensemble," Lecture Notes in Computer Science, vol. 7951, pp. 447-454, 2013. doi:10.1007/978-3-642-39065-4_54

[21] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote-sensing image classification," Remote Sensing Letters, vol. 4, no. 9, pp. 853-862, 2013. doi:10.1080/2150704x.2013.805279

[22] M. J. McDonnell, "Box-filtering techniques," Computer Graphics and Image Processing, vol. 17, no. 1, pp. 65-70, 1981. doi:10.1016/s0146-664x(81)80009-3

[23] P. Saint-Marc, J. S. Chen, and G. Medioni, "Adaptive smoothing: a general tool for early vision," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, pp. 618-624, 1989. doi:10.1109/cvpr.1989.37910

[24] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pp. 629-639, 1990. doi:10.1109/34.56205

[25] L. Yaroslavsky, Digital Picture Processing - An Introduction. Springer-Verlag Berlin Heidelberg, 1985. doi:10.1007/978-3-642-81929-2

[26] J.-S. Lee, "Digital image smoothing and the sigma filter," Computer Vision, Graphics, and Image Processing, vol. 24, no. 2, pp. 255-269, 1983. doi:10.1016/0734-189x(83)90047-6

[27] C. Tomasi and R. Manduchi, "Bilateral Filtering for gray and color images," in ICCV '98, Sixth International Conference on Computer Vision, Washington DC, 1998. doi:10.1109/ICCV.1998.710815

[28] S. M. Smith and J. M. Brady, "SUSAN - A New Approach to Low Level Image Processing," International Journal of Computer Vision, vol. 23, no. 1, pp. 45-78, 1997. doi:10.1023/a:1007963824710

[29] M. Elad, "On the origin of the bilateral filter and ways to improve it," IEEE Trans Image Process, vol. 11, no. 10, pp. 1141-51, 2002. doi:10.1109/TIP.2002.801126

[30] A. Buades, B. Coll, and J. M. Morel, "A Review of Image Denoising Algorithms, with a New One," Multiscale Modeling & Simulation, vol. 4, no. 2, pp. 490-530, 2005. doi:10.1137/040616024

[31] J. V. Manjon, P. Coupe, L. Marti-Bonmati, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," J Magn Reson Imaging, vol. 31, no. 1, pp. 192-203, Jan 2010. doi:10.1002/jmri.22003

[32] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," IEEE Trans Med Imaging, vol. 27, no. 4, pp. 425-41, Apr 2008. doi:10.1109/TMI.2007.906087

[33] B. M. ter Haar Romeny, Front-End Vision and Multi-Scale Image Analysis. Springer Netherlands, 2003. doi:10.1007/978-1-4020-8840-7

[34] H. L. Shang, "A survey of functional principal component analysis," AStA Advances in Statistical Analysis, vol. 98, no. 2, pp. 121-142, 2013. doi:10.1007/s10182-013-0213-1

[35] J.-L. Wang, J.-M. Chiou, and H.-G. Müller, "Functional data analysis," Annual Review of Statistics and Its Application, vol. 3, pp. 257-295, 2016. doi:10.1146/annurev-statistics-041715-033624

[36] J. O. Ramsay and B. W. Silverman, Functional Data Analysis, 2nd Ed. New York: Springer, 2005. doi:10.1007/b98888

[37] J. Yang, J.-y. Yang, D. Zhang, and J.-f. Lu, "Feature fusion: parallel strategy vs. serial strategy," Pattern Recognition, vol. 36, no. 6, pp. 1369-1381, 6// 2003. doi:10.1016/S0031-3203(02)00262-5

[38] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe. 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3 [Online]. Available: doi:10.4231/R7RX991C

[39] . AVIRIS Salinas Valley and ROSIS Pavia University Hyperspectral Datasets. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

[40] R. Roscher and B. Waske, "Shapelet-Based Sparse Representation for Landcover Classification of Hyperspectral Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 3, pp. 1623-1634, 2016. doi:10.1109/tgrs.2015.2484619

[41] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, "Hyperspectral Image Classification With Robust Sparse Representation," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 5, pp. 641-645, 2016. doi:10.1109/lgrs.2016.2532380

[42] W. Li, Q. Du, and M. Xiong, "Kernel Collaborative Representation With Tikhonov Regularization for Hyperspectral Image Classification," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 1, pp. 1-5, 2014. doi:10.1109/lgrs.2014.2325978

[43] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of Hyperspectral Images by Exploiting Spectral–Spatial Information of Superpixel via Multiple Kernels," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 12, pp. 6663-6674, 2015. doi:10.1109/tgrs.2015.2445767

[44] K. Xudong, L. Shutao, F. Leyuan, L. Meixiu, and J. A. Benediktsson, "Extended Random Walker-Based Classification of Hyperspectral Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 1, pp. 144-153, 2015. doi:10.1109/tgrs.2014.2319373

[45] J. Li, M. Khodadadzadeh, A. Plaza, X. Jia, and J. M. Bioucas-Dias, "A Discontinuity Preserving Relaxation Scheme for Spectral–Spatial Hyperspectral Image Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 2, pp. 625-639, 2016. doi:10.1109/jstars.2015.2470129